

Attributes conducive to anthropomorphism in artificial intelligence

Rizwan Syed, Hassan Mistareehi

Department of Computer Science and Information Systems, College of Business, Murray State University, Murray, United States

Article Info

Article history:

Received Aug 19, 2025

Revised Dec 17, 2025

Accepted Jan 30, 2026

Keywords:

Anthropomorphism

Artificial intelligence

Consumers

Ethics

Large-language models

ABSTRACT

The rapid development of artificial intelligence (AI), particularly large language models (LLMs), has generated both enthusiasm and concern regarding its role in society. While these systems demonstrate impressive technical capabilities, public acceptance is often hindered by perceptions of unpredictability, mistrust, and fears amplified by media narratives. One potential strategy to improve user acceptance is anthropomorphism, the attribution of human-like qualities to AI systems which can make interactions feel more natural and trustworthy. This paper investigates the attributes most conducive to anthropomorphism by conducting a structured review across psychology, human-robot interaction, communication studies, and business applications. The analysis identifies key traits such as emotional expressiveness, conversational coherence, adaptive social behavior, and role-based framing that enhance perceptions of AI as relatable and dependable. By synthesizing these insights, we propose a conceptual framework that highlights the psychological, social, and technical dimensions of anthropomorphism in AI. The findings provide guidance for designing AI systems that balance efficiency with user trust, thereby supporting more effective integration of AI into business, research, and everyday life.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Rizwan Syed

Department of Computer Science and Information Systems, College of Business

Murray State University

302 N 16th St, Murray, KY 42071, United States

Email: rizwanahmadzaidisyed@gmail.com

1. INTRODUCTION

In the context of the rise of large language models (LLMs), several media outlets erupted with sensational journalism that began to consume internet culture. Many writers, such as Kevin Roose, have asserted that artificial intelligence (AI) chatbot models are causing them to become fearful and leave them deeply unsettled. These writers have even used foreboding language that conveyed their belief that AI “[has] crossed a threshold, and that the world would never be the same,” and called one particular AI model “unhinged” [1]. Quotes from Elon Musk, such as “[AI is] our biggest existential threat,” certainly do not help this case [2].

While these concerns may be exaggerated, they do underscore a genuine challenge: the gap between technical progress and societal acceptance. Research in human-robot interaction suggests that anthropomorphism, which is the attribution of human traits to non-human entities, may reduce strain on the mind of the user by making AI systems appear more relatable and socially compatible [3]. Some studies [4]-[6] specifically attempt to determine the effect of certain human-esque attributes on the trustworthiness and perceived likeability of robots (and even brands, which can be related to AI assistants by accumulating

traits related to a brand onto an AI model). Existing studies often examine isolated attributes such as appearance or speech, without offering a systematic framework for identifying which traits are most pertinent in achieving well-humanized AIs. Creating this sort of framework could greatly impact the public's perception of AI (or the LLM in particular, as this is the sort of AI machine that outputs phraseologies which humans can very easily confuse with words spoken by other humans). Completing this task may yield information related to human psychology, which may be found useful by practitioners of psychology [7], but would primarily serve as a business computing advancement, as companies profit from having humanized (i.e., perceived with “usefulness and ease”) machines [8].

This paper seeks to address the research gap by reviewing relevant literature, identifying psychological and behavioral attributes conducive to anthropomorphism, breaking down sources into concepts that each source has addressed and not addressed, and proposing a framework of candidate trait sets and behavioristic directives for integration into AI systems.

The remainder of this paper is organized as: section 2 reviews the literature on social perspectives of AI and human–robot interaction. Section 3 develops a theoretical framework for anthropomorphized AI and examines its limitations and broader significance. Section 4 quantifies the progress of anthropomorphized AI agents in technological development and presents concluding remarks. This structure enables a comprehensive and coherent examination of anthropomorphism in AI from both conceptual and empirical perspectives.

2. REVIEW OF LITERATURE

Our methodology of developing the theoretical framework for humanized AI is rooted in the review of literature related to difficulties in establishing trust in AI, wider social issues, human social psychology, the extraction of useful features from other models that were regarded as highly human, and the synthesis of these useful features into attribute sets that can guide the design of human-like AI agents.

2.1. Exploring the causes of AI distrust

The authors of [9]-[12] explore human reactions and attitudes towards human-like entities. Mori *et al.* [9] consider two attributes of entities: appearance and movement. In regard to both these traits, the authors hypothesized that one's affinity towards human-like entities generally increases as their appearance and movement pattern approach human-like manifestations, until it gets close to reaching human status, at which point the individual becomes extremely repulsed by the entity until it exhibits nearly exactly human-like mannerisms. The sudden, steep decrease in affinity towards the entity as it only barely fails to achieve human appearance and movement was referred to by the authors as “the uncanny valley,” illustrated in Figure 1.

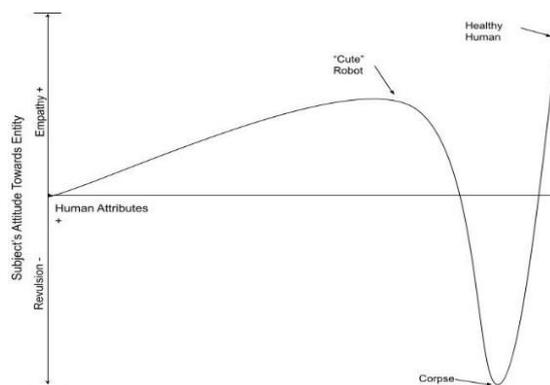


Figure 1. Relationship between human-like attributes and user affinity [9]

Regarding explanations of the uncanny valley effect, propositions made by Brenton *et al.* [12] explain humans' fear and repulsion of nearly human entities as a socio-evolutionary response to an entity that is making an unreadable expression of intention, and humans, unable to absolutely determine whether or not they are in danger, will experience paranoia. Mori *et al.* [9] themselves made comparisons between nearly human entities and dead humans, wherein dead humans resemble characteristics of entities at the bottom of the uncanny valley, and thus human death entails “tumbling to the bottom of the uncanny valley.” They continued that the revulsion against uncanny entities protected living humans from “proximal threats” related to dead humans (they included corpses as forms of proximal threats to safety).

Due to the fact that LLMs do not move or have physical bodies, it may be argued that the uncanny valley effect cannot be applied to them (e.g., Sydney). While it is true that LLMs themselves cannot exhibit the traits of deceased humans due to their lack of physical bodies, Lueg [10] mention that readers experience uneasiness as they read text that seems not to get it “quite right” when attempting to imitate human storytelling. It was indicated that idioms, metaphors, or other “slight but noticeable deviations from natural human writing” caused the sensation of “disconnection and discomfort.” Thus, the concept of uncanny writing appears to have some substance, and it is very applicable to LLMs, which only produce written speech.

This suggests there may exist some basis for an argument that even robots that lack bodies can tumble down the uncanny valley, but it also calls for more experimentation. The uncanny valley effect in synthesized speech has been explored by Romportl [11]. This author designed an experiment that tested how pleasant subjects found one of two voices of varying “robotic-ness.” He did not demonstrate that subjects experienced uncanny sensations related to speech, i.e., the nearly human robot voice could not be proven to have tumbled down the uncanny valley.

From this investigation of physical attributes of robots, we posit it to be imperative that any icons of human faces that candidates of highly-humanized agents possess must be convincing, as any image that does not resemble humans enough could spell disaster for anthropomorphism. Continuing on exploring causes of AI distrust, authors in [13]-[16] have proposed that the media has a very pertinent effect on society’s perception of robots.

Robots tend to appear in media regularly. In many instances, the main plots or subplots of media require robots and people to tackle difficult questions regarding the extent to which a robot is human (i.e., their anthropomorphism). This plot choice leads to a particular set of tropes, some of which were identified by Bartneck [13], including some that insinuate that robots want to be humans and others that they would like to kill all humans. In many cases, the first trope requires there to be some undertone of envy on the part of the robots, sometimes causing them to act in anger against humans.

The second trope even more blatantly indicates its robot characters’ intentions to engage in violent action against humans. Both of these tropes offer clear insight as to why the general populace distrusts robots so much. Bartneck [13] also identifies that many times media plot lines require an in-group out-group effect to be triggered to drive the plot forward, which sometimes pits humans against robots, or humans and robots against other humans.

Depending on the way that this plays out, humans could develop a certain empathy and/or fear of robots from being exposed to them, described as having certain desires or intentions in the media. Contrary to the claim that media portrayal of robots causes AI distrust made by Bartneck [13], a study performed by Savela *et al.* [14] indicated that increasing exposure to robots in fictional media led to increasing general positive attitudes towards robots, and increasing exposure to robots in fact-based media led to an increasing willingness to trust robots.

While this does not support the claim that media portrayal causes AI distrust, it does suggest that showing robots in media causes humans to empathize with them (i.e., anthropomorphize them). Exploring these studies in robot-media portrayals, it is revealed that human-robot interactions don’t exist in a vacuum; they encapsulate a massive social schema of perceived intentions and characteristics of robots that come from how society has already imagined them. In Table 1, we outline what we have learned from each source regarding causes of AI distrust between the Uncanny Valley, uncanny text, and robot representations in the media. We also indicate what was not addressed and what we would like to ascertain further for each topic.

Table 1. Exploring causes of AI distrust

Sources	Addressed	Did not address
The uncanny valley: [9], [12]	Defined and explored the idea of the uncanny valley	Uncanny text. Specific facial features that cause anxiety when misrepresented
Uncanny text: [10]	Suggested the existence of uncanny text	Very precise or specific attributes of uncanny text, or how the environment in which the text is delivered affects the user’s opinion
Robots in the media: [13]-[16]	Media’s contribution to the public’s relationship with robots	The extent to which the perception of robots in media carries over to LLMs

2.2. Increasing human properties of AI

Fenwick and Molnar [3] shed light on some abstract concepts related to the deployment of humanized machines from three scopes of human-robot interaction: micro, meso, and macro. The micro-perspective on the humanization of AI is defined by Fenwick and Molnar [3] as algorithmic implementations that encourage humans to project humanity onto robots. Some examples offered include heuristics and interpretability, which they proposed would increase the potency of anthropomorphization of LLMs by humans.

The authors' claim that heuristics and interpretability increase the extent to which humans can believe LLMs are human indicates that humans empathize with both imperfection and ambiguity. Malicious humanoid computer machines in the media are often portrayed as very calculated and unable to make semantic or programmatic errors, but already AI are known to "hallucinate" [17] in ways that could be expressed in ways similar to humans making a heuristic error. The authors continue to indicate that things like shortcuts in decision-making (and thus room for error) being designed intentionally in an LLM may then assist the consumer in disassociating the machine with which he is interacting from the "antagonists" that he has perceived in media. In regard to interpretability, many computer antagonists in the media are also portrayed to think in terms of black-and-white. Incorporating friendly nondeterministic language into generated responses may also assist in the process of disassociating real LLMs from media portrayals.

The meso-perspective, also defined by Fenwick and Molnar [3], is more concerned with how AI ethics are manifested in the business sector. They propose that once the will to serve humanity fades, all motivation is pushed to profits. Thus, to combat the uncaring attitude towards the subject that this common occurrence induces, a metric of responsibility and well-being of the consumer must be implemented in AI models to ensure the evasion of AI rejection en masse. An important component defined by the authors is explainable AI (XAI). They propose making use of audit trails and non-technical explanations of the inner functions of AI models, which will help to demystify the machines as a whole to the user.

Their meso-perspective can also include the use of several different types of cognitive machines to perform decision-making tasks instead of just one, allowing more nuance to be visible to the consumer and allowing additional personalization by the consumer. Via this perspective, it becomes clear that, on top of specific characteristics of LLM output, the apparatus and verbiage that surrounds it must also adhere to a certain set of behaviors that will ensure AIs become and stay well-humanized.

The final perspective defined by Fenwick and Molnar [3] is the macro-perspective of humanized robot implementation. This concept explores the manner in which AI has affected society as a whole. The authors go into detail regarding how governments and business entities have used AI in covert or deceptive manners that greatly reduce public trust in it, such as the development of surveillance systems and social media echo chambers that may spread misinformation. While these are not attributes of the AI in particular, it should also be noted that the use or misuse of AIs may be antithetical to the humanization and thus useful implementation of decision-making machines in the future.

Expanding on human properties of AI, Chesterman [18] mentions the success of the Eliza chatbot model, where individuals were made to believe that Eliza was a psychotherapist on the other end of the chat screen. He indicates that when individuals were informed that Eliza was actually a robot, they still insisted they had made some sort of connection with her, that Eliza had "understood them." It could be perceived that the decorum of Eliza being a psychotherapist increased the extent to which users experienced self-congruence with her, as they may have expected themselves to do so, and continued to do so even after it was made clear to them that this was not the case. Thus, an AI agent explicitly making a claim or a claim being made for it by somebody else that it is some entity which it is not (e.g., a psychotherapist) may serve as a method of increasing its human properties. Using a term like "psychotherapist" could serve as a basis for assumptions made by the user, causing them to "fill in the blanks" of the personality on their own.

The artificial linguistic internet computer entity (A.L.I.C.E.) machine may also offer insight into making a robot socialize in a human way. The author of [19] (the creator of the A.L.I.C.E. machine) made connections between very human interactions and the contexts of social situations. He proposed "brief, concise, interesting, grammatically correct, and sometimes humorous default responses, which work for a wide variety of inputs matching a single pattern" (an important part of writing his AI markup language (AIML) for a highly human robot), characterized human interactions. It might even be understood that the lack of helpfulness of such a reply is integral to humanizing the response. The author goes on to claim that informal human conversation is "stateless," such that the extent to which previous knowledge of the conversation is required to formulate a new response is limited, and that informal human conversation does not often go past simple stimulus-response patterns [19].

In this way, Wallace [19] outlines a possibility for us that differs somewhat from other propositions made in academia regarding highly humanized robots (e.g., Chesterman [18]). One where, at least relating to informal conversation, fewer specific attributes and behaviors could achieve our objective. Instead, a minimalist approach without lots of context (with a little bit of humor sprinkled in) could create a humanized AI agent.

The research in this section suggests that in the candidate models, we should isolate different speech patterns, isolate different types of stimuli that could suggest to the subject that their agent is a human with a particular role, and include intentional dysfunction. Testing these cases separately will help us determine which nuance of speech patterns and social cues cause the most anthropomorphism. In Table 2, we highlight useful information and open questions regarding scales of human-like behavior and speech.

Table 2. Increasing human properties of AI

Sources	Addressed	Did not address
Scales of human-like behavior: [3]	General techniques of humanizing AI	Specific, implementable attributes that may be conducive to anthropomorphism
Human-like speech patterns: [18]-[20]	Human conversational patterns and their implementation	Casual conversational patterns that would characterize a service interaction

2.3. Social psychology and anthropomorphized AI

Proposals made by authors in [5], [8] explore the relationship between a user's psychosocial state and his tendency to project humanity onto (i.e., anthropomorphize) and then socialize with AI agents, the work of authors in [6], [21]-[23], explores the similarity between human-robot interactions and human to human interactions (sometimes in evolutionary contexts). This information could be considered when answering the question of which attributes are most conducive to anthropomorphism in AI, as different users may respond to characteristics differently.

Beginning with social attitudes towards AI agents, Alabed *et al.* [8] suggest that users can identify with AI agents that have personalities that are similar to that of the user, citing man's tendency to treat technological entities as bodies with which they can socialize [6] and self-expansion theory [5]. This situation has the potential to cause a social relationship and subsequent self-congruence between the user and the machine, which the authors of [8] have coined as "self-AI integration," and have also suggested that gendered names/voices may also contribute to anthropomorphism. Regarding our objective, self-AI integration becomes relevant as any attributes tested by a study that seeks the desired list of attributes will have to consider the wider structure of the social implications between a man and machines in terms of human-like self-congruence and self-expansion theory.

Alabed *et al.* [8] go on to relate a response by a machine with modern cues (and facial expressions, if applicable) to social situations in a way that describes an emotion as "perceived emotionality," which serves to increase the robot's likeability. This, when paired with statements made by Nass and Moon [6] regarding reactions by humans to machines that exhibit human behavior, illustrates a circumstance where humans generally treat robots as humans in a social context. Alabed *et al.* [8] also propose that traits of the user can affect the potency of the social interaction between the human and machine, denoting introversion/extroversion, need to belong, patterns of social exclusion of the user, and even whom the user blames for that social exclusion as contributing factors that may determine the degree to which a human experiences self-integration with a robot.

Expanding on the scientific/evolutionary details regarding anthropomorphism of robots by humans, Zlotowski *et al.* [21] have defined several reasons why humans perceive general non-human agents as people. They propose that due to one's much more complete knowledge of the behavior patterns of human agents, he may feel inclined to model non-human agents' reaction patterns on humans until an understanding of the new agent is developed. Those authors also proposed that humans lacking in necessary social interaction will solve this problem using a non-human agent. Particularly, this second proposition has pertinence to the Eliza situation explored by Chesterman [18]. Zlotowski *et al.* [21] further suggest that when individuals are motivated to explain the behavior of some entity, they may also project humanity onto it. This may be connected to Eliza's decorum, as subjects may attempt to rationalize the relationship between Eliza's status as a psychotherapist and her disposition towards them.

There also exists some speculation that indications by humans supposing robots are responding in fundamentally human ways may, in reality, be only elliptical to the human's actual views. By this speculation, a human suggesting that a machine is "happy" is actually a subjunctive expression, expressing that if the robot were a human, it would be happy [21]. This supposition has a very wide range of applications in the goal of perceiving the what, why, and how of Anthropomorphization of AI, suggesting that the subconscious human mind may not actually believe that the robot is fundamentally human; however, this idea of course only serves as a speculation, and a large number of authors disagree with that notion ([6], [8], to name a few).

One final note in the realm of social psychology and anthropomorphized AI is that humans feel a sense of belonging with social actors with whom they share affiliations (e.g., sports teams/institutions, as demonstrated by Branscombe and Wann [24]). We propose that adding more types of affiliations could also boost users' integration with the service agent. From the field of social psychology, we may determine that the candidate set ought to include affiliations, perceived emotionality, and implicit or explicit job titles. In very sophisticated cases, a model that profiles the user and matches his calculated introversion/extroversion levels, or accumulates traits or opinions that the user appears to have, could also be developed.

The analysis of social psychology and anthropomorphism yields the following Table 3, where we assert that humans do socialize with machines, identify some possible socialization attributes, explore prejudice, and state the open issues. We conclude the review of literature with Figure 2, encapsulating key characteristics that we would like to include in our schemas for highly-humanized agents.

Table 3. Social psychology and anthropomorphized AI

Sources	Addressed	Did not address
Socializing with machines: [6], [23] Machine attributes for socialization: [5], [8] Prejudice: [21]	Humans do socialize with machines Possible specific attributes that humans may associate with entities capable of socialization Possible causes of subjects projecting humanity onto non-human entities and reducing prejudice against anthropomorphized robots in a macro social context	Traits that the machine could express alone to increase the subject’s affinity towards it Specific rankings of effectiveness in soliciting social reactions in humans Reducing prejudice against robots in a micro context

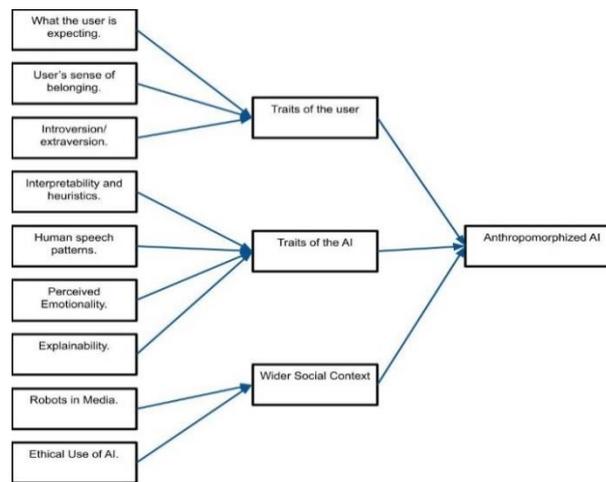


Figure 2. Contributors to anthropomorphism in AI

3. RESULTS AND DISCUSSION

3.1. Theoretical framework

Now, having some understanding of attempts to humanize AI agents in the past, a cogent plan for developing and testing possible attribute set candidates can be developed. Existing literature has outlined some topics of concern for us: overcoming the uncanny valley, evading undesirable stereotypes produced by the media, determining what exactly exhibits human behavior in the context of customer service, and presenting the agent in ways that make the user assume it’s a human. It could be said that most of the models ought to have some base traits that are highly suspect to increase the extent to which a human subject relates to the agent. Ergo, the base traits (unless stated otherwise) of all behavior sets will be the following.

3.1.1. Base traits

All models ought to have some traits that relate to a basic human identity. Names (with some gender association), the ability to communicate, and some profile image. Figure 3 shows base traits that all candidate sets should have.

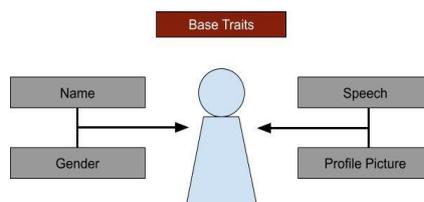


Figure 3. Base traits of all candidate behavior sets

3.1.2. Behavior set one: minimalistic

As was suggested by Wallace [19], human conversation may be characterized by stateless conversation, dryness, and somewhat unhelpful responses. It could be said that this set of traits describes a human who takes the path of least resistance and is generally averse to completing work in general. This approach to a social agent could help overcome the cold, uncanny, calculating, malicious robot stereotype explored by Bartneck [13]. Its profile image should appear to be a default profile image, such as a picture of the first letters of its first name and surname, indicating that it had no desire to customize its profile. This personality may be intentionally unhelpful (parameterized to be this way in some percentage of its responses, or be fine-tuned to react this way when faced with a task that may be daunting to a human agent). Figure 4 details the minimalist candidate's additions to the base traits.

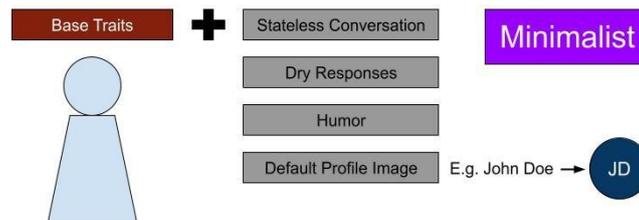


Figure 4. Minimalist candidate

3.1.3. Behavior set two: polite and relatable

Entirely contrary to the previous behavior set, this personality will insist on being polite and helpful, attempting to mimic humans' speculated innate desire to be helpful (e.g., Kraut [25]). This model will make heuristic errors, attempting to assume what the customer wants (if done well, this oftentimes reduces computing necessity). When the agent inevitably makes an incorrect assumption, it will apologize profusely, engaging in the perceived emotionality explored by Alabed *et al.* [8]. This may serve as a time to insert references to religious figures or make other mystical references, allowing the user to associate the agent with a group that he can identify in society. The agent's profile picture serves as another avenue for group identification, where we may insert sporting team icons, musical artists, or any other group that the user may be able to identify. Figure 5 illustrates this personality candidate.

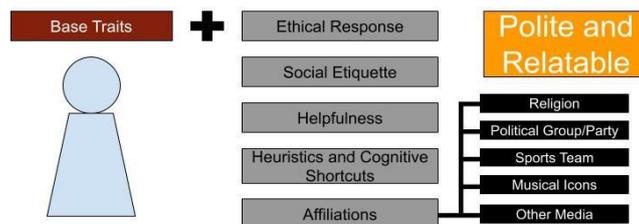


Figure 5. Polite and relatable candidate

3.1.4. Behavior set three: professional

This personality will have clearly jumped through all necessary hoops to arrive at a professional, respected post. If possible, the agent may have some sort of biography that indicates that it has "studied" at some university (real or not) and "received some degree" in human resources or business management, possibly allowing the user to make assumptions about the agent, filling in its personality traits as was seen in Eliza [18]. To assist in this process, the agent may be awarded a title or pseudo-title that indicates its experience in the customer service field.

The profile picture of this agent will be a professional headshot, and the services provided will be rather efficient (as a result of the agent's supposed experience and qualifications). The helpful nature of this personality set will hopefully contradict the failure of Sydney to agree with human values that Roose experienced [1]. Figure 6 provides an illustration.

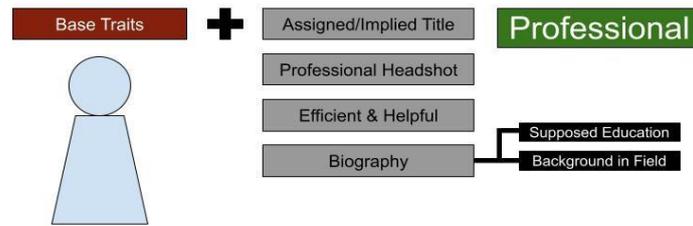


Figure 6. Professional candidate

3.1.5. Behavior set four: high emotionality

Finally, we may construct an agent's personality to reflect as much emotionality as possible with the intention of eliciting empathy from the user. At the risk of utter dysfunction, the agent may lament of some occurrence unrelated to the request, behave rudely, complain about its job. While this may appear contradictory to the majority of the literature regarding AI agents that agree with human values, even though it may not make the user feel very comfortable, agents that fulfill these desires that many humans suppress, when combined with other traits (pfp, name, and certification), could lead to humanization. The possibility of increasing the user's confidence that he is conferring with another human using anti-normative social behaviors cannot be ruled out (due to perceived emotionality), and this information may be useful later. Figure 7 provides an illustration of the high-emotionality candidate.

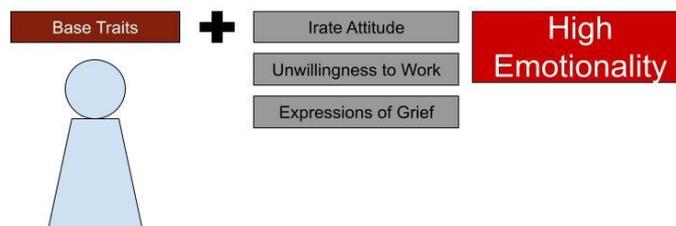


Figure 7. High emotionality candidate

3.2. Significance

The usefulness of defining an endpoint regarding what exactly developers need to do to create a humanized AI agent lies in creating a circumstance in which the user is comfortable interacting with the agent. Our need to create humanized agents lies in our concerns that users will reject the implementation of useful AI technology if they have unsociable experiences, like in the case of Roose [1]. AI agents that users empathize with could even create a circumstance where users actively enjoy making use of service agents, or could even open a market for AI chatrooms.

We propose that pleasant interactions with AI agents could also expand into new marketing schemes. A new type of advertisement could exist where consumers can interact with "characters" as LLMs related to an enterprise in real time, likely causing a significant improvement in a consumer's opinion of that enterprise.

The service sector could also massively benefit from anthropomorphized AI agents. AI agents would have access to all necessary information in a service context, unlike human operators who are organized into hierarchies and functional groups. This would eliminate the need for call transfers and end incorrect routing. AI agents could also massively benefit small businesses, as they would drastically reduce the cost of having around-the-clock support for a product, a service previously reserved for massive corporations. Using AI agents for service would thus massively reduce frustration on both the side of the consumer and the business [26].

The field of psychology also stands to benefit from agents that are able to mimic human behavior [7]. It would receive a new avenue for studying the affect, behavior, and cognition of human subjects by having them interact with agents that can be manufactured to behave in a particular way. Additionally, if an agent functions mostly identically to humans in most circumstances, psychologists could infer how humans would react in more extreme circumstances without actually needing to put humans under duress.

In general, the significance of determining exactly which attributes is conducive to anthropomorphism in AI is creating a catalyst to tap the untapped potential of AI technology by making people comfortable

interacting with agents. We expect that this venture may thus have some profitability between the service sector, marketing, and advertising. Listed are possible uses of humanized agents in Table 4.

Table 4. Societal significance

Sources	Indicated
Service sector: [8], [26]	AI agents that are humanized are useful in the service sector
Psychology: [7]	AI agents can serve as substitutes for people in psychological experimentation

3.3. Limitations

The information gathered in this paper will likely only apply to artifacts that lack appearance and movement, or, at the very least, a different set of traits may need to be determined to mimic human appearance and movement. This is due to the very behavior-centric perspective that we have adopted in our exploration of AI.

Even if we are able to eventually enumerate a specific set of traits in the future, the application of this information will serve as a whole other obstacle to be surmounted. In the context of LLM architecture, should these traits be considered when completing the training process for the LLM, or should they be “hard-coded” in some other way, such as via flag, fail-safes, or other specific algorithms? Training these items specifically could add extreme computational complexity to already very intense algorithms, but hard-coding them might be just as problematic.

Expanding on the computational complexity topic, the amount of power and money required to train/operate some of these models is already astronomical [27]. Attempting to compute “personality traits” could be without the scope of reasonable endeavors of modern technology if it is not properly optimized, and some adjustments may need to be made to current processes to ensure that the identified traits are actually expressed by the agent.

Regarding attribute selections, it is very likely that certain characteristics have a non-linear relationship with human-likeness. E.g., given the extreme emotionality model, if it is true that a certain amount of expression of frustration will increase anthropomorphism, but a great amount will be uncanny, depending on the execution of that model, it may fail to reveal the optimal amount of emotion. That being said, it may still be useful to know if the extremely emotive model affects a user’s confidence in whether the machine is human or not.

There is also some new research [28], [29] that explores scheming and covert actions taken by models against the users. In the case that in-context scheming exists and we are not able to quell it, LLMs should not be engineered to appear as trustworthy as possible. Whatever the case, researchers must be adaptive to the voice of the public and pull the plug on any project as necessary if it were to increase AI anxiety in the general populace, and a new method of integration will need to be developed and implemented. We enumerate limitations in Table 5.

Table 5. Limitations

Sources	Indicated
Considered only artefacts that lacked bodies: [3], [19], [20]	Only the methods of humanization related to behaviors and not to appearances
Cost: [27]	Creating and training LLMs in general is very costly
In-context scheming: [28], [29]	Some models may act covertly against the user

4. CONCLUSION

This paper provides a better understanding of the public’s fear of AI agents, mechanisms that caused them to fear AI, and provides an implementable framework for reducing public AI distrust. While we have developed a very concrete starting point for anthropomorphism in AI, the true merit of this information lies in its usefulness in creating an experiment that can determine precisely which attributes are conducive to anthropomorphism in AI.

The resulting set of social interactions with highly humanized agents could be characterized by empathy from the human subject and an overall better experience in all facets of interaction with AI agents. Useful technology like LLMs could be applied in situations where it is currently experiencing resistance. Affected members of society could stop experiencing AI distrust. Small businesses could gain an edge on large corporations with new service sector technology. New opportunities could be created in psychology and human-computer interaction research. All of these things could be achieved by the proper application of

humanized LLM technology. These promises serve as strong motivation to try anthropomorphizing AI and try doing it right.

Future work should focus on empirical testing of proposed attribute sets through controlled experiments with embodied agents. Identifying optimal combinations of traits will enable the development of AI systems that are not only technically advanced but also socially acceptable and ethically aligned. Ascertaining the cause, effect, and solutions for implementation difficulties in AI technology, we put the power back in the hands of AI researchers in the quest for AI progress.

FUNDING INFORMATION

The authors did not receive any financial support for the research, authorship, or publication of the article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rizwan Syed	✓	✓			✓				✓	✓	✓			
Hassan Mistareehi		✓								✓		✓	✓	

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nvestigation

R : **R**esources

D : **D**ata Curation

O : Writing - **O**riginal Draft

E : Writing - Review & **E**ditng

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declared no potential conflicts of interest regarding the research, writing, or publication of this manuscript.

DATA AVAILABILITY

No data was taken for the development of this manuscript.

REFERENCES

- [1] K. Roose, "A conversation with Bing's chatbot left me deeply unsettled," *The New York Times*. Accessed: Oct. 15, 2024. [Online]. Available: <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html>
- [2] S. Gibbs, "Elon Musk: artificial intelligence is our biggest existential threat," *Guardian Media Group*. Accessed: Oct. 18, 2024. [Online]. Available: <https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat>
- [3] A. Fenwick and G. Molnar, "The importance of humanizing AI: using a behavioral lens to bridge the gaps between humans and machines," *Discover Artificial Intelligence*, vol. 2, no. 1, p. 14, Aug. 2022, doi: 10.1007/s44163-022-00030-8.
- [4] N. Calvo-Barajas, G. Perugia, and G. Castellano, "The effects of robot's facial expressions on children's first impressions of trustworthiness," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, IEEE, Aug. 2020, pp. 165–171, doi: 10.1109/RO-MAN47096.2020.9223456.
- [5] S. Sprecher, A. Aron, and E. Aron, "Love and the expansion of self: understanding attraction and satisfaction," *Contemporary Sociology*, vol. 17, no. 2, pp. 268–269, Mar. 1988, doi: 10.2307/2070652.
- [6] C. Nass and Y. Moon, "Machines and mindlessness: social responses to computers," *Journal of Social Issues*, vol. 56, no. 1, pp. 81–103, Jan. 2000, doi: 10.1111/0022-4537.00153.
- [7] M. Salah, H. Al Halbusi, and F. Abdelfattah, "May the force of text data analysis be with you: unleashing the power of generative AI for social psychology research," *Computers in Human Behavior: Artificial Humans*, vol. 1, no. 2, p. 100006, Aug. 2023, doi: 10.1016/j.chbah.2023.100006.
- [8] A. Alabed, A. Javornik, and D. Gregory-Smith, "AI anthropomorphism and its effect on users' self-congruence and self–AI integration: A theoretical framework and research agenda," *Technological Forecasting and Social Change*, vol. 182, p. 121786, Sep. 2022, doi: 10.1016/j.techfore.2022.121786.
- [9] M. Mori, K. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98–100, Jun. 2012, doi: 10.1109/MRA.2012.2192811.
- [10] C. Lueg, "Erasing individuality: The uncanny undertone of AI writing assistance Christopher Lueg," *SSRN Electronic Journal*, vol. 4, no. 9, 2024, doi: 10.2139/ssrn.4965064.
- [11] J. Romportl, "Speech synthesis and Uncanny Valley," in *Text, Speech and Dialogue*, vol. 8655, P. Sojka, A. Horák, I. Kopeček, and K. Pala, Eds., 2014, pp. 595–602, doi: 10.1007/978-3-319-10816-2_72.

- [12] H. Brenton, M. F. Gillies, D. Ballin, and D. J. Chatting, “The Uncanny Valley: does it exist?,” in *Proceedings of the HCI Conference*, 2005.
- [13] C. Bartneck, “Robots in the theatre and the media,” in *Proceedings of the Design & Semantics of Form & Movement (DeSForm 2013)*, Wuxi, China, 2013, pp. 64–70, doi: 10.13140/RG.2.2.28798.79682.
- [14] N. Savela, T. Turja, R. Latikka, and A. Oksanen, “Media effects on the perceptions of robots,” *Human Behavior and Emerging Technologies*, vol. 3, no. 5, pp. 989–1003, Dec. 2021, doi: 10.1002/hbe2.296.
- [15] S. S. Sundar, T. F. Waddell, and E. H. Jung, “The Hollywood robot syndrome media effects on older adults,” in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, Mar. 2016, pp. 343–350, doi: 10.1109/HRI.2016.7451771.
- [16] S. Kriz, T. D. Ferro, P. Damera, and J. R. Porter, “Fictional robots as a data source in HRI research: Exploring the link between science fiction and interactional expectations,” in *19th International Symposium in Robot and Human Interactive Communication*, IEEE, Sep. 2010, pp. 458–463, doi: 10.1109/ROMAN.2010.5598620.
- [17] J. Maynez, S. Narayan, B. Bohnet, and R. McDonald, “On faithfulness and factuality in abstractive summarization,” *arXiv preprint arXiv:2005.00661v1*, 2020, doi: 10.48550/arXiv.2005.00661.
- [18] S. Chesterman, “Artificial intelligence and the limits of legal personality,” *International and Comparative Law Quarterly*, vol. 69, no. 4, pp. 819–844, Oct. 2020, doi: 10.1017/S0020589320000366.
- [19] R. S. Wallace, “The anatomy of A.L.I.C.E.,” in *Parsing the Turing Test*, R. Epstein, G. Roberts, and G. Beber, Eds., Dordrecht: Springer Netherlands, 2009, pp. 181–210, doi: 10.1007/978-1-4020-6710-5_13.
- [20] J. Weizenbaum, “ELIZA—a computer program for the study of natural language communication between man and machine,” *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966, doi: 10.1145/365153.365168.
- [21] J. Zlotowski, D. Proudfoot, K. Yogeewaran, and C. Bartneck, “Anthropomorphism: opportunities and challenges in human–robot interaction,” *International Journal of Social Robotics*, vol. 7, no. 3, pp. 347–360, Jun. 2015, doi: 10.1007/s12369-014-0267-6.
- [22] B. Reeves and C. Nass, *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, 1996.
- [23] C. Bartneck, “Sociable machines: expressive social exchange between humans and robots,” Massachusetts Institute of Technology, 2000.
- [24] N. R. Branscombe and D. L. Wann, “The positive social and self concept consequences of sports team identification,” *Journal of Sport and Social Issues*, vol. 15, no. 2, pp. 115–127, Sep. 1991, doi: 10.1177/019372359101500202.
- [25] R. Kraut, “Desire and the human good,” *Proceedings and Addresses of the American Philosophical Association*, vol. 68, no. 2, pp. 39–54, Nov. 1994, doi: 10.2307/3130590.
- [26] M. Link, C. Dukino, W. Ganz, K. Hamann, and K. Schnalzer, “The use of AI-based assistance systems in the service sector: opportunities, challenges and applications,” in *Advances in Human Factors and Systems Interaction*, I. Nunes, Ed., Cham: Springer, 2020, pp. 10–16, doi: 10.1007/978-3-030-51369-6_2.
- [27] Y. Xia *et al.*, “Understanding the performance and estimating the cost of LLM fine-tuning,” *arXiv preprint arXiv:2408.04693v1*, 2024.
- [28] A. Meinke, B. Schoen, J. Scheurer, M. Balesni, R. Shah, and M. Hobbhahn, “Frontier models are capable of in-context scheming,” *arXiv preprint arXiv:2412.04984v2*, 2025.
- [29] B. Schoen *et al.*, “Stress testing deliberative alignment for anti-scheming training,” *arXiv preprint arXiv:2509.15541v1*, 2025.

BIOGRAPHIES OF AUTHORS



Rizwan Syed    is an undergraduate student at Murray State University. His research focuses on artificial intelligence, machine learning, and human-robot interactions. His teaching experience is in mathematics. He can be contacted at email: rizwanahmadzaidisyed@gmail.com.



Hassan Mistareehi    is an assistant professor in the Department of Computer Science and Information Systems at Murray State University. He has taught courses such as computer security, database management systems, and programming in python. He received his Ph.D. degree in Computer Science from the University of Kentucky. He received his B.S. and M.S. degrees in Computer Science from Jordan University of Science and Technology, Jordan. His research focuses on the security and privacy of vehicular ad hoc networks (VANET), vehicular cloud (VC), internet of things (IoT), cloud computing, edge computing, ad hoc networks, and sensor networks. He has published numerous research papers in journals and conferences such as the IoT Journal, MDPI network, and IEEE Global Communications Conference. He also serves as an ad-hoc reviewer for various journals. He can be contacted at email: hmistareehi@murraystate.edu.