SYSTEMATIC REVIEW | Article in Press

Emotions and anthropomorphism towards robots: a scoping review

Cecilia Roselli · Ivano Zanzarella · Pasquale Grieco · Edoardo Datteri

Received: 16 June 2026 / Accepted: 16 October 2026
© The Author(s) 2026

Abstract

Human interaction with non-human entities, such as robots, has long been characterised by anthropomorphism – namely, the tendency to attribute human-like mental states to inanimate objects. While humans appear evolutionarily hardwired to use anthropomorphism as the default strategy for navigating interactions with their peers, significant gaps remain in the Human-Robot Interaction (HRI) field in defining and operationalising "emotions" and their relationship to psychological anthropomorphism towards robots. This scoping review aims to systematically map the existing HRI literature and see how emotions are investigated in relation to robots. By reviewing relevant empirical studies, it addresses four core research questions focusing on: (i) the conceptualisation of emotions and anthropomorphism towards robots; (ii) prevailing theoretical models and definitions; (iii) methodologies used to investigate emotions in HRI; and (iv) key findings in the literature regarding the emotions-anthropomorphism link. This review contributes to broaden current discussions about how emotions are conceptualised and analysed in HRI research, by which it would be possible to contribute to extend knowledge about how people emotionally interact with robots by considering their anthropomorphism.

Keywords Emotion · Anthropomorphism · Human-robot interaction · Scoping review

Introduction

In 1944, Heider and Simmel conducted a pivotal study in which they created a short, silent animation using four simple geometric shapes: a large triangle, a small circle, and a rectangle with a moving segment. All these shapes moved across the scene, and participants were asked to observe them and then describe what they saw. Surprisingly, what the authors found was that people displayed a strong tendency to impose narrative meanings even on simple geometric shapes, i.e., by attributing social contexts, meaningful interactions, and even emotions to them [1]. For example, the circular movements of the circle and the small triangle around each other, and their touching each other, were interpreted by some participants as expressions of joyful emotions. Otherwise, when asked why the big triangle suddenly broke the house (i.e., when the rectangle's segment moved, giving the impression of a door opening), the majority of participants reported that it was the effect of the big triangle getting mad and thwarted because the circle and the small triangle had escaped [1].

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

https://doi.org/10.1186/s40359-026-05844-0

Published online: 10 November 2026

In the last decades, with the advance of technology, the investigations into individuals attributing lifelike qualities, including emotions, to other (technological) entities beyond humans have moved from simple geometrical shapes to more sophisticated artefacts, such as robots. People's fascination towards automata, or "self-moving machines", was established a few centuries ago, starting with Vaucanson's life-sized flute player (1738), which was presented as one of the most groundbreaking biomorphic automata – it genuinely played the instrument using the same physical principles as a human musician [2]. The mimicry of some of its biological functions (i.e., the automaton physically exhaled air through mechanical lungs, shifted lips, and its moving metal tongue moved to articulate notes) led people to perceive it as possessing a form of artificial or simulated life [3, 4], and attribute soulfulness, sensibility, and emotional intentionality to its performances [3].

More recently, Breazeal developed Kismet, a pioneering robot head that was largely recognised as the first truly social robot, specifically designed to engage in natural, expressive, face-to-face social interactions with humans (2004). Kismet served as a seminal case study to test and guide human attribution of emotions. Indeed, when interacting with it, Breazeal pointed out that people did not treat Kismet as a purely mechanical device but started treating it as a social agent [5], by attributing internal complex emotions to it (e.g., loneliness).

Interestingly, the rationale behind Kismet's design was the following: Breazeal argued that humans are evolutionarily hardwired to interpret expressive behaviour as social signals, and that a robot with a socially evocative design could facilitate anthropomorphic attribution. In other words, humans tend to *anthropomorphise* technology – that is, they tend to attribute human-like mental states to technological, non-human artefacts [6]. A similar observation was also previously made by Heider and Simmel [1]: they noticed that participants tended to use a few "anthropomorphic" words to describe the behaviour of the geometrical shapes, since a description in purely geometrical terms would be too complicated. This may be because attributing lifelike, anthropomorphic characteristics to non-human entities is the default strategy to interpret others' behaviour, i.e., the strategy that humans usually exploit to navigate others and explain what they do in a fast and efficient manner [7].

The above-mentioned examples seem to show that a relationship between emotions and anthropomorphism towards robots exists. However, the current literature shows at least two blurred points and gaps. The first, more theoretical point is that these two theoretical constructs seem to be conceptually and, in some cases, definitionally connected in some of the most influential frameworks. For example, Epley and colleagues [6] treat anthropomorphism as the broader construct: the tendency to attribute human-like characteristics, *including emotions*, to the real or imagined behaviours of non-human agents. In this case, "emotions" fall under this umbrella rather than standing apart. Lately, Kühne and Peter's [8] multidimensional model has made this even more explicit by grounding it in Wellman's Theory-of-Mind (ToM) framework (1990): according to this view, anthropomorphism in HRI consists of attributing thinking, perceiving, desiring, choosing, and *feelings* to a robot. Thus, the relationship between emotions and anthropomorphism is conceptual, and not essentially empirical, in these studies. However, other evidence suggests that the conceptual distinction between the two can be sharpened. For example, in their study, Spatola and Wudarczyk , [9] distinguished between primary vs. secondary emotions: the latter (e.g., guilt, jealousy), unlike primary emotions (e.g., fear, joy), are ascribed uniquely to humans. The authors found that, at the explicit level, the attribution of secondary emotions to robots was linked to higher anthropomorphism; at the implicit level, higher primary/secondary emotion attribution is linked to lower anthropomorphism towards robots. In other words, the authors established an "empirical" link between emotion and anthropomorphism, by which they suggested that these two constructs are conceptually and logically independent. Thus, emotions (in terms of emotion attribution) cannot be considered only a sub-dimension of anthropomorphism, but also a marker – which is graded, with the type of emotion (primary vs. secondary) differentially predicting anthropomorphic perception. Taken together, what emerged from this mixed evidence is that, at present, HRI evidence shows different manners in which these two constructs, i.e., emotions and anthropomorphism, can (or cannot) be linked to each other.

One potential approach is to disentangle various dimensions of anthropomorphism, which stands for at least two different kinds of concepts that usually get conflated: (i) *anthropomorphism by design*, i.e., the robot's built-in

human-like features (for example, voice, face, and gestures); and (ii) *psychological anthropomorphism*, i.e., humans' attribution of anthropomorphism to others, in this case robots (e.g., "I believe that the robot feels") ([10]; see [11] for a review). This distinction matters because empirical results diverge sharply depending on which one is chosen. For example, Schömbs and colleagues [12] directly manipulated anthropomorphism by design and found no significant influence of anthropomorphism on participants' empathetic behaviours towards robots. However, their exploratory analysis indicated that individuals' tendency to anthropomorphise robots might be crucial for empathy towards robots, suggesting that while anthropomorphism by design failed as a predictor of emotions towards robots (in terms of empathetic engagement), psychological anthropomorphism may succeed. There is also contrasting evidence, such as the study of [13]; which manipulates the robot's appearance (a design factor) and found that people more frequently display positive emotions towards robots (i.e., empathise with them) if they have a more human-like appearance. Thus, current HRI literature does not cleanly point toward null effects of anthropomorphism by design on emotions, but rather on this latter having inconsistent effects across studies and task contexts, whereas "attributional" anthropomorphism (i.e., psychological anthropomorphism) seems to show more consistent predictive value (e.g., Darling et al., 2015; Natarajan and Gombolay, 2020; Wu et al., 2025).

In this regard, the term "emotion" seems to denote various kinds of theoretical constructs in the literature. The few examples described above show that the concept of "emotion" can be intended as the attribution of emotions to other non-human entities, being either geometric shapes (as in the aforementioned experiment by Heider and Simmel, 1994), human-like automata such as Vaucanson's Flute Player [2], or human-like robots such as Kismet [5, 14]. This conceptualisation has prospered in part of the HRI literature (e.g., Marchesi et al., [15], Thellman, [16]. However, "emotions" can also be used to refer to the human-like emotions expressed by robots [17], or to emotion elicited in the human by the robot, i.e., the user's affective response [18]. Therefore, it seems that the term "emotion" does not admit of a unitary meaning in the current HRI literature, which creates conceptual confusion and further difficulties in describing, operationalising, and measuring this concept. Given this complexity and apparent heterogeneity, in the present review we decided to focus on both conceptual and terminological aspects of emotions, i.e., two aspects that (as highlighted throughout the manuscripts) represent two separate but linked features. In RQ0, we proposed an analytical taxonomy to classify the emotion-related phenomena, regardless of the terminological aspects (i.e., which labels were used to call "emotion" in the reviewed papers). In RQ1, we mapped the explicit definition of "emotion" in the reviewed papers, regardless of the specific focus of analysis (based on our proposed taxonomy) the researchers aimed to investigate.

Despite the various definitions and operationalisations of both concepts (emotions and anthropomorphism), current HRI literature provides a set of interdisciplinary frameworks integrating multiple fields (neuroscience, cognitive science, robotics, social psychology, and philosophy of mind) to connect them. For example, Nass and Moon [19] proposed the CASA (Computers Are Social Actors) paradigm as a socio-psychological framework arguing that humans mindlessly apply social and emotional scripts to computers and robots, via the automatic activation of social heuristics rather than conscious beliefs in machine minds. Subsequently, Epley and colleagues' [6] SEEK model explains, from a socio-cognitive perspective, why people anthropomorphise; crucially, it treats emotions such as loneliness and social disconnection as direct psychological drivers of anthropomorphising non-human agents, making emotion a cause rather than only a consequence of anthropomorphism. Lately, Giannopulu , [20] introduced the concept of "enrobotment", i.e., the idea that the creation of robots externalises users' internal mental/affective representations, and that their emotional and affective engagement with a robot reflects a projection of internalised mental states onto an external, technological artefact. Phillips and Visser, [21] recently proposed an integrative conceptual framework distinguishing robot human likeness, intended as a design property, from anthropomorphism, intended as a psychological inference process, and situating emotion attribution as one of the inferential outputs of the latter.

That said, the primary aim of this scoping review is to map how current HRI literature has conceptualised "emotion", and operationalised and investigated it in the context of the existing relationship with the concept of "anthropomorphism towards robots", which carries different meanings according to current literature. Note that,

unlike systematic reviews, we did not focus on a clearly predefined question nor hypothesise a direction of the effect; instead, we provided a narrative and exploratory overview of the results of HRI studies examining the relationship between emotions and anthropomorphism towards robots. We did so by considering empirical studies involving interactions between humans and robots. By narrowing our focus to this domain, we concentrated on robots as non-humans, technological artefacts to which the concept of "emotions" can be applied. By definition, a robot is a programmable, physical, automated machine that can perceive and interpret information from the physical environment and can act upon this environment [22]. Robots have three types of components: (i) a control system (e.g., the controller board), (ii) the sensors, which can read information from the surrounding environment, and (iii) the actuators, which produce an effect on the environment of the robot. To stick to this definition, we based our review only on technological artefacts that met this definition, and thus we excluded other entities such as computers and disembodied avatars.

To meet our aim, we proposed the following research questions as guidance for our scoping review:

RQ0 Is the term 'emotion' used to refer to a single theoretical construct in current HRI literature?

RQ1 What are the main theoretical frameworks/models of emotion and anthropomorphism towards robots? How have both emotions and anthropomorphism been defined in the current literature?

RQ2 Which methods have been used to explore the relationship between emotion and anthropomorphism towards robots?

RQ3 What are the main findings regarding the relationship between emotion and anthropomorphism towards robots?

It is important to note that, while systematic reviews gather empirical evidence to address a specific research question [23], our review falls more within the category of scoping reviews, which are intended to map the existing literature on a given topic and summarise research methodologies and findings in a given field (in our case, HRI), without assessing the quality of studies or answering a specific, focused question. Since the purpose is to map the range and nature of current evidence, identify how concepts are defined and operationalised, it is not designed to estimate pooled effects or appraise the strength of evidence [24]. To provide a structured literature review, we have decided to incorporate in our scoping review both the recommendations of the PRISMA 2020 Statement guidelines [25], which served as a guideline for reporting specific aspects of the review, and the protocol developed by Thellman and colleagues [16], which was successfully applied to HRI reviews. This choice was meant to incorporate structured methods into a descriptive review and ensure the replicability of our rigorous and step-by-step methodology. Furthermore, we would like to highlight here that we have not decided a priori criteria to identify our emotion-related categories (RQ0) or map the terminology that HRI researchers used to refer to our constructs of interest (emotion and anthropomorphism; RQ1). Rather, all the findings and observations reported below genuinely came out from our reflections, and thus cannot fully encompass complexity and heterogeneity of the phenomena we examined – especially because they would need further support from statistical data giving a precise estimate of the effects.

Methods

The protocol we followed includes the following steps: (i) defining the theoretical background as well as the rationale of the review; (ii) listing the research questions that the review aims to address; (iii) clearly describing the literature research strategy, including query strings and search systems; (iv) defining the study selection

criteria and procedures; and (v) formulating the data extraction strategy, while specifying how each piece of information required to address our research questions has been obtained. While steps (i) and (ii) have already been described in the Introduction section, below we detail the three remaining steps.

Data source and research strategy

First and foremost, we primarily identified the academic research systems that are suitable for a literature review. Following the recent work of Gusenbauer and Haddaway [26], we decided that Google Scholar is inappropriate as a search system, as its use may negatively impact research validity; for example, it is not reproducible in reporting identical results for repeated identical queries [26]. Thus, we used other academic search systems, which proved to be appropriate for reviews because they are bibliographic databases with a curated catalogue of information. During the time that this review was conducted (July 2025/February 2026), all these search systems indexed publications from the research areas that the authors identified as relevant: Psychology, Social Sciences, Computer Sciences, Neuroscience, Engineering, and Human-Robot Interaction.

For all the selected search systems, a query string was used; it included the relevant terms associated with emotion, anthropomorphism, and robot, and their various permutations, to address the research questions of the present work. It resulted in "emot*", "anthropomorph*", and "robot*", which were combined into a functional string using the AND operator. We point out here that our choice to use only the term "anthropomorph*" instead of including other relevant and similar constructs such as "agency", "mind attribution", "theory of mind" and "intentional stance" was motivated as follows. From a conceptual perspective, there is still confusion in HRI literature about whether considering "anthropomorphism" as conceptually independent from other closely related constructs, such as "agency", "mind attribution", "theory of mind" and "intentional stance". On the one hand, some authors suggest that the term "anthropomorphism" can be used as a construct incorporating the other constructs [6, 27–29]. On the other hand, there is also opposing evidence suggesting that the construct of anthropomorphism differs from others, such as "intentional stance" [30] and "theory of mind" [31]. Since we were not able to disentangle the degree to which anthropomorphism can encompass these related constructs, we put aside all the conceptual discussion around this term (which goes beyond the scope of this review) and simply adopt "anthropomorphism" as a generic search term which can potentially comprise other relevant terms/constructs.

The search was conducted in July 2025 and resulted in several lists, including 55 (PubMed), 296 (ScienceDirect), 77 (IEEE Xplore), 24 (ACM Digital Library), 207 (Wiley Online Library), and 304 (Scopus) publications, for a total of $N = 963$ publications.

Study selection

For the study selection, we followed the review protocol that Thellman and colleagues [16] have successfully developed for HRI reviews.

The 963 publications were identified based on the following inclusion and exclusion criteria. The inclusion criteria were: (i) papers reporting empirical studies; (ii) written in English; (iii) published in peer-reviewed journals and proceedings; (iv) publication date (2000–2025), where it was possible to select this option in the search system; (v) open-access publication. The exclusion criteria were: i) publications not subjected to a peer-reviewed process; (ii) non-empirical papers (e.g., other systematic reviews and/or meta-analyses); (iii) publications in languages other than English; (iv) unpublished papers (e.g., preprints stored in repositories such as ArXiv and OSF). All publications obtained following these inclusion and exclusion criteria ($N = 963$) were examined for their relevance to the research questions we aimed to address following a three-step study selection as conceived by Thellman and colleagues [16] for reviews conducted in the HRI domain (see Fig. 1 for a summary of the study selection process).

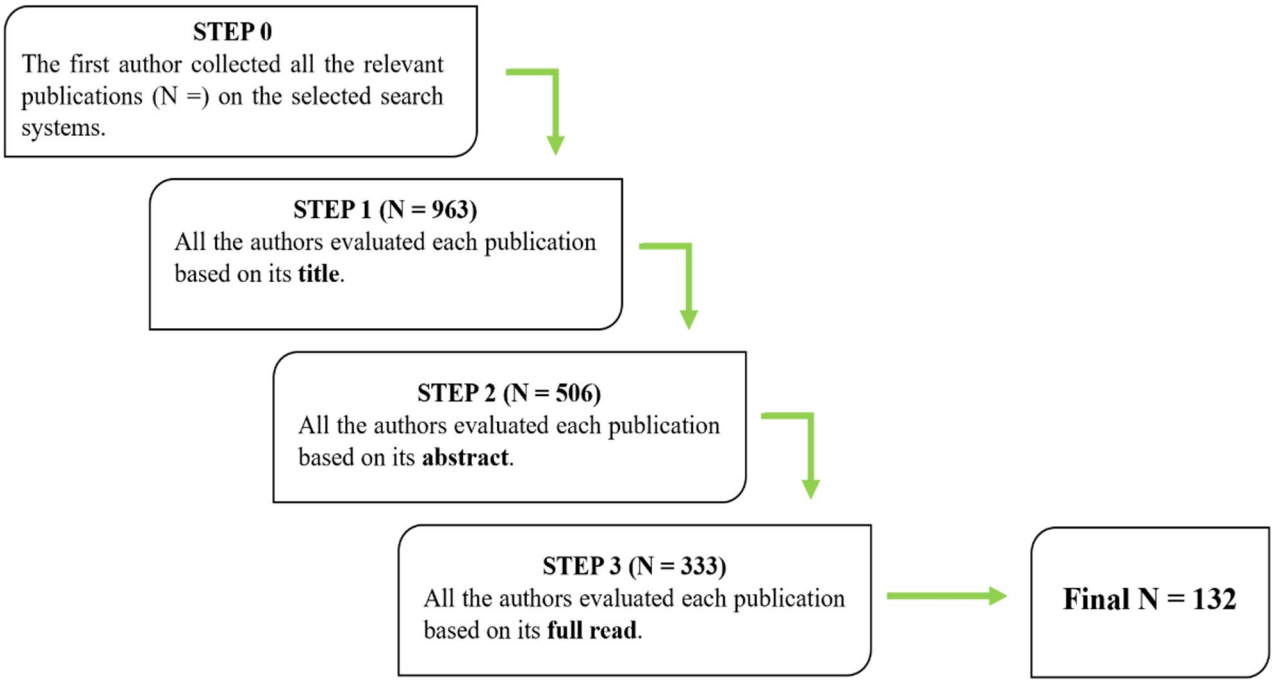


Fig. 1 Schematic representation of the search selection process

Step 1

All the authors (four in total) independently judged each publication (N = 963) as relevant or irrelevant based on its title. In other words, all authors independently judged whether the title suggested that the paper addressed the topic of interest (emotion and anthropomorphism in HRI). Publication titles that were perceived as difficult to judge were conservatively marked as relevant. All publications judged as relevant by at least two authors out of four were kept; the rest were discarded.

Step 2

All the authors independently judged each of the remaining publications (N = 506) as relevant or irrelevant based on their abstracts. In other words, as for Step 1, all the authors independently judged whether the abstract suggested that the paper addressed the topic of interest (emotion and anthropomorphism in HRI). Abstracts that were perceived as difficult to judge were conservatively marked as relevant. All publications judged as relevant by at least two authors were kept; the rest were discarded.

Step 3

All the authors independently judged each of the remaining publications (N = 333) as relevant or irrelevant based on a full read-through. As for the previous steps, all the authors independently judged whether the full paper was relevant or irrelevant with reference to the three relevant terms associated with emotion, anthropomorphism, and robots. The publications judged as relevant by at least two authors were kept; this resulted in N = 152 publications. After removing duplicates, i.e., publications that appeared identically in two or more search systems, the final number of publications was N = 132 (see the corresponding OSF repository at the following link: https://me-l.co/oluak9y1 for a summary of the included publications, divided across the three steps of selection).

Note that four raters independently classified each of the 132 papers as "relevant" or "irrelevant", following Thellman and colleagues' [16] procedure. Inter-rater variability was assessed with Fleiss' kappa [32], Light's kappa [33], and the mean of all pairwise Cohen's kappa, together with the percentage of papers on which all four raters agreed. 95% CIs were obtained by nonparametric bootstrap over items (2000 resamples). Analyses were run in RStudio v.4.3.3 (RStudio [34]), using the *irr* package [35]. Agreement was interpreted using Landis and Koch's framework (1977) for the measurement of observers' agreement. Results showed that raters classified 67% of items (papers) as relevant. Inter-rater agreement was moderate (Fleiss' $\kappa = 0.51$, 95% CI [0.42, 0.60]; Light's $\kappa = 0.52$, 95% CI [0.43, 0.61]; mean pairwise agreement = 84%). Results indicate that the relevance judgments may be reliable as a measure of shared constructs among the raters, with moderate/poor differences in how the raters interpreted them.

Data extraction

Through the data extraction process, we collected all the necessary information to address the research questions as listed in the Introduction section (RQ0-RQ3). Data were extracted from each of the publications included in the review ($N = 132$). The categories of the extracted data were inspired by those used by Thellman and colleagues [16] and adapted to the specific scopes of this review (a full table with a summary of each category is also reported in Table 1).

Variables of interest

The first step consisted of listing how the two variables of interest (emotion and anthropomorphism towards robots) have been defined and operationalised in each of the included publications.

Participants

Information regarding participants was collected with reference to (i) sample size, to check for robustness of results, (ii) age and sex/gender, to see whether people's attribution of emotion to robots changed as a function of age, sex/gender, or both. We followed the basic distinction between children and adults (for age) and males, females and others (for sex/gender).

Study setting

This category refers to where each study was conducted (Lab/Field/Online/Mixed/Not applicable). Studies categorised under the "Lab" label took place in a controlled environment (e.g., a laboratory at the university) and under the supervision of one or more experimenters; studies that were carried out in the field (e.g., in the street or in another public space) were labelled as "Field study". Studies conducted by recruiting participants through an online platform and thus did not foresee their physical presence in the experimental setting were labelled as "Online studies". Finally, "Mixed" refers to a mixed study setting, e.g., a study may comprise experiments conducted online and others in a controlled laboratory setting, whereas "Not applicable" (NA) stands for studies in which the authors did not report any information regarding the study setting.

Robot characteristics: presentation

We collected three types of information related to robot characteristics. The first is *robot presentation*, i.e., how the robot has been presented to participants in each study (Text/Image/Video/Physical/Mixed/Not applicable). "Text" refers to a verbal description of the robot provided to participants, e.g., through vignettes; "Image" refers to a picture of the robot; "Video" refers to a robot presented through video; finally, "Physical" refers to a robot

Table 1 Categories and subcategories of data extracted from each study included in the review ($N = 132$). They were both inspired by the categorisation introduced and adopted by Thellman and colleagues [16]

Category	Subcategories
Definitions	[Emotion] [Anthropomorphism]
Participants	[Sample size] [Sex] [Age]
Study setting	[Lab] [Field] [Online] [Mixed] [Not applicable]
Robot Characteristics: Presentation	[Text] [Image] [Video] [Physical] [Mixed] [Not applicable]
Robot Characteristics: Morphology	[Anthropomorphic] [Zoomorphic] [Functional] [Mixed] [Not applicable]
Robot Characteristics: Behaviour	[Interaction with participants] [Interaction with others (non-participants)] [Non-social] [No behaviour] [Mixed] [Not applicable]
Measure: emotion	[Verbal] [Behavioural] [Neurological] [Mixed] [Not Applicable]
Measure: anthropomorphism	[Verbal] [Behavioural] [Neurological] [Mixed] [Not Applicable]
Findings	[Yes] [No]

being physically present in the experimental setting. "Mixed" refers to studies combining various presentation modalities (e.g., the robot was presented first as a static image and then as a dynamic stimulus in a video), whereas "Not applicable" (NA) refers to studies in which the authors did not report any information regarding the robot's presentation modality.

Robot presentation: morphology

The second information related to the robot's characteristics relates to its morphology, i.e., the shape of the robot presented in each study (Anthropomorphic/Zoomorphic/Functional/Mixed/Not applicable). "Anthropomorphic" refers to a robot resembling a human-like shape; "Zoomorphic" refers to a robot having an animal-like shape; "Functional" refers to a robot that does not have an anthropomorphic or zoomorphic shape. Finally, "Mixed" refers to studies involving various robot morphologies, e.g., the participants were presented with both anthropomorphic and functional robots. "Not applicable" was used if the authors did not provide any information regarding the shape of the robot they used in their study.

Robot presentation: behaviour

The third characteristic related to robots is the *robot behaviour* (Interaction with participants/Interaction with others (non-participants)/Non-social/No behaviour/Mixed/Not applicable). "Interaction with participants" means that participants interacted with the robot; "Interactions with others (non-participants)" means that participants did not directly interact with the robot, but they observed an interaction between a robot and another human (e.g., an experimenter); "Non-social" means that the study did not include an interaction between the robot and a human (neither the participant nor another person); and finally, "No behaviour" means that the robot did not display any behaviour (for example, when the robot was presented in the form of pictures). "Mixed" stands for studies combining various modalities of robot behaviour, for example, the participants first observed the robot interacting with others (non-participants), and then they directly interacted with it; "Not applicable" stands for studies which did not report any information related to the behaviour of the robot.

Measure

This category comprises the measurement of the two variables of interest, i.e., emotion and anthropomorphism. In other words, it refers to the type of data collected in each study included in this review, related to both participants' emotions and their likelihood to anthropomorphise robots (anthropomorphism) (Verbal/Behavioural/Neurological/Mixed/Not applicable). "Verbal" refers to participants' subjective reports, quantitative methods such as ratings on Likert scales, as well as qualitative methods. "Behaviour" refers to non-verbal data only, e.g., data collected through quantitative methods (e.g., reaction times, accuracy ratings. "Neurological" refers to data obtained by electrophysiological and functional techniques (e.g., electroencephalography, functional magnetic resonance, etc.). Finally, "Mixed" refers to combined approaches of measurement, for example, participants completed questionnaires (which go under the "Verbal" label) and then a behavioural task; "Not applicable" stands for studies in which one or both variables were not measured, or the type of measurement adopted was not clearly specified.

Findings

This category refers to the findings of the study (yes/no). "Yes" refers to the emergence of a relationship between emotion and anthropomorphism towards robots; conversely, "no" refers to a lack of relationship between the two variables of interest. Note that, throughout the review, we considered various kinds of relationships between emotions and anthropomorphism, recalling the conceptual issues that we discussed in the Introduction.

Overview of the included publications

Of the 132 publications included in the review, the majority (59.39%, namely, 79 publications) reported studies conducted in a laboratory setting, i.e., a controlled environment in which participants were under the supervision of one or more experimenters. About a quarter of the publications (24.81%, namely 33 publications) reported studies conducted online, i.e., they comprised controlled experiments leveraging digital platforms (for example, Amazon Mechanical Turk or Prolific) to reach a wide audience. A small amount of the shortlisted publications (9.02%, namely, 12 publications) reported studies conducted in the field, i.e., conducted in natural and/or real-world settings rather than a controlled laboratory environment. Finally, the remaining publications (6.01%, namely, 8 publications) comprised a mixed study setting, e.g., combining laboratory and online settings (e.g., [36, 37]; Yamaguchi et al., [36]) or laboratory and field settings (e.g., Yam et al., [38].

This information is summarised below in Fig. 2.

Moreover, Fig. 3 highlights that, in the selected period (2000–2025), there was a significant increase in published studies investigating the topic of emotions and their relationship with anthropomorphising robots

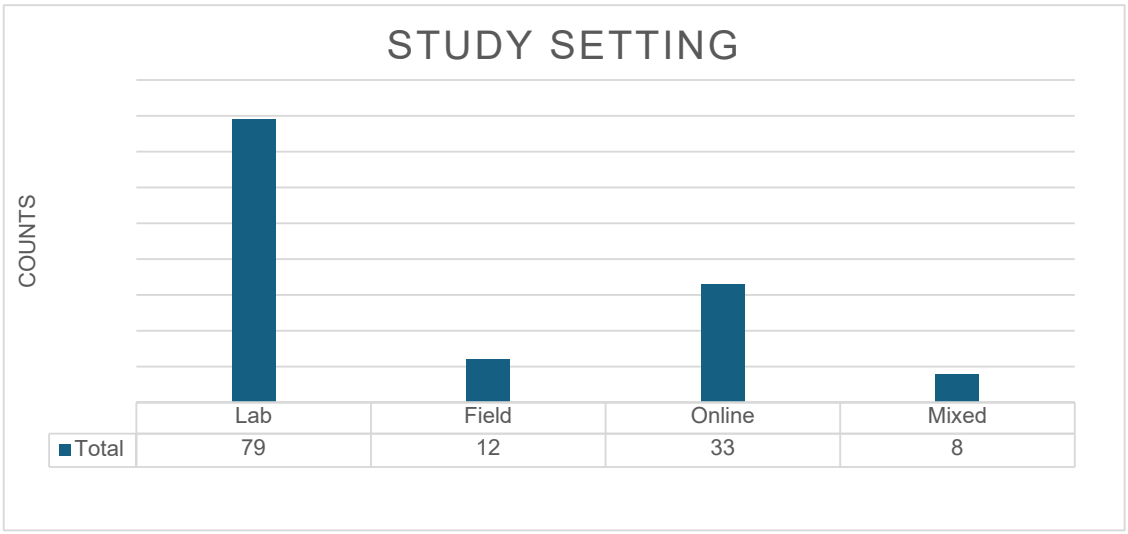


Fig. 2 Graphical representation of the shortlisted publications based on the study settings. Note that "Mixed" stands for combined study settings within a single study (e.g., laboratory + online experiments)

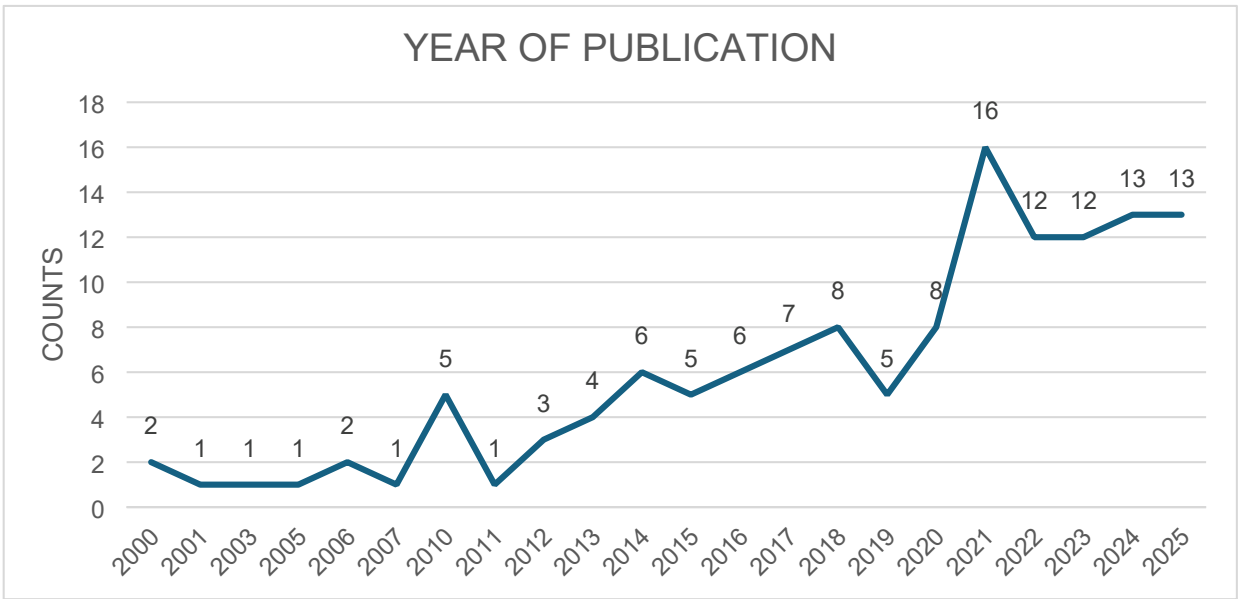


Fig. 3 Schematic representation of the shortlisted publications, plotted separately for each year of the selected period (2000–2025)

(especially from 2021 onwards). It is relevant data, potentially suggesting an increase in interest in these topics, given the increasingly massive presence of robots (think of Alexa and Google Home) in human social spaces.

Regarding participants, the majority of publications (88.72) reported studies involving neurotypical adults, except for one, which involved individuals diagnosed with schizophrenia [39]. 10 studies involved neurotypical children ([40–48] ; [49]), whereas 3 studies involved samples composed of elderly people [50–52].

Literature review

To address the aim of this review, i.e., to understand whether there is a relationship (and, if so, of what sort) between emotions and people's anthropomorphism towards robots, thus addressing RQ0 – RQ3, we analysed the publications included in the review. The next question presents the results of this analysis.

RQ0: is the term 'emotion' used to refer to a single theoretical construct in current HRI literature?

The first important point that we identified is that, in the last two decades, the term "emotion" has been used heterogeneously. What we have noticed is that this term seems to comprise various foci, which do not represent equivalent manifestations of a single construct but rather different levels of analysis. Notably, all these levels of analysis refer to the Human, to the Robot, or to the Interaction, i.e., the three fundamental aspects that are investigated in the Human-Robot Interaction domain.

The first level is embedded in the human ("H") factor. Humans' interaction with robots is shaped by their background conceptions and ideas about them, which in turn modulate behavioural expectations. Thinking of anthropomorphism, for example, familiarity with sci-fi narratives may induce people to think that robots can have a mind of their own, or determine an automatic tendency to interpret their behaviour as caused by ascribed mental states. Therefore, from the perspective of the human interacting with the robot, the term "emotion" can refer to *people's attribution of emotions to robots*, as also highlighted in the Introduction [5]. But reference to 'emotion' may also refer to people's ability to correctly and accurately identify and *recognise emotional expression when displayed by a robot*. Although similar, there is a core difference between the two levels. Whereas emotion recognition can be defined as the ability to identify and decode emotional signals from an agent's behaviour, i.e., via facial expressions, body language, or the tone of voice (Demenescu et al., 2012; [53]), emotional attribution is an inferential process by which people assign an emotional state to another agent [54, 55].

Another level that emerged from the analysis is related to the robot ("R") factor. It concerns the characteristics of the robot (e.g., in terms of behaviour and/or appearance) and the way they promote successful interactions between humans and robots. In the context of anthropomorphism towards robots, it means that, if a robot exhibits a human-like morphological shape (e.g., it has eyes and/or its body resembles a human-like shape), it may automatically elicit people's anthropomorphism. Therefore, when considering the R factor, the expression "emotion" translates into the level of the *robot's emotional design* (i.e., emotion expression and generation). In other words, it focuses on robots to explore how they can simulate and exhibit emotions to appear (for example) more likeable, trustworthy, or intelligent.

Last but not least, one cannot neglect the importance of the "I" factor, i.e., the interaction itself. In the context of anthropomorphism towards robots, if an interaction promotes social bonding between humans and robots through the opportunity to share the same physical and social space, the human may have the impression that the robot is "like him/her" and thus be more likely to ascribe anthropomorphic traits to it. Thus, when considering the I factor, the expression "emotion" focuses on humans' emotional reactions to the robot's presence and/or behaviour. Therefore, the fourth level of analysis is *emotional reaction*. A summary of the four levels of analysis across all studies is reported in Fig. 4 below.

As Fig. 4 above shows, more than one-fourth of the studies used the expression "emotions with reference to the first level of analysis, i.e., people attributing emotions to robots (37/132 studies; 28.03%). The same percentage emerged for studies focusing on the level of the robot's emotional design, i.e., how the robot can simulate and exhibit emotions to promote a natural interaction between humans and robots. In short order, studies are also focusing on the other "H" factor-level of analysis, i.e., people's ability to recognise emotions when displayed by a robot (29/132 studies; 21.97%); the same percentage for studies focusing on emotional reactions to robot appearance and/or behaviour.

In sum, the mapping of the selected publications highlighted that the expression "emotions" comprises (at least) four primary levels of analysis, which do not correspond to equivalent categories or meanings. To valorise

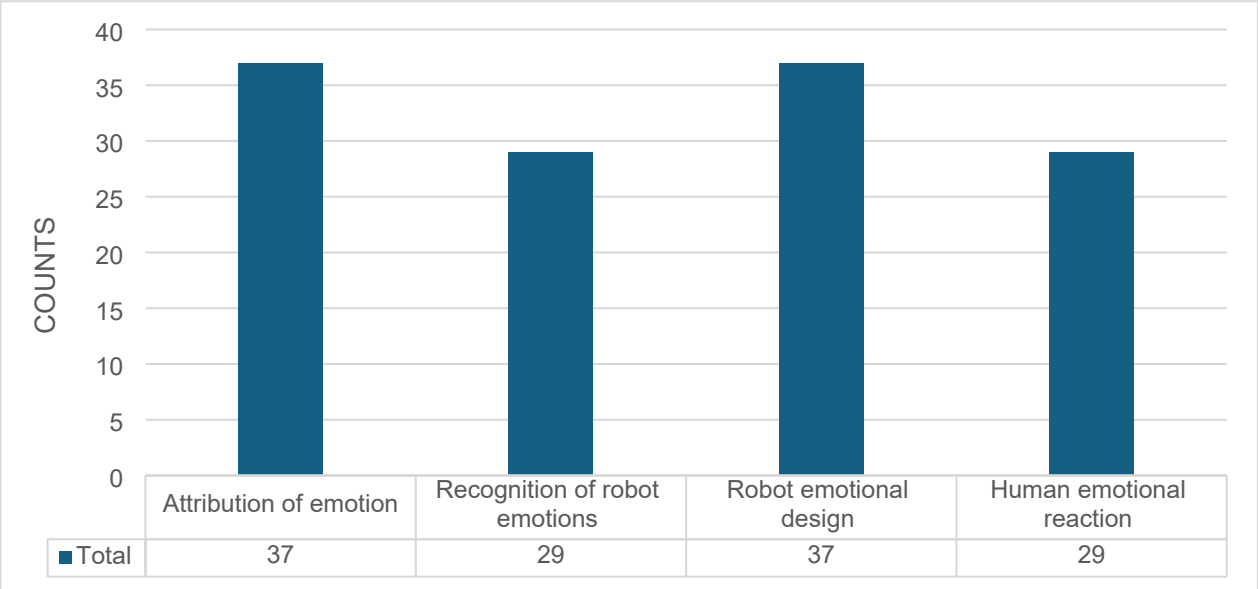


Fig. 4 Graphical representation of the shortlisted publications based on the four categories, i.e., the various meanings assumed by the expression "emotions" in the HRI domain

Table 2 Schematic summary of the most frequent definitions of emotion in the reviewed papers. Note that the category "type" refers to either a generic definition of emotion provided by the authors ("generic" label) or to a single, specific emotion investigated in a given study ("specific" label)

Definition	Type	References
Experience: the ability to feel emotions and bodily sensations.	General	Chi et al., [36], Koban, [67], Swiderska, [68], Yam, [38]
Intelligence: the ability to perceive and understand different emotions.	General	Grundke et al., 2023; [44]
Affective communication: expression of emotions through speech.	General	Baza et al., 2022; Heger et al., 2006; [69, 70]; Thorstenson and Tian, 2025
Affective impressions/responses (specifically, liking): sensory reactions when encountering another agent, e.g., a robot.	General	Angel-Fernandez and Bonarini, [71], Koch, [43], Letheren, [72], Li, [73]; Yamaguchi et al., [36]
Multidimensional construct: self-reported states and brain activity in regions like the limbic system.	General	Rosenthal-Von der [74]; Wei and Zhao, 2016; [75]
Human-like states: internal states, both physiological and psychological, that are uniquely human.	General	Becker et al., 2023; Breazeal, 2001; [31]; Taminami et al., 2020; Trieu and Nguyen, 2025; [76]
Empathy: affective emotional response.	Specific	Augello, [40], Mattiassi, [77]; Tsumura and Yamada, 2023
Humour: an affective skill, also known as a type of human-like emotional response.	Specific	Xie et al., [78], Yang, [79]

Table 3 Schematic summary of the most frequent definitions of emotion in the reviewed papers

Definition	References
Humanness: the quality or condition of being human, or having characteristics typically human.	Chang et al., 2022; [44, 73, 80, 89]; Waytz and Norton, 2014; [79]
Mind perception/attribution: attribution of mental states such as beliefs, intentions, thoughts, and emotions to others.	Grundke et al., 2023; [90]
Definition taken from Epley and colleagues' [6] work: tendency of imbuing the real or imagined behaviour of nonhuman agents with human-like characteristics, motivations, intentions, or emotions.	Crowell et al., [91]; Eyssel et al., 2010; Eyssel and Kuchenbrandt, 2012; [92]; Mara and Appel, 2015; [38];

the conceptual heterogeneity that emerged when assessing how researchers addressed "emotions" in the HRI domain, we decided to apply these four categories of interest to the main findings of this review (RQ3). Notably, the potential link with anthropomorphism towards robots will be treated separately for each of the four levels.

RQ1: what are the main theoretical frameworks/models of emotion and anthropomorphism towards robots? How have both emotion and anthropomorphism been defined in the current literature?

This section collects the main theoretical frameworks/models of the two variables of interest (i.e., emotions and anthropomorphism towards robots), and the definitions of both derived from the empirical findings of the short-listed studies.

Emotions: models

Some recurring models can be identified across the 132 studies included in the review.

Circumplex model of affect (CMA) [56] According to Russell, emotions are not distinct and isolated from each other, but exist in a continuous space and can be mapped using two primary axes: (i) valence, i.e., the degree to which an emotion is pleasant (as in the case of joy and pleasure) or not (as in the case of sadness and fear); and (ii) arousal, i.e., the level of physiological activation associated with a specific emotion. For example, excitement and tension foresee a high level of arousal, whereas low arousal is typical of boredom or relaxation.

Pleasure-Arousal-Dominance (PAD) model [57] This model describes emotions as points within a three-dimensional space: (i) pleasure, i.e., the valence or degree of "positivity" of an emotion; (ii) arousal, i.e., the level of physical and/or mental alertness, and (iii) dominance, i.e., the degree of control that a person experiences. By combining these three dimensions, it is possible to define specific emotional states; for example, anger is characterised by low pleasure (it is an unpleasant emotion) combined with high arousal (high activation) and high dominance (feeling powerful/assertive).

Plutchik's wheel of emotion [58] Plutchik identified eight primary bipolar emotions, arranged in opposing pairs (joy-sadness; trust-disgust; fear-anger; surprise- anticipation). These represent the "primary" feelings that form the basis of all other complex emotional experiences, thus having secondary and tertiary emotions in a wheel-shaped mechanism. In the wheel, emotions are organised following three main features: (i) intensity, shown through the vertical dimension – emotions at the centre are more intense, whereas those at the edges are the mildest; (ii) similarity, i.e., emotions that are close to each other on the wheel are more similar; and (iii) polarity, i.e., opposite emotions sit directly across the wheel.

Model of the six basic emotions [59] According to this model, humans across all cultures share six universal and innate emotions (happiness, sadness, anger, fear, disgust, and surprise). Notably, they correspond to specific, hardwired facial muscle configurations. Therefore, they developed a system called the Facial Action Coding System (FACS) [60], which represents a comprehensive, automatic catalogue for describing observable facial movements. The rationale behind this was that facial expression breaks down into single Action Units (AUs), which correspond to specific groups of muscles. Once FACS was established, the authors identified which specific combinations of AUs consistently appeared when one of the six basic emotions appeared; this gave birth to the "Emotion FACS" (EMFACS), which served as a dictionary that mapped individuals' facial expressions into the six basic emotions [61].

Dehumanisation model of emotions Deriving from the intergroup socio-cognitive literature [62], it revolves around the idea that there are cases in which people view outgroup members as less humans by denying them the spectrum of human emotions. According to the work of Leyen and colleagues (2000; 2001), the core of the model is the distinction between primary and secondary emotions. The former refers to emotions (such as anger, fear, and joy) shared between animals and humans, which appeared early in life/evolution and are simple, reac-

tive and physiological. Conversely, the latter refers to emotions (such as melancholy, shame, and hope) which are exclusively human; thus, they develop at a later stage, are culturally situated, and are more cognitive and moral.

Biopsychological model of mood Thayer [63] proposed a model aiming to overcome the "happy vs. sad" dichotomy, arguing that humans' mood is primarily driven by two independent dimensions of activation (or arousal). In other words, the model identifies two distinct types of activation that function on separate continua: (i) the energetic arousal, which is strictly tied to vitality, and (ii) the tense arousal, which is tied to humans' "fight or flight" mechanism and thus the perception of danger and pressure.

Appraisal theory There are two major influential models. One is Scherer's Component Process Model (CPM), which considers emotions the result of a dynamic process consisting of five interrelated subsystems: (i) appraisal, i.e., evaluation of objects and events; (ii) physiology, i.e., bodily changes such as heartbeat; (iii) motivational component, i.e., the preparation for action; (iv) the motor expression component, which involves facial expressions, vocal changes, and gestures; and finally (v) subjective feelings, i.e., the internal experience of feeling and labelling an emotion [64]. The core of the CPM is a sequence of Stimulus Evaluation Checks (SECs) that allow us to process information in a specific, functional order; being a continuous process, it can explain a vast spectrum of emotions rather than just basic categories like happiness or sadness.

The other relevant model is the Appraisal Model proposed by Moors and colleagues [65], according to which emotions are adaptive responses that reflect our appraisals of features from the external environment/internal events, which are significant for our well-being. In other words, emotions are elicited by the evaluations (appraisals) of how events and situations relate to our important goals, values, and concerns.

Emotion as a key to the theory of mind (ToM) A distinction between cognitive and affective ToM was introduced [66]: while the former refers to the understanding of what others *think* or *believe* and thus involves mechanisms such as perspective-taking and understanding false beliefs, the latter involves the understanding of what others *feel*, thus requiring empathy and the ability to distinguish one's emotional states from someone else's. In other words, we are not only able to recognise and attribute emotions to others; we can also infer *why* they are feeling a specific emotion based on their mental states and environmental information.

Emotions: definitions

The definitions of emotions outlined in the shortlisted publications were various and heterogeneous. To favour a better understanding for readers, we summarised the most recurring definitions in Table 2 below.

It is important to note that, firstly, this list comprises only the most frequent conceptualisations of emotions that can be found in the reviewed papers; thus, it is not expected to be fully exhaustive of all the definitions of emotions; this also does not imply that the levels of analysis of the "emotion" construct (as described under RQ0) are only those four that we have proposed here. Secondly, "emotions" are identified in the HRI literature not only with *attitudes*, i.e., mental dispositions matured through experience and impacting behavioural, verbal, and emotional reactions [80], but also with *adaptive responses* that reflect our appraisals of environmental events [81]. From a neurological perspective, they have also been described as *interrelated*, *synchronised changes* [82] in response to evaluations of internal or external stimuli. When considering studies that focused on a single emotion, the core of investigation could be extremely varied: from embarrassment, i.e., the anticipation of negative evaluation by others [83], to anxiety, i.e., the negative emotional reaction that can arise (for example) during robot usage [52], to the "wow effect", i.e., a heightened emotional experience going beyond the usual understanding, and characterised by surprise and delight [84].

Anthropomorphism: models

As for emotion, several definitions, models, and theories of anthropomorphism can be found in the reviewed papers. Note that some of the models have been shortly described in the Introduction already, as they represent relevant models in the current HRI literature on anthropomorphism towards robots.

Three-factor model of anthropomorphism (SEEK theory) This model, also known as SEEK Theory, was proposed by Epley and colleagues [6] to explain why and when humans attribute human-like mental states to non-human agents, including technological artefacts such as robots. In the authors' view, the ascription of mental states is a motivated process driven by three psychological determinants: (i) a cognitive factor that they called the "elicited agent knowledge", i.e., the accessibility and applicability of anthropocentric knowledge; and two motivational mechanisms, namely (ii) the effectance motivation and (iii) the sociality. With the former (effectance motivation), the authors refer to the human intrinsic motivation to understand the behaviour of other agents, whereas the latter (sociality) refers to the desire for social contact and affiliation. In sum, the model predicts that humans are more likely to anthropomorphise other agents when anthropocentric knowledge is accessible and applicable, when they are motivated to behave like social agents, and when lacking a sense of social connections to other humans [6].

Media equation theory The core principles of the Media Equation Theory, which are then translated into the Computers Are Social Actors (CASA) framework [85], posit that humans' interactions with technological artefacts such as computers, television, and new media are fundamentally natural. This theory stems from the observation, later supported by empirical evidence (e.g., Gambino et al., [86], Nass, [19], Xu, [87], that humans tend to respond to machines using the same social rules and scripts that they mindlessly apply to other people, even when they are fully aware that machines lack feelings and intentions.

Dual model of dehumanisation In the context of social psychology, Haslam extensively investigated the conditions under which humans perceived others as "less than human". To do so, he conceptualises a dual model of dehumanisation that distinguishes between two distinct ways of defining what it means to be human [88]. Specifically, he identified two categories of human attributes that, when denied to others, can result in two different forms of dehumanisation. The first category comprises the Uniquely Human (UH) characteristics, i.e., a set of attributes that distinguish humans from animals (for example, rationality, civility, and moral sensibility). These attributes are typically learned in early childhood, can be culturally transmitted, and involve high-order cognitive abilities. When denied, the result is the "animalistic dehumanisation": the other is seen as unintelligent, lacking self-control, and their behaviours are considered driven by instinct. The second category comprises the Human Nature (HN) characteristics, i.e., traits that are essentially human (for example, agency, cognitive flexibility, and interpersonal warmth). These features are seen as innate, natural, and shared by all humans regardless of their cultural background. When denied, the result is a "mechanistic dehumanisation": the other is seen as object-like, lacking soul or depth.

Anthropomorphism: definitions

Where present, the definitions of anthropomorphism were quite consistent across the shortlisted publications. As for emotion, we summarised the most recurrent definitions in Table 3 below.

Note that the first listed definition of anthropomorphism (humanness) encompasses various aspects, such as physical appearance (5), behaviour (29), and embodiment [80]. More specifically, in some publications, anthropomorphism was declined as *anthropomorphic trustworthiness*, i.e., a general impression-based trust [93]; *social presence*, i.e., being in the company of a social entity [52]; or more implicitly, i.e., by underlining the distinction between superficial (appearance) and essential (mind, cognition, and emotion) human-like characteristics [94].

RQ2: which methods have been used to explore the relationship between emotion and anthropomorphism towards robots?

Our second research question relates to the methods used in the reviewed publications to explore the relationship between emotions and anthropomorphism towards robots. "Methods" included both stimuli and measures, as reported in Table 1. In particular, "stimuli" refers to the robot characteristics that we divided into three categories (presentation, morphology, and behaviours), whereas "measures" refers to how emotion and anthropomorphism have been measured, and thus the data type. To facilitate readers in following the flow of arguments, we will detail stimuli and measures separately in the next subsections.

RQ2a: what types of stimuli have been used to explore the relationship between emotion and anthropomorphism towards robots?

As mentioned above, "stimuli" refers to the characteristics of the robot used in the shortlisted publications of this review. Following Thellman et al.'s methodology (2022), they were divided into three subcategories: presentation, morphology, and behaviour.

Robot presentation

The most frequent modality involved the presentation of a physical robot, i.e., a robot sharing the same physical experimental space as participants (51/132 studies; 38.63%). The second most frequent modality was the presentation of a video of the robot (33/132 studies; 25%), immediately followed by the robot presented in the form of a static image (26/132 studies; 19.69%). A mixed presentation was also selected (20/132 studies; 15.03%), for example, participants were exposed to the robot both in the form of a static image and a dynamic video stimulus (as in [68]), or both in the form of video recordings and then in its embodied, physical version (as in [95]). The less frequent modality is text, i.e., when the robot was presented using a verbal description (for example, sentences or full stories) provided to participants (2/132 studies; 1.51%). A summary of robot presentation across all the studies is reported in Fig. 5

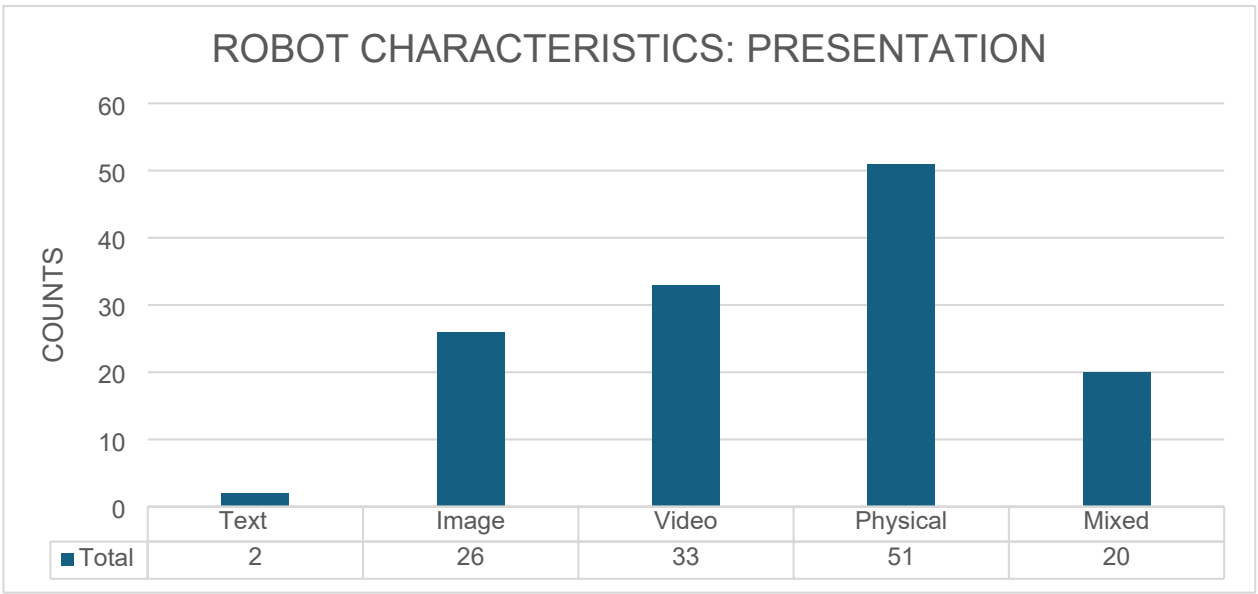


Fig. 5 Publications included in the review, sorted according to the subcategories of robot presentation

Robot morphology

The majority of studies (74/132 studies; 56.06%) involved robots with anthropomorphic features. This category is followed by studies employing robots with different morphologies (35/132 studies; 26.51%), e.g., anthropomorphic and functional features (as in the case of [96]). Only 17 studies out of 132 (12.87%) used a robot having functional features, i.e., a mechanical shape; 3 studies employed a robot clearly described as zoomorphic (2.27%). Finally, in 3 studies (2.27%), it was not possible to collect information regarding the robot morphology because the authors did not provide details about it. A summary of robot morphology across all studies is reported in Fig. 6

Robot behaviour

In almost half of the studies, the authors explicitly stated that, during the experiment(s), the robot directly interacted with the participants, for example, through voice and facial expressions (as in the case of [97]), or via haptic contact (as in the case of Sato et al., 2020). A large number of the remaining studies used robots displaying no behaviour, for example, because the robot was presented in the form of static images (35/132 studies, 26.51%). Only 15 studies foresaw the robot interacting with humans, who, however, were not the participants (15/132 studies; 11.36%); a similar number of studies (18/132 studies; 13.63%) used a robot programmed to display some behaviours, although they were not social. Finally, 9 studies (6.81%) used robots displaying various kinds of behaviours, for example, interacting both with participants and non-participants (as in the case of Li et al., 2020). A summary of robot behaviour across all studies is reported in Fig. 7

RQ2b: what types of measures have been used to explore the relationship between emotion and anthropomorphism towards robots?

As mentioned above, "measures" refer to how the two variables of interest, namely emotion and anthropomorphism towards robots, have been measured in the shortlisted publications of this review; and thus, following Thellman et al.'s methodology (2022), what type of data was collected.

Measure: emotion

In 79.5% of the reviewed studies (105/132 studies), the author(s) employed verbal measures to collect information from participants regarding their emotions. In most cases, they administered some items to participants (Likert-like or open questions), for example, asking them to rate the perceived robot's emotional expression in terms of intensity or identify the emotion displayed by the robot. However, in some cases, the authors used validated

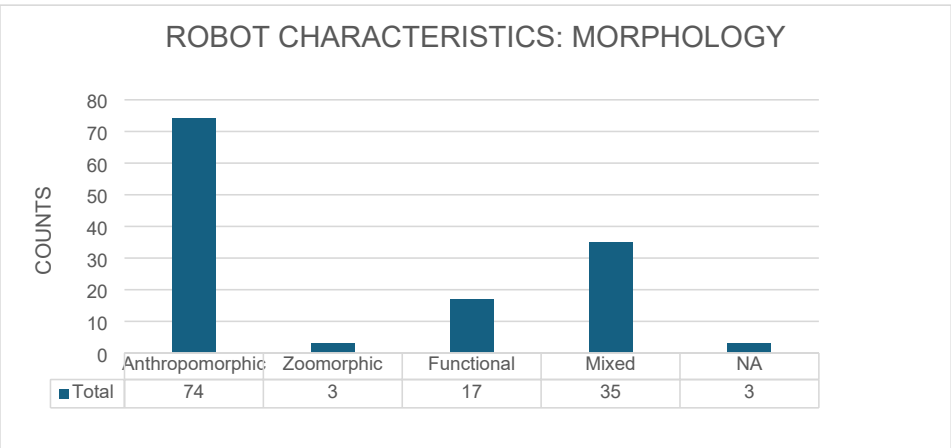


Fig. 6 Publications included in the review, sorted according to the subcategories of robot morphology

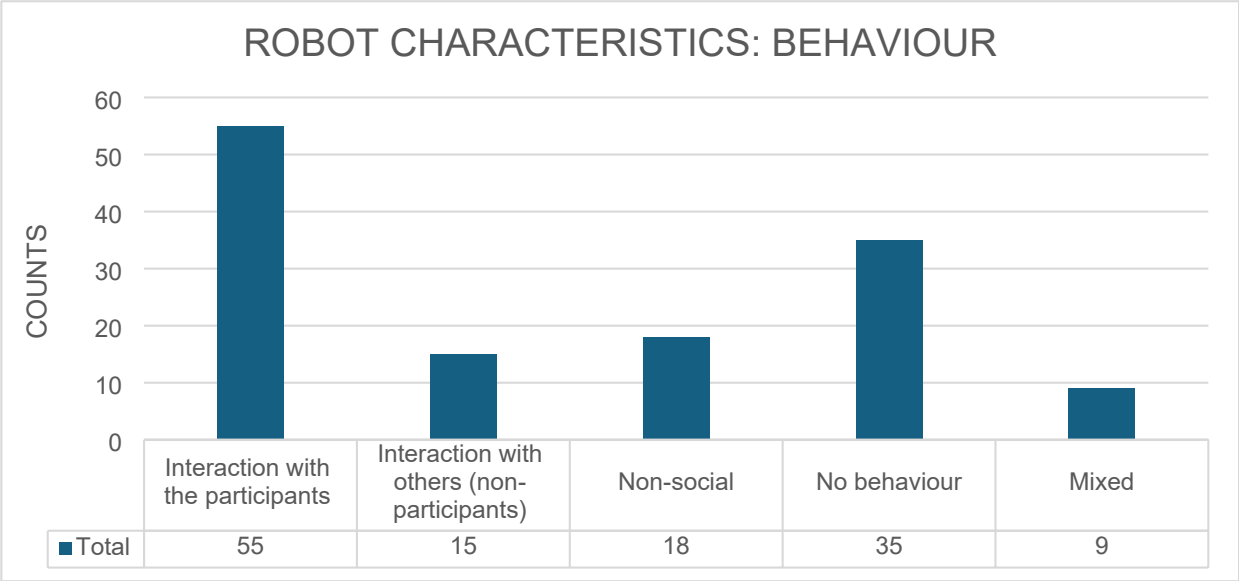


Fig. 7 Publications included in the review, sorted according to the subcategories of robot behaviour

instruments, such as the Self-Assessment Manikin (as in the case of [70]), the Mean Opinion Score [98] or the Pleasure-Arousal-Dominance (PAD) Emotional Scale (Xie and Liu, 2023). In other cases, the authors used subscales of questionnaires specifically targeting a certain emotional dimension, such as the Emotional Interdependence subscale, i.e., a scale taken from the Networked Minds Social Presence Inventory (as in the case of [99]), or some subscales of the larger Mind Perception Scale [92]. Only 6 studies (4.54%) employed non-verbal, behavioural measures, for example, reaction times and accuracy rates when participants had to discriminate whether the robot displayed a neutral or emotional expression [100]; similarly, only 6 studies (4.54%) reported neurological measures, for example, fMRI scanning of brain regions (such as the ventromedial prefrontal cortex) associated with emotions (Rosenthal-Von der [74]). Finally, 9 studies (6.81%) reported a mixed methodology, e.g., combining both verbal and behavioural/neurological measures. Notably, 7 studies (5.3%) did not measure emotions and thus were classified as "NA". A summary of emotion measures across all studies is reported in Fig. 8

Measure

anthropomorphism. In 71.21% of the studies (94/132 studies), the authors measured anthropomorphism towards robots using verbal measures, such as subjective reports and questionnaires. In some cases, participants were asked to describe the robot using their own words, which the authors later analysed for themes like perceived realism and human-likeness (as in the case of [89]). In some other cases, the authors used validated instruments to capture the degree to which participants anthropomorphised robots. For example, the anthropomorphism scale of the RoSAS was used the study by [101]. Note that, in Spatola and Wudarczyk's work (2021a), the authors used RoSAS by explicitly stating that it assesses the perception of robots with regards to social characteristics (anthropomorphism), including warmth, competence, and discomfort dimensions. Another instrument that has been used is the Godspeed questionnaire, and specifically the anthropomorphism subscale (e.g., Kühnlenz and Kühnlenz, 2020; [95]). Of the remaining studies, 9 employed a diverse methodology: 5 collected behavioural data (3.78%), 2 collected neurological data (1.51%), and 2 combined mixed kinds of data (for example, verbal and behavioural, as in the case of [37]). Finally, the 22.72% (30/132 studies) were classified as "NA", because anthropomorphism was not directly measured – for example, because it was directly in the experiment by presenting robots with different degrees of anthropomorphic features. A summary of anthropomorphism measures across all studies is reported in Fig. 9

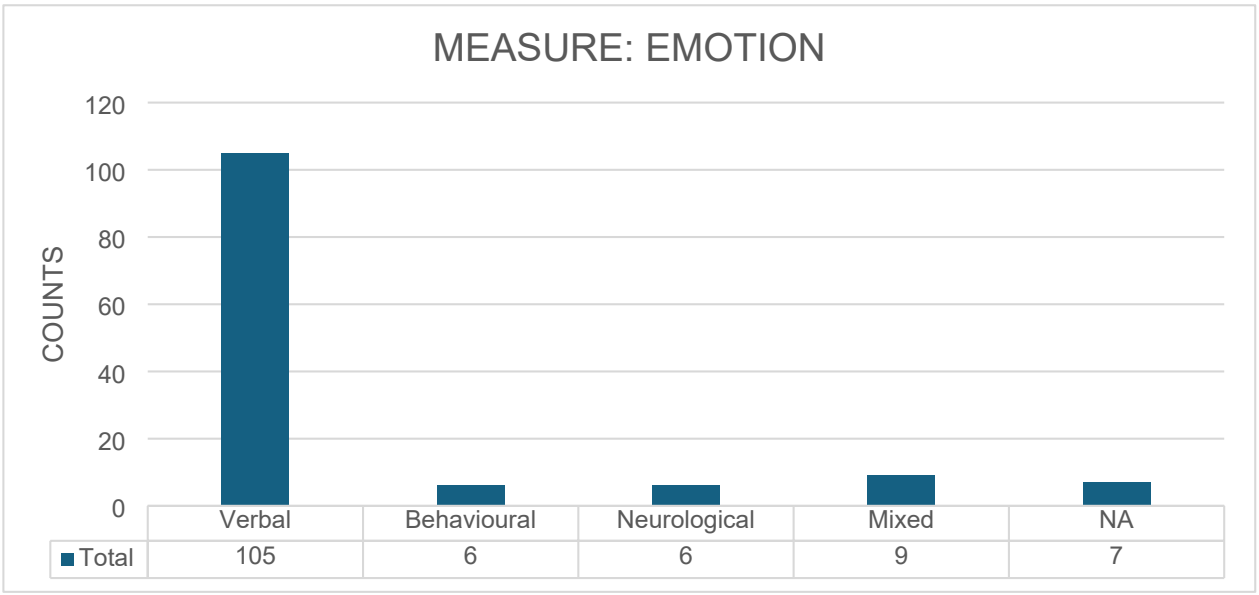
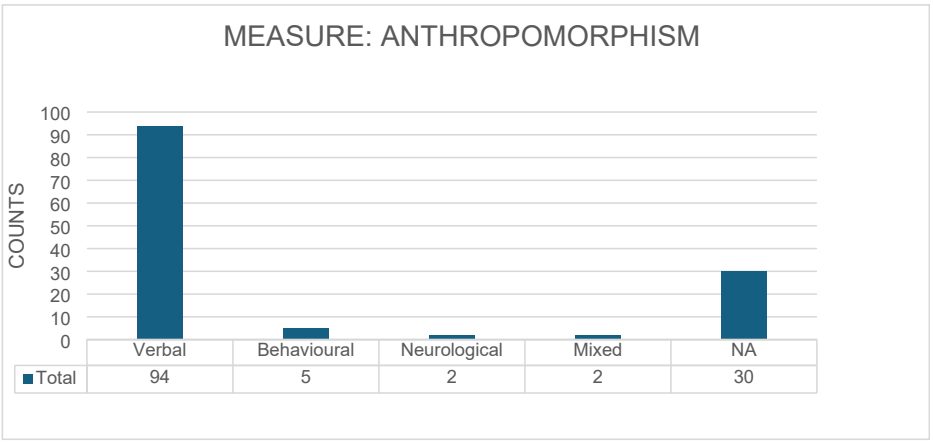


Fig. 8 Publications included in the review, sorted according to measures of emotion

Fig. 9 Publications included in the review, sorted according to the measures of anthropomorphism towards robots



RQ3: what are the main findings regarding the relationship between emotion and anthropomorphism towards robots?

In this section, we present research findings on the relationship between the two variables of interest, namely, emotions and anthropomorphism towards robots. As already shown in Table 1 we first divided the results of the shortlisted publications into two categories. The "Yes" category includes studies demonstrating an existing relationship between human emotions and anthropomorphism towards robots; conversely, the "No" category includes studies showing a lack of relationship between the two. Note that, unlike systematic reviews, we did not focus on a clearly predefined question nor hypothesise a direction of the effect; instead, we provided a narrative and exploratory overview of the results of HRI studies examining the relationship between emotions and anthropomorphism towards robots.

Overall, the majority of the reviewed studies (105/132 studies; 79.5%) reported the existence of a statistically significant relationship between human emotions and people's tendency to anthropomorphise robots. Conversely, only one-third of the reviewed studies (27/132 studies; 20.5%) reported a lack of relationship between the two (see Fig. 10).

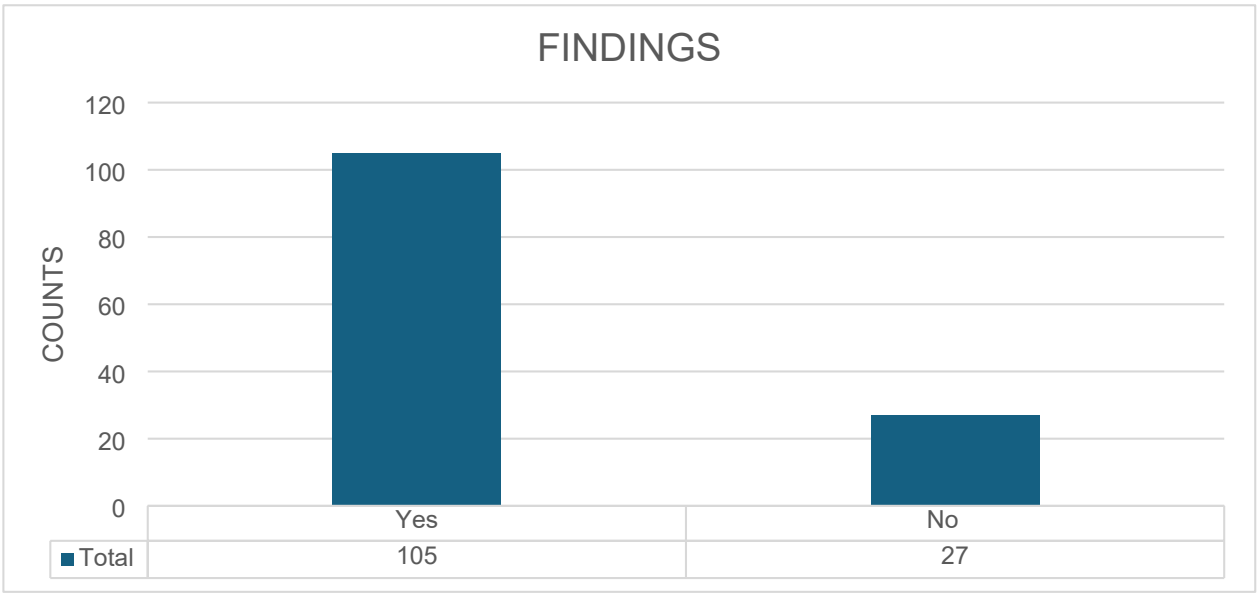


Fig. 10 Publications included in the review, sorted according to the presence (Yes) or absence (No) of findings related to the relationship between human emotions and people's anthropomorphism towards robots

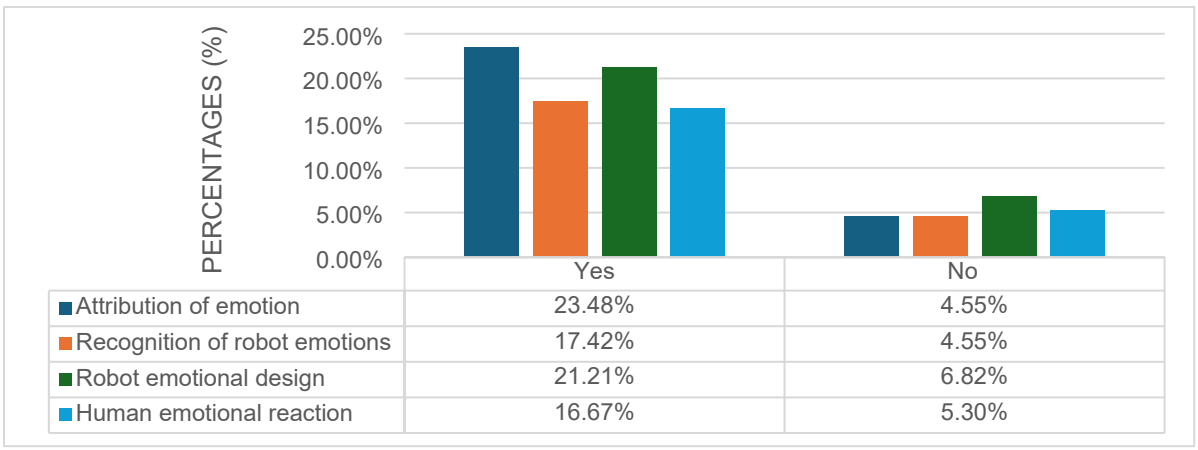


Fig. 11 Percentages (%) of the reviewed studies, reported separately for (i) the presence (Yes) or absence (No) of findings, and (iii) according to the subgroup of findings (one of the four categories reported in the table below the plot)

However, as detailed under the RQ0 section, what we found was that the term "emotion" encompasses various levels of analysis, also depending on which factor of human-robot interactions researchers are focusing on (the "H" factor, the "R factor, or the "I" factor). This heterogeneity is itself an interesting finding of the review, which may help one to achieve a finer-grained representation of the relationship between emotion and anthropomorphism as it emerges in the literature. Notably, the term "emotion" is used in the HRI literature to refer to people's attribution of emotions to robots and their recognition when expressed by a robot (the "H" factor), to the effectiveness of robots' design to convey human emotional expressions (the "R" factor), and to individuals' emotional reactions to robots' behaviours (the "I" factor). For this reason, the findings are grouped according to the level of analysis of "emotion" in the next paragraphs. As shown in Fig. 11, the subcategory comprising the majority of studies is the one labelled under "attribution of emotions", which historically seems to be the most investigated layers in the HRI field; it was closely followed by the "robot emotional design" category.

Note that some studies are difficult to classify in specific categories; for example, studies exploring whether the participants' perceived anthropomorphism affected people's attributed emotions to robots, which, however, were programmed to display emotions by design (i.e., via voice, speech, and/or haptic contact). In these ambiguous cases, the classification criteria were based on explicit and hierarchical decision rules, applied in this order: (i) the stated aim: a given study is assigned to each category according to the research question or main hypothesis explicitly formulated by the authors; (ii) the manipulated variable: if the study aim is unclear, the study is assigned according to the main variable – for example, if the authors manipulated the robot's expressive channel, it falls under the "robot's emotional design category"; otherwise, if they manipulate the participants' perceive anthropomorphism, the study falls under the "attribution of emotions to robots" category; and (iii) the role of the emotion-related measures: if a measure counts towards categorization only if it is the main dependent variable of the study's central hypothesis. In other words, if the "emotion" (whatever level of analysis) serves only as a manipulation check or as evidence that the design worked – for example, participants' attribution of emotion used to verify that the robot's expressions were effective – it does not determine the category.

When needed to distinguish between two "emotion" levels of analysis, because a given study was in between two or more of them and thus it could not be unambiguously classified in a single category, we applied the following explicit criteria. For example, in a study involving both emotion attribution and recognition measured through the same questionnaire (e.g., with items comprising "does the robot feel the emotion X" and "what emotion is the robot showing?", respectively), the rule applied was to look at whether the items concern matching a label to an expression (recognition) or judging the robot's internal state (attribution), and categorise the study according to dominant item wording in the questionnaire used. Instead, when disentangling between emotion attribution and emotional reactions (e.g., a study measures whether participants attribute a given emotion to the robot and whether it facilitates or inhibits their prosocial behaviour towards the robot), the rule applied involved identifying the dependent variable in the study's stated hypotheses. For example, if the hypothesis (and the framing presented in the Introduction) revolves around the attribution process, the study goes under the "emotion attribution" group, with the prosocial behaviour reported as a consequence. Conversely, if the hypothesis centers on behavioural and attitudinal effects (e.g., helping), and, for example, attribution is measured only as a mediator, then the study goes under the "emotional reaction" group. Despite the boundaries among groups can be thin, below we report all the reviewed studies ($N=132$) as divided into the four groups, according to our analytical taxonomy as described under RQ0.

Attribution of emotion to robots

The focus of this subgroup of studies ($N=37$) was people's attribution of emotions to robots, and whether it varied as a function of the degree to which they ascribe human-like, anthropomorphic traits to robots.

Overall, the majority of these studies (31/37 studies; 83.7%) reported the existence of a relationship between the degree of attributed emotions to robots and people's likelihood of anthropomorphising them (see Fig. 12).

Within the "Yes" category, most studies involving both adults (e.g., [38, 75, 77]) and children [44] reported a *positive* relationship between the two variables of interest, in such a way that the more the robot is perceived as anthropomorphic, the more the participants attribute emotions to it. In some cases, although positive, the directionality of this relationship is opposite, with the emotional and affective capacity ascribed to robots being a strong and consistent predictor of anthropomorphism towards them [67, 102]. However, it is important to note that Crowell et al., [91] reported a *negative* relationship between the two variables, i.e., people were more likely to attribute emotions to non-humanoid robots as compared to humanoids (i.e., robots resembling human-like features).

A minority (6/37 studies; 16.3%) of studies can be also found showing an absence of relationship between attributed emotions to robots and attributed anthropomorphism towards them (the "No" category). In some of them, it was argued that anthropomorphic features are not strictly necessary for humans to attribute emotions to

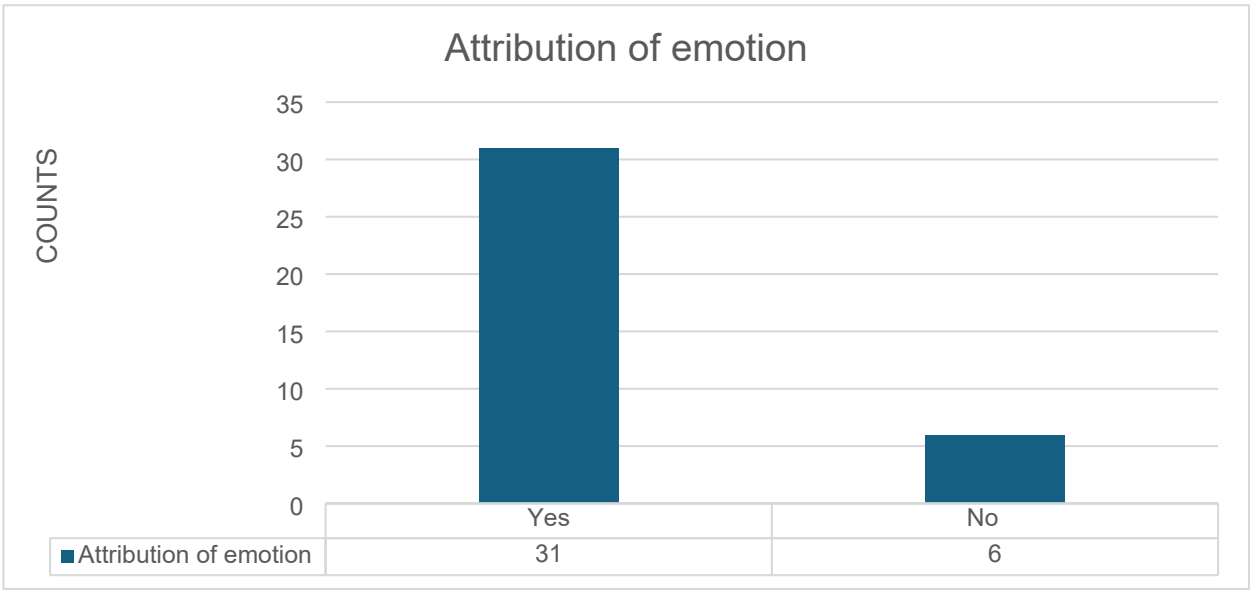


Fig. 12 Subgroup of publications ($N=37$) included in the review and involving studies related to people's attribution of emotions to robots as a function of anthropomorphism; sorted according to whether they reported the existence (Yes) or not (No) of a relationship between humans' attribution of emotions and their tendency to anthropomorphise robots

robots, and that biologically inspired movements [71], for example, haptic cues (Sato et al., 2020), would be sufficient to convey emotional meanings.

When evaluating whether the overall pattern of results still holds in this subgroup of studies, we made some interesting observations. First, verbal measures still represent the most used type of measure for both emotions and anthropomorphism (both over 85%), regardless of whether the studies included in this subgroup revealed the presence or absence of a relationship between emotion and anthropomorphism. Regarding the robot's characteristics (robot presentation), the studies included in this subgroup equally selected a physical, embodied robot (10 studies) as well as a robot presented in the form of an image (10 studies) or video (10 studies); the remaining studies (7) opted for a mixed presentation (for example, video+text or image+physical). Regarding the robot's morphology, almost half of the studies used a robot with human-like anthropomorphic shape (17/37 studies), while the other morphologies were less frequently selected; regarding the robot's behaviour, the two most preferred modalities were either a robot in direct interaction with participants (9/37 studies) or a robot not exhibiting any behaviour, for example, because presented in the form of an image (10/37 studies).

Recognition of robots' emotions

The focus of this subgroup of studies ($N=29$) was people's recognition of emotions displayed by the robot, and whether it varied as a function of the extent to which people attribute human-like, anthropomorphic traits to robots.

Overall, almost 80% of these studies (23/29 studies; 79.3%) reported the existence of a relationship between people's capacity to recognise emotions displayed by robots and their likelihood of anthropomorphising them (see Fig. 13).

Within the "Yes" category, findings showed the existence of a *positive* relationship between the two variables of interest, in such a way that the more participants anthropomorphised robots, the better and faster they were in accurately identifying and recognising the correct emotion when displayed by a robot, when tested both with adults (Breazeal, 2001; [103]; Zhang et al., 2014; [76]) and children [41, 43]. Interestingly, this was also true for participants affected by schizophrenia, although only one among the reviewed studies recruited this type of

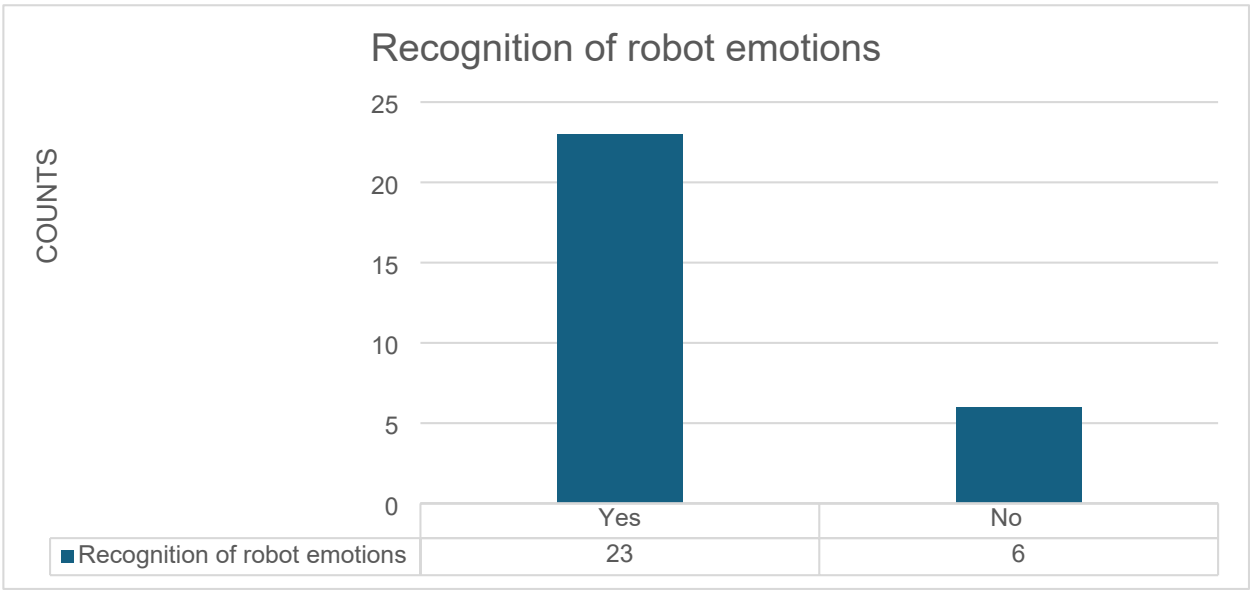


Fig. 13 Subgroup of publications ($N=29$) included in the review and involving studies related to people's ability to recognise emotions displayed by robots as a function of anthropomorphism towards them; sorted according to whether they reported the existence (Yes) or not (No) of a relationship between humans' recognition of emotions and their tendency to anthropomorphise robots

sample [39]. However, it is important to mention that, within this category, there was a study reporting a *negative* relationship between the two variables of interest, i.e., the less participants anthropomorphised robots, the better they recognised emotions displayed by a robot; indeed, Ko and colleagues [104] observed that participants recognised emotions significantly more accurately when they were expressed by an animal-like robot (Pleo, which does not have an anthropomorphic, human-like shape) than by the humanoid robot NAO.

The "No" category, albeit poorly populated (6/29 studies; 20.7%), offers evidence showing no relationship between the participants' ability in recognising emotions displayed by a robot and the anthropomorphism attributed to it. For example, Tsiourti and colleagues (2017) argued that a higher degree of anthropomorphic features does not necessarily lead to higher accuracy in emotion recognition, since they observed a similar recognition rate between the abstract robot Hobbit and the humanoid robot Pepper. However, the Hobbit robot is indeed described as having a human-like appearance, although its features are more "abstract" than Pepper's (Tsiourti et al., 2017). Along a similar line, other studies reported that the emotional expressions displayed by a robot were correctly identified above the chance level, regardless of the anthropomorphism exhibited and attributed to it (Wiltgen et al., 2010; Beck et al., 2012; [105]; Alimardani et al., 2024; Ito et al., 2024).

When evaluating whether the overall pattern of results still holds in this subgroup of studies, we made some interesting observations. First, verbal measures still represent the most used type of measure for both emotions (79.3%; 23/29 studies) and anthropomorphism (72.4%; 21/29 studies), and this regardless of whether the studies included in this subgroup revealed the presence or absence of a relationship between emotion and anthropomorphism. Regarding the robot's characteristics, in terms of robot presentation, we observed a prevalence of studies opting for the video presentation of the robot (11 studies). Regarding the robot's morphology, more than half of the studies used a robot resembling a human-like anthropomorphic shape (20/29 studies), while the other morphologies were less frequently selected; regarding the robot's behaviour, the two most preferred modalities were either a robot in direct interaction with participants (10/29 studies) or a robot not exhibiting any behaviour, for example, because presented in the form of an image (11/29 studies).

Robots' "emotional" design

The focus of this subgroup of studies ($N=37$) was on the robot's display of emotions, in terms of the implemented robotic design to convey emotions, as a function of people's attributed anthropomorphism towards it. In other words, whether the way the robots are designed and their behaviours implemented to convey emotions varies depending on the extent to which people attribute anthropomorphic features to them.

Overall, more than 75% of these studies (28/37 studies; 75.7%) reported the existence of a relationship between people's emotional reactions and their likelihood of anthropomorphising them (see Fig. 14).

The "Yes" category comprises studies in which anthropomorphic features of robots are shown to positively influence human behaviours; for example, when a robot highly resembles a human-like shape, it gives participants greater positive impressions, which in turn are processed more easily than neutral ones [73]. It seems to be true also in elderly people, since anthropomorphic robot expressions (defined as "natural") and gestures successfully conveyed the six basic emotions or happiness, sadness, anger, fear, disgust, and surprise [50].

However, in most publications of this subgroup, the relationship between the two variables of interest appears to be reversed, meaning that "emotional features" are found to positively impact the perception of anthropomorphism. For instance, adding tactile sensations (cutaneous sensations) allowed for more natural human-robot interaction, making the robot appear friendlier, as well as what other humans can do (Takanishi et al., 2000). Similarly, when a robot is programmed to display more positive facial expressions, it elicits greater "anthropomorphic trustworthiness" in participants [93]. Still, emotional gait patterns implemented in a robot positively affected the perceived anthropomorphism (Destephe et al., 2014). The same happens with implementing non-verbal emotional responses in robots, e.g., when they mimic human emotions such as happiness or fear (Eyssel et al., 2010), or with adding emotion-specific vertical oscillations, such as "bouncy" movements for happiness or "stomping" for anger [98].

However, a minority of studies (9/37 studies; 24.3%) suggest that emotional behaviours and reactions in a robot are not associated with attributed anthropomorphism. For example, Zinina and colleagues [106] argued that exhibiting non-verbal social cues (i.e., mimicking social and emotional behaviour) is sufficient to make a robot attractive and intuitive to humans without needing a realistic, anthropomorphic, human face. Similarly, Sial and

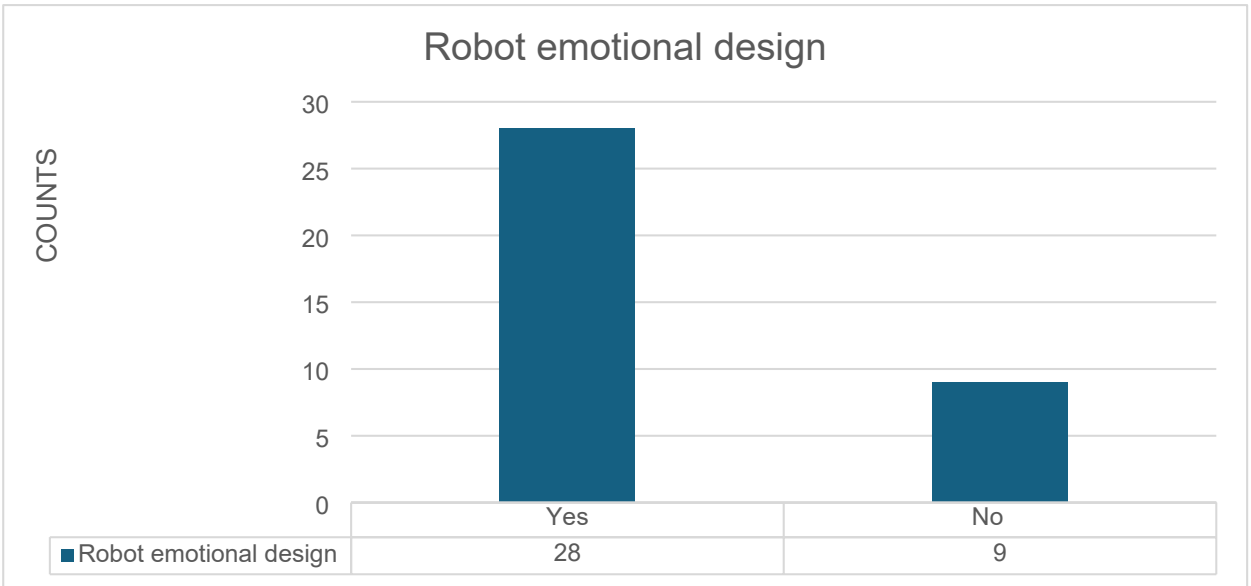


Fig. 14 Subgroup of publications ($N=37$) included in the review and involving studies related to robots' "emotional" design as a function of anthropomorphism towards them; sorted according to whether they reported the existence (Yes) or not (No) of a relationship between robots' "emotional" design and their tendency to anthropomorphise robots

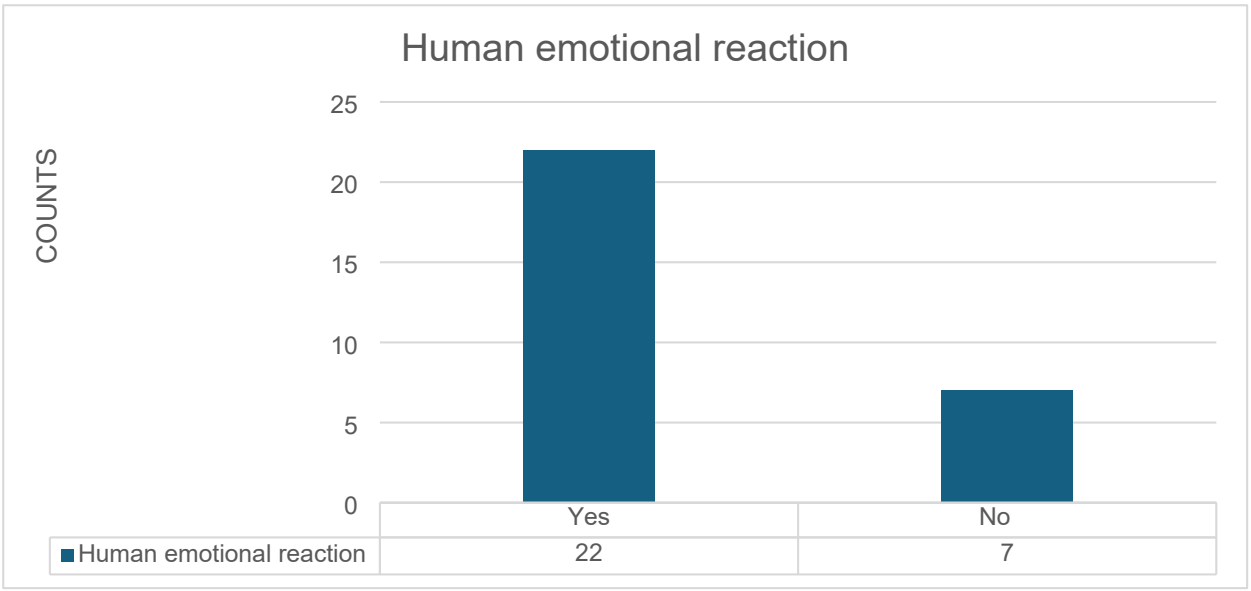


Fig. 15 Subgroup of publications ($N=29$) included in the review and involving studies related to people's emotional reactions as a function of anthropomorphism towards them; sorted according to whether they reported the existence (Yes) or not (No) of a relationship between people's emotional reactions and their tendency to anthropomorphise robots

colleagues [107] argued that anthropomorphic features are not required for a robot to communicate emotions effectively. Indeed, a "smooth" motion is sufficient for a mechanoid robot to appear more natural and predictable, enhancing the effectiveness of the interaction.

When evaluating whether the overall pattern of results still holds in this subgroup of studies, we made the following observations. First, verbal measures still represent the most used type of measure for both emotions (70.3%; 26/37studies) and anthropomorphism (59.4%; 22/37studies) regardless of whether the studies included in this subgroup revealed the presence or absence of a relationship between emotion and anthropomorphism. Regarding the robot's characteristics, in terms of robot presentation, the studies included in this subgroup showed a prevalence for the physical presentation of the robot, i.e., an embodied robot physically sharing the experimental settings with participants (23/37 studies). Regarding the robot's morphology, more than half of the studies used a robot resembling a human-like anthropomorphic shape (22/37 studies), while the other morphologies were less frequently selected; regarding the robot's behaviour, the most preferred modality was the robot in direct interaction with participants (22/37 studies).

Human emotional reactions

The focus of this subgroup of studies ($N=29$) was on people's emotional reactions as a function of their attributed anthropomorphism to them. In other words, whether people's emotional reactions to robots vary depending on the extent to which they attribute anthropomorphic features to them.

Overall, more than 75% of these studies (22/29 studies; 75.9%) reported the existence of a relationship between people's emotional reactions to and their likelihood of anthropomorphising them (see Fig. 15).

Within the "Yes" Category, the majority of studies reported the existence of a *positive* relationship between people's emotional reactions and their likelihood of anthropomorphising them; i.e., people were more likely to experience emotional reactions when interacting with robots who were assigned a higher level of anthropomorphic attributes. More specifically, the general trend seems to be that interacting with a robot having highly attributed anthropomorphic features increases people's positive feelings and decreases negative ones. For example, participants' sense of efficacy and emotional engagement are significantly powered by ascribing human-like

characteristics to robots [108]. From the user's perspective, their positive emotional reactions, such as satisfaction [78] and trust (Pinney et al., 2023; [69]), increase. Conversely, their negative reactions, such as fear and uneasiness [109], as well as anxiety [52] and the uncanny (Alimardani et al., 2024), decrease.

Notably, similar effects emerged when focusing on people's emotional reactions directed at robots. Indeed, some of the reviewed studies showed that a higher degree of anthropomorphic attribution to robots leads to a significant increase in people's positive attitudes towards them, e.g., a greater liking [72, 110, 111]; empathy, for example, when observing a harmed robot [90]; and intimacy between humans and robots [112].

While the "No" category comprises a minority of studies (7/29 studies; 24.1%), a lack of relationship between people's emotional reactions and their tendency to anthropomorphise robots is documented. For example, in their study, Hofree and colleagues [89] argued that robots' attributed anthropomorphic features do not modulate people's emotional reactions, in terms of the facial expression (positive or negative) they exhibited when performing an interactive game with a robot. Similarly, the robot's anthropomorphic appearance did not significantly influence participants' emotional reaction, in terms of trust exhibited towards the robot [80] or their sharing behaviour during a human-robot interaction task, in the case of children [113].

When examining this subgroup of studies, we observed that, firstly, verbal measures still represent the most used type of measure for both emotions (82.6% 24/29 studies) and anthropomorphism (70%; 20/29 studies) and this regardless of whether the studies included in this subgroup revealed the presence or absence of a relationship between emotion and anthropomorphism. Regarding the robot's characteristics (robot presentation), the studies included in this subgroup showed a prevalence for the physical presentation of the robot, i.e., an embodied robot physically sharing the experimental settings with participants (10/29 studies), although the robot presented in the form of an image was the second preferred modality (8/29 studies). Regarding the robot's morphology, more than half of the studies used a robot resembling a human-like anthropomorphic shape (15/29 studies), while the other morphologies were less frequently selected; regarding the robot's behaviour, the most preferred modality was the robot in direct interaction with participants (14/29 studies).

Discussion

This scoping review aimed to map the existing HRI literature dealing with the relationship between emotions and anthropomorphism towards robots. To do so, we reflected on how these two constructs have been conceptualised, operationalised and measured, to see whether emotions relate, and in what sense, to people's likelihood to anthropomorphise robots.

The first significant finding of this work is that the term "emotion" does not carry a single, univocal meaning in the HRI literature. Rather, as documented in response to RQ0, it operates as a conceptually plural construct, encompassing at least four different foci depending on the relevant dimensions of human-robot interactions that researchers are interested in: the human ("H" factor), the robot (the "R" factor), or the interaction itself (the "I" factor). Considering the H (human) factor, the term "emotion" can refer to either people's *attribution of emotions to a robot* during an interaction, or people's ability to correctly and accurately *recognise emotional expressions displayed by a robot*. Considering the R (robot) factor, the level of analysis is represented by the *robot's emotional design* – in terms of appearance and implemented behaviour – which effectively conveys human-like emotions, thus making emotional attribution and recognition more straightforward. Considering the interaction between the two systems (the "I" factor), the level of analysis is represented by the emotional reactions that people experience when observing or interacting with a robot that exhibits certain behaviours or has a certain appearance. Far from being merely terminological, this plurality reflects the various levels of analysis that the term "emotion" carries on in the HRI domain, as well as deep conceptual tensions in how the domain has approached the study of emotions in nonhuman agents. Indeed, as our findings show, the following four questions are substantially different from one another, both empirically and theoretically: do people attribute emotions to robots? Do people recognise emotions displayed by robots? Do people react emotionally to robots? Can robots effectively display emotions?

This complexity has largely gone unacknowledged in the HRI literature. Greater conceptual precision in what level of analysis researchers are targeting in their studies when approaching emotions would therefore be not only desirable, but necessary for the HRI field to advance coherently.

This fragmentation into four different levels of analysis, however, may raise a more philosophical question about the possible dependency of these four foci of "emotions" on one another. Consider, for example, the H-factor levels of analysis: attribution and recognition of emotions. It is not obvious that the two can be logically separated, since (*cognitively*) attributing an emotion to a robot may presuppose some a priori act of (*perceptual*) recognition – whereby it is, however, still debated whether perception itself necessarily implies prior cognitive categories [114]. Or, consider the relationship between emotional reaction and attribution: empathetic responses, for instance, may already implicate attributive processes rather than constituting a distinct psychological outcome that follows from them. Relatedly, this also introduces another, and perhaps more fundamental, problem, i.e., whether something like an "emotional" design in robots can be conceived at all independently of any act of attribution or recognition: each behavioural expression (in artificial agents, as well as in humans and nonhuman animals) becomes emotionally significant only for someone who has the relevant interpretative resources for "decoding" and classifying it as such – much as some strange scribbles on the wall of an ancient temple remain meaningless to those who do not know they are part of a writing system (and this even before trying to understand their semantic meaning). Such resources are not reducible to design choice alone and depend on many subjective and cultural factors [115]. Addressing these questions would require a philosophical work that goes beyond the scope of the present review. Yet we do believe that the HRI field would benefit from confronting these topics explicitly, rather than leaving them embedded in unexamined terminological choices.

These considerations have direct implications for how we should understand the relationship between emotions and anthropomorphism toward robots. From a human-centred perspective, anthropomorphism can be considered a psychological phenomenon that significantly affects people's interaction with non-human entities such as robots [116]. Specifically, it promotes a more intuitive interaction because it allows people to draw on scripts that are well-known and routinely adopted in human-human interactions (Hoffman and [14]; Clodic et al., 2017). However, as our conceptual mapping reveals, which emotional phenomenon is at stake – attribution, recognition, reaction, or design – substantially shapes both the nature of the relationship with anthropomorphism and its theoretical interpretation.

When assessing the theoretical conceptualisations of the two constructs of interest (emotions and anthropomorphism towards robots), in terms of both their theoretical framework and definitions (RQ1), the 132 reviewed studies revealed the following. Regarding *emotions*, without a clear prevalence of one theoretical framework over another, what clearly emerges is that, in the HRI domain, the tendency is to apply the same few models of emotions that are widely known and adopted in human studies, for example, the model of Six Basic Emotions [59, 60], the Circumplex Model of Affect (CMA; [56]) and the Pleasure-Arousal-Dominance (PAD) Model [57]. A greater variety is observed when focusing on the definitions of emotions provided in the reviewed papers. When available, since many studies did not supply a formal definition of emotion, they could be either generic (e.g., emotions defined as experience, intelligence, a form of affective communication of response, a multidimensional construct, a human-like state) or more specific, if a given study focuses on a single emotion (for example, empathy). This heterogeneity of definitions is consistent with the conceptual plurality documented in response to RQ0: the absence of a shared definition of "emotion" across the reviewed studies is not incidental but reflects the same underlying ambiguity that motivated the distinction between attribution, recognition, reaction, and design in the first place. What seems to emerge is that – not surprisingly – the investigation of emotions in the HRI domain heavily takes inspiration from the study of human-human interactions, which offers building blocks to understand how emotions can be conceptualised, transmitted, and investigated during human-robot interactions.

Unlike the case of emotion, the theoretical frameworks addressing *anthropomorphism towards robots* were consistent across the reviewed publications. Indeed, one of the most cited is the three-factor model of anthropomorphism proposed by Epley and colleagues [6], which proposes an explanation of why and when humans attribute anthropomorphic features to generic "nonhuman entities"; it was closely followed by the Media Equation

Theory [85], which derives from studies on computers and media, and the Dual Model of Dehumanisation [88], which revolves around the context of social psychology, As for emotions, the study of anthropomorphism in HRI has largely borrowed from other disciplines investigating interactions between humans. We also observed consistency when focusing on the definitions of anthropomorphism provided in the reviewed studies, which mostly refer to three main categories: humanness, mind perception/attribution, and the definition taken from Epley and colleagues' study ("the tendency of imbuing the real or imagined behaviour of nonhuman agents with human-like characteristics, motivations, intentions, or emotions", 2007).

However, two important aspects are worth highlighting here. First, what we observed was that the majority of the meanings provided in the reviewed studies defined "anthropomorphism" in terms of *psychological anthropomorphism* as described by Fischer [10]; that is, the attribution of human-like traits to either human or nonhuman agents. Instead, the definition of anthropomorphism intended as the way of implementing human-like traits when designing a robot (*anthropomorphism by design*, [10]) is not adopted in the reviewed publications, despite some of the studies sticking to this definition and directly manipulated the degree of anthropomorphism expressed by a robot as a variable in the design (e.g., Bartneck et al., [83], Letheren, [72], Xie, [78], Yang, [79]. Second, and related to the previous point, some of the definitions of (psychological) anthropomorphism provided in the reviewed studies, for example, mind perception/attribution, may overlap (at least partially) with the two primary constructs of this review, i.e., emotions and anthropomorphism. Thus, it makes it difficult to disentangle whether a connection between the two may be due to a definitional and/or measurement overlap, or to a reliable empirical relationship between separable constructs. For example, in the model of Mind Perception [28, 29], the Experience dimension covers the capacity to feel, including emotional states, and it is distinct from the Agency dimension. Therefore, attributed emotions may be considered one component of the attributed mind, and the association between emotion attribution and psychological anthropomorphism may partly reflect a definitional or measurement overlap rather than an empirical relationship between separable constructs. One possibility to disentangle the constructs would be to focus on the level of analysis theorised by Fischer [10], i.e., anthropomorphism by design; this kind of anthropomorphism is separable, in both theoretical and methodological terms, from the concept of "emotion", reducing the possibility of the conceptually problematic overlap that might emerge when focusing on psychological anthropomorphism.

When focusing on the *methods* used to explore the relationship between emotions and anthropomorphism towards robots (RQ2), we observed consistent patterns in the choice of methods. First, it is important to mention that the majority of the reviewed publications included studies conducted in a controlled laboratory setting, which allows for higher control and replicability of results with respect to, for example, field or online studies. Regarding the *stimuli*, the reviewed studies reflected a large variety of choices with respect to robots' characteristics, in terms of presentation, morphology, and implemented behaviours. However, overall, there were some preferred choices over others. The most preferred modality of robot presentation was using a physical robot, i.e., an embodied robot sharing the same experimental setting as participants in a real-time scenario (38.63%). Furthermore, in the majority of the reviewed studies, the robot used resembled an anthropomorphic shape (56.06%), and it was programmed to directly interact with participants (41.66%).

What these methodological choices can tell us about the current state of HRI research on emotion and anthropomorphism towards robots is that the field seems to appropriately prioritise embodied and situated interactions, recognising that emotional responses are more likely to emerge when the robot is physically present and directly engaged with the user, as opposed to more abstract or purely virtual settings. At the same time, however, considering that emotions also involve cognitive and context-dependent processes [117], exploring settings that go *beyond* (i.e., *not* in place of) the robot's physical presence – such as narrative-based scenarios, delayed or reflective evaluations, and contexts requiring interpretation – may be equally valuable, especially for investigating more complex forms of emotion that are not directly tied to immediate behavioural responses.

Regarding *measures*, i.e., how emotion and anthropomorphism towards robots were measured in the reviewed publications, what emerges seems to be a prevalence of verbal measures (e.g., participants' subjective ratings, evaluations collected via Likert scales, or qualitative methods such as semi-structured interviews) for both

variables of interest. This methodological choice perfectly aligns with what Corneille and Gawronski recently argued in their paper (2024), where they discussed the advantages of using self-report measures instead of indirect behavioural measures. They claim that self-report measures should be considered superior because, on average, they are more reliable, are generally better at predicting people's behaviour, and are more flexible, allowing for the assessment of people's thoughts and feelings that, when related to complex issues, are more difficult to capture with the use of implicit behavioural measures [118]. However, it is also likely that self-report and verbal measures in general offer only a partial view, especially considering that other types of measures (behavioural and neurological) are almost completely disregarded in the reviewed literature. As a consequence, little is known about mechanisms regulating emotions in HRI contexts at both the behavioural and brain levels. From our perspective, an integrated approach would be preferred. Indeed, using indirect behavioural and neural measures would allow to extend current knowledge regarding the phenomenon without completely relying on participants' subjective reports, which are still exposed to numerous biases such as social desirability and/or a lack of conscious access to the relevant thoughts and feelings (Nisbett and Wilson, 1977; Greenwald and Banaji, 1995).

Regarding the findings about the relationship between emotions and anthropomorphism towards robots (RQ3), what emerged is that the majority of the reviewed studies (105/132) reported the existence of a statistically significant relationship between the two variables of interest, and this regardless of which of the four levels of "emotion" – as documented in response to RQ0 - was at stake in a given subgroup of studies. However, it is important to note here that this proportion does not estimate a single homogeneous relationship between two separately defined constructs; rather, it describes the frequency with which the corpus of the reviewed studies reports some association between heterogeneous levels of analysis and phenomena grouped under the terms "emotion" and "anthropomorphism", respectively. More importantly, since the two constructs are not entirely separable (i.e., partial terminological and conceptual overlaps between "emotion" and anthropomorphism" have emerged when examining the definition provided by the current HRI literature), it can also be that the observed association may be predictable by construction. To support (or disconfirm) the existence of a relationship, further studies will need to provide statistical data giving a precise estimate of the effect, which, however, falls outside the scope of this scoping review.

Nonetheless, since we clearly showed the heterogeneity that investigations of the constructs of "emotion" and "anthropomorphism" carry on, at both theoretical (RQ0-RQ1) and methodological levels (RQ2), we also examined whether and how the nature of the relationship between emotion and anthropomorphism varied across the four foci of emotions. This issue falls outside the original research questions; however, we do believe that it clearly demonstrates the analytical value of the taxonomy we proposed, and also indicates a possible future direction to follow in HRI studies when investigating the potential links between these two constructs.

The first observation we made is that this pattern (i.e., prevalence of a positive relationship linking emotion and anthropomorphism) holds across studies focusing on people's attribution of emotions to robots, since the majority of these studies (more than 80%) report that higher levels of anthropomorphism were associated with more likely attribution of emotions to robots. A similar trend emerged also in studies focusing on people's recognition of emotions displayed by the robot, whose majority (almost 80%) showed that higher levels of anthropomorphism were associated with greater accuracy and speed in recognising the correct human emotion when displayed by the robot. Likewise, studies focusing on human emotional reactions to robots' behaviour or appearance showed a comparable pattern. Indeed, the majority of these studies (~ 75%) showed that higher levels of anthropomorphism were associated with a higher likelihood for participants to experience and report emotional responses. More in depth, interacting with or observing robots with anthropomorphic features appears to increase positive feelings and decrease negative ones, both self-directed (e.g., increased sense of self-efficacy, reduced anxiety) and robot-directed (e.g., greater liking, less fear and uneasiness). Notably, the only cluster of studies that showed a slightly different pattern was the one focusing on robot emotional design. Indeed, most of these studies (~ 75%) still reported a positive relation between the two constructs, i.e., when robots are implemented with emotional features (e.g., facial expressions, haptic feedback, body motion), these features tended to enhance their perceived anthropomorphism, making them appear more natural and human-like. It is important to note here that focusing

on and assessing a precise direction of the relationship between emotion and anthropomorphism falls outside the scope of this scoping review, which seeks to evaluate only the presence/absence of a relationship between these two constructs (according to the HRI review protocol developed by Thellman and colleagues, [16]). However, we do believe that it might offer an interesting insight for empirical research in the HRI domain. However, sticking to our analytical taxonomy (RQ0) and thus evaluating how the patterns of findings vary across the four levels of analysis for emotion (attribution of emotion to robots, recognition of emotions when displayed by the robot), we observed that the overall pattern of results is still reflected at the subgroup levels. This means that, across subgroups, the reviewed studies show a strong prevalence of physically embodied, anthropomorphic robots situated in direct interaction with participants; information on both constructs (emotion and anthropomorphism) is predominantly collected using verbal measures. Note that this pattern does not fully represent the complexity of the current HRI literature on these phenomena; it simply suggests that, at present, most studies in this subfield share methodological characteristics that serve as a starting point for drawing conclusions. Another important point to mention is that, in terms of robots' characteristics, the subgroup of studies focusing on the attribution of emotion to robots is the one that seems to reveal higher internal complexity, since there is no clear prevalence – in terms of descriptive frequency – among modalities (e.g., robot presentation, morphology, and behaviour). By paralleling the psychological dimension of anthropomorphism (i.e., the user attributes anthropomorphic traits to robots), we may suggest that also the *psychological dimension of emotions* (i.e., the user attributes emotions to robots) may be the one more internally nuanced and deserving further attention in the HRI field. However, the other "emotion" foci – recognition of emotions, robot's emotional design and user's emotional reactions – are worthy of further attention; and, overall, these preliminary conclusions need to be statistically confirmed, for example through meta-analyses that quantitatively combine data from those studies to produce a precise estimate of effect.

Conclusions and limitations

Taking these findings together, our scoping review suggests that the relationship between anthropomorphism and emotions in HRI is currently being studied under a variety of different conceptual and methodological frameworks. While this variety of perspectives is clearly useful for achieving a nuanced and realistic representation of the complexities of human-robot interaction processes and should be welcomed, it also implies that more studies and statistical data are needed to explore the various aspects of the phenomenon in depth and integrate them into a coherent model.

This complexity is not surprising, especially if we consider that (i) "emotion" can partially overlap with (and be included in) anthropomorphism – after all, when humans attribute human-like qualities to robots, these qualities may well include mental states but likely also desires, beliefs, and emotions; and that (ii) the term "emotion" itself encompasses a variety of levels of analysis, which does not lead it to a single, univocal meaning and operationalisation. Overall, our findings suggest that anthropomorphism towards robots is recognized as a central topic for understanding human-robot interactions [118], and deserves further investigation especially with reference to the "affective" sub-domain revolving around the study of (human) emotions.

Although this review offers valuable insights for the HRI community, it is important to underline two important aspects that may further limit the generalisations of our findings. First and foremost, the majority of studies (88.7%) involved neurotypical adults; a small portion of studies involved neurotypical children or the elderly, and only one study involved people diagnosed with schizophrenia. Therefore, our results must be taken with caution, as they cannot be easily generalised to the whole general population. Second, it is fundamental to clarify that, although the majority of studies suggested the existence of a relationship between the two variables of interest (emotion and anthropomorphism towards robots), it is equally true that other evidence showed divergent findings. Thus, these findings must also be taken into account to fully understand what is going on between human emotions and robots' anthropomorphic features.

Beyond the two limitations already mentioned, several additional constraints should be acknowledged. First, the near-exclusive reliance on verbal self-report measures limits our understanding of the implicit and unconscious mechanisms underlying emotional responses in HRI contexts, where behavioural and neurological data would provide complementary and arguably more objective evidence. Second, the considerable heterogeneity across the included studies – in terms of sample size, experimental setting, and measurement tools – makes direct comparisons between findings particularly challenging. Third, restricting the search to English-language publications may have introduced a cultural bias, potentially overrepresenting Western perspectives on anthropomorphism, which is known to vary substantially across socio-cultural contexts. Fourth, as with any review, publication bias cannot be ruled out: studies reporting null results may have been less likely to be published, possibly leading to an overestimation of the relationship between human emotions and anthropomorphism towards robots. Fifth, the definitional choice to exclude computers and disembodied avatars, while theoretically motivated, leaves aside a growing and relevant body of literature on digital and conversational agents – such as voice assistants and chatbots – that are increasingly present in people's everyday lives. Finally, the predominance of laboratory-based studies limits the ecological validity of the findings, as emotional responses elicited in controlled experimental settings may not fully reflect those arising in real-world human-robot interactions, such as in domestic, clinical, or industrial environments.

Acknowledgements Not applicable.

Author contributions CR conceptualised the study and the methodology, performed the literature search and analysis, and wrote the manuscript (original draft preparation). IZ conceptualised the study and methodology, and wrote the manuscript (original draft preparation). PG conceptualised the study and methodology, and wrote the manuscript (original draft preparation). ED conceptualised the study and methodology, administered the project, and wrote the manuscript (original draft preparation). All the authors revised the manuscript.

Funding This work was supported by the European Union under the Next Generation EU program, in the framework of the project "Fostering STEM inclusion among girls with migrant background: the role of educational robots" (PRIN 2022 PNRR, grant n. P2022B2MN2).

Data availability An Excel file containing detailed information regarding the articles included in this review, as well as its summary (in a separate Excel file), are made available at the following repository: https://osf.io/e8dxu/overview? view_only=a2eb214ce8e84c298600b8903c79a01c.

Declarations

Ethics approval and consent to participate The present manuscript is a review summarising the existing literature, and thus did not require human participants. For this reason, documented approval from the institutional review board of the University of Milano-Bicocca has not been asked.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Heider F, Simmel M. An experimental study of apparent behaviour. Am J Psychol. 1944;57(2):243–59.
2. Solis J, Ozawa K, Takeuchi M, Kusano T, Ishikawa S, Petersen K, Takanishi A. Biologically-inspired control architecture for musical performance robots. Int J Adv Rob Syst. 2014;11(10):172.
3. Spiers BM. Music and the Spectacle of Artificial Life. The University of Chicago; 2020.
4. Wood S. Living dolls: a magical history of the quest for mechanical life. Faber and Faber; 2002.
5. Breazeal C. Regulation and entrainment in human—robot interaction. Int J Robot Res. 2002;21(10–11):883–902.
6. Epley N, Waytz A, Cacioppo JT. On seeing human: a three-factor theory of anthropomorphism. Psychol Rev. 2007;114(4):864.
7. Perez-Osorio J, Wykowska A. Adopting the intentional stance toward natural and artificial agents. Philosophical Psychol. 2020;33(3):369–95.
8. Kühne R, Peter J. Anthropomorphism in human–robot interactions: a multidimensional conceptualization. Communication Theory. 2023;33(1):42–52.
9. Spatola N, Wudarczyk OA. Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. Comput Hum Behav. 2021a;124:106934.
10. Fischer K. Tracking anthropomorphizing behavior in human-robot interaction. ACM Trans Human-Robot Interact (THRI). 2021;11(1):1–28.
11. Jansen S, Rehren O, Jahn K, Ohler P, Rey GD. Make It More Human! A Systematic Literature Review of the Anthropomorphic Processes on Empathy. Hybrid Societies: Hum Interact Embodied Technol. 2026;1:187–96.
12. Schömbs S, Klein J, Roesler E. Feeling with a robot—the role of anthropomorphism by design and the tendency to anthropomorphize in human-robot interaction. Front Rob AI. 2023;10:1149601.
13. Chakrabarti B, Robinson P, Riek LD. (2009). How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*.
14. Breazeal C. Designing sociable robots. MIT Press; 2004.
15. Marchesi S, Ghiglino D, Ciardo F, Perez-Osorio J, Baykara E, Wykowska A. Do we adopt the intentional stance toward humanoid robots? Front Psychol. 2019;10:450.
16. Thellman S, De Graaf M, Ziemke T. Mental state attribution to robots: A systematic review of conceptions, methods, and findings. ACM Trans Human-Robot Interact (THRI). 2022;11(4):1–51.
17. Breazeal C. Emotion and sociable humanoid robots. Int J Hum Comput Stud. 2003;59(1–2):119–55.
18. Zhang T, Kaber DB, Zhu B, Swangnetr M, Mosaly P, Hodge L. Service robot feature design effects on user perceptions and emotional responses. Intel Serv Robot. 2010;3(2):73–88.
19. Nass C, Moon Y. Machines and mindlessness: Social responses to computers. J Soc Issues. 2000;56(1):81–103.
20. Giannopulu I. Neuroscience, robotics and virtual reality: internalised vs externalised mind/brain. Springer; 2019.
21. Phillips EK, de Visser EJ. Humanlikeness as design, anthropomorphism as inference: a conceptual framework for human–robot interaction. Front Cognition. 2026;5:1786256.
22. Russell S, Norvig P. Artificial intelligence: a modern approach. 3rd ed. Upper Saddle River, EUA: Prentice-Hall; 2010.
23. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res synthesis methods. 2014;5(4):371–85.
24. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol. 2018;18(1):143.
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Moher D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, *372*.
26. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res synthesis methods. 2020;11(2):181–217.
27. Airenti G. The cognitive bases of anthropomorphism: from relatedness to empathy. Int J Social Robot. 2015;7(1):117–27.
28. Gray HM, Gray K, Wegner DM. Dimensions of mind perception. Science. 2007;315(5812):619–619.
29. Waytz A, Gray K, Epley N, Wegner DM. Causes and consequences of mind perception. Trends Cogn Sci. 2010;14(8):383–8.
30. Spatola N, Marchesi S, Wykowska A. The intentional stance test-2: How to measure the tendency to adopt intentional stance towards robots. Front Rob AI. 2021b;8:666586.
31. Hortensius R, Kent M, Darda KM, Jastrzab L, Koldewyn K, Ramsey R, Cross ES. Exploring the relationship between anthropomorphism and theory-of-mind in brain and behaviour. Hum Brain Mapp. 2021;42(13):4224–41.
32. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971;76(5):378.

33. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychol Bull. 1971;76(5):365.
34. Team RC. (2014). A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020. *URL*https://www.R-project.org. *R-project. org*
35. Gamer M, Lemon J, Gamer MM, Robinson A, Kendall's W. Package 'irr'. Var coefficients interrater Reliab Agreem. 2012;22:1–32.
36. Chi OH, Chi CG, Gursoy D. Seeing personhood in machines: conceptualizing anthropomorphism of social robots. J Service Res. 2025;28(1):78–92.
37. Terada K, Yamada S. Mind-reading and behavior-reading against agents with and without anthropomorphic features in a competitive situation. Front Psychol. 2017;8:1071.
38. Yam KC, Bigman YE, Tang PM, Ilies R, De Cremer D, Soh H, Gray K. Robots at work: People prefer—and forgive—service robots with perceived feelings. J Appl Psychol. 2021;106(10):1557.
39. Raffard S, Bortolon C, Khoramshahi M, Salesse RN, Burca M, Marin L, Capdevielle D. Humanoid robots versus humans: How is emotional valence of facial expressions recognized by individuals with schizophrenia? An exploratory study. Schizophr Res. 2016;176(2–3):506–13.
40. Augello A. Unveiling the reasoning processes of robots through introspective dialogues in a storytelling system: A study on the elicited empathy. Cogn Syst Res. 2022;73:12–20.
41. Beck A, Cañamero L, Hiolle A, Damiano L, Cosi P, Tesser F, Sommavilla G. Interpretation of emotional body language displayed by a humanoid robot: A case study with children. Int J Social Robot. 2013;5(3):325–34.
42. Istenič A, Podpečan V, Rosanda V, Zhai X. Social robot radical innovation: capacities and status attributed to the NAO robot by 11 to 12-year-old students and preservice teachers' perceptions. Cogent Educ. 2025;12(1):2489820.
43. Koch SA, Stevens CE, Clesi CD, Lebersfeld JB, Sellers AG, McNew ME, Hopkins MI. A feasibility study evaluating the emotionally expressive robot SAM. Int J Social Robot. 2017;9(4):601–13.
44. Pasupuleti D, Sasidharan S, Manikutty G, Das AM, Pankajakshan P, Strauss S. Co-designing the embodiment of a minimalist social robot to encourage hand hygiene practises among children in India. Int J Social Robot. 2023;15(2):345–67.
45. Peretti G, Manzi F, Di Dio C, Cangelosi A, Harris PL, Massaro D, Marchetti A. Can a robot lie? Young children's understanding of intentionality beneath false statements. Infant Child Dev. 2023;32(2):e2398.
46. Saylor MM, Somanader M, Levin DT, Kawamura K. How do young children deal with hybrids of living and nonliving things: The case of humanoid robots. Br J Dev Psychol. 2010;28(4):835–51.
47. Torpegaard J, Knudsen LS, Linnet MP, Skov MB, Merritt T. Preschool children's social and playful interactions with a play-facilitating cardboard robot. Int J Child-Comput Interact. 2022;31:100435.
48. Woods S. Exploring the design space of robots: Children's perspectives. Interact Comput. 2006;18(6):1390–418.
49. Mao K, Li Z, Lu Z, Nishimura Y, Li T. Nao, I can read your mind: Preschool-aged children construct theory of artificial mind (TOAM). Int J Human–Computer Interact. 2025;41(21):13555–68.
50. Jo S, Hong S. (2025). The development of human-robot interaction design for optimal emotional expression in social robots used by older people: design of robot facial expressions and gestures. IEEE Access.
51. Liu SX, Shen Q, Hancock J. Can a social robot be too warm or too competent? Older Chinese adults' perceptions of social robots and vulnerabilities. Comput Hum Behav. 2021;125:106942.
52. Torta E, Werner F, Johnson DO, Juola JF, Cuijpers RH, Bazzani M, Bregman J. Evaluation of a small socially-assistive humanoid robot in intelligent homes for the care of the elderly. J Intell Robotic Syst. 2014;76(1):57–71.
53. Hippolyte L, Barisnikov K, Van der Linden M, Detraux JJ. From facial emotional recognition abilities to emotional attribution: A study in Down syndrome. Res Dev Disabil. 2009;30(5):1007–22.
54. Berio L, Newen A. I expect you to be happy, so I see you smile: A multidimensional account of emotion attribution. Philos Phenomenol Res. 2025;110(2):552–75.
55. Haas BW, Anderson IW, Filkowski MM. Interpersonal reactivity and the attribution of emotional reactions. Emotion. 2015;15(3):390.
56. Russell JA. A circumplex model of affect. J Personal Soc Psychol. 1980;39(6):1161.
57. Mehrabian A, Russell JA. An approach to environmental psychology. The MIT Press; 1974.
58. Plutchik R. A general psychoevolutionary theory of emotion. Theories of emotion. Academic; 1980. pp. 3–33.
59. Ekman P. An argument for basic emotions. Cognition Emot. 1992;6(3–4):169–200.
60. Ekman P, Friesen WV. Facial action coding system. Environmental Psychology & Nonverbal Behavior; 1978.
61. Ekman P, Rosenberg EL, editors. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press; 1997.
62. Mackie DM, Smith ER, Ray DG. Intergroup emotions and intergroup relations. Soc Pers Psychol Compass. 2008;2(5):1866–80.
63. Thayer RE. The biopsychology of mood and arousal. Oxford University Press; 1990.
64. Scherer KR. Appraisal considered as a process of multilevel sequential checking. Appraisal processes emotion: Theory methods Res. 2001;92(120):57.

65. Moors A, Ellsworth PC, Scherer KR, Frijda NH. Appraisal theories of emotion: State of the art and future development. Emot Rev. 2013;5(2):119–24.
66. Shamay-Tsoory SG, Aharon-Peretz J, Perry D. Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain. 2009;132(3):617–27.
67. Koban K, Banks J. It feels, therefore it is: Associations between mind perception and mind ascription for social robots. Comput Hum Behav. 2024;153:108098.
68. Swiderska A, Küster D. (2020). Robots as malevolent moral agents: Harmful behavior results in dehumanization, not anthropomorphism. Cogn Sci, 44(7), e12872.
69. Klüber K, Onnasch L. Beyond the monotonic: Enhancing human-robot interaction through affective communication. Computers Hum Behavior: Artif Hum. 2025;3:100131.
70. Mirnig N, Strasser E, Weiss A, Kühnlenz B, Wollherr D, Tscheligi M. Can you read my face? A methodological variation for assessing facial expressions of robotic heads. Int J Social Robot. 2015;7(1):63–76.
71. Angel-Fernandez JM, Bonarini A. Robots Showing Emotions: Emotion Representation with no bio-inspired body. Interact Stud. 2016;17(3):408–37.
72. Letheren K, Jetten J, Roberts J, Donovan J. Robots should be seen and not heard… sometimes: Anthropomorphism and AI service robot interactions. Psychol Mark. 2021;38(12):2393–406.
73. Li M, Guo F, Ren Z, Duffy VG. A visual and neural evaluation of the affective impression on humanoid robot appearances in free viewing. Int J Ind Ergon. 2022;88:103159.
74. Pütten R-Vder, Krämer AM, Maderwald NC, Brand S, M., Grabenhorst F. Neural mechanisms for accepting and rejecting artificial social partners in the uncanny valley. J Neurosci. 2019;39(33):6555–70.
75. Wu J, Du X, Liu Y, Tang W, Xue C. How the degree of anthropomorphism of human-like robots affects users' perceptual and emotional processing: Evidence from an EEG study. Sensors. 2024;24(15):4809.
76. Yagi S, Nakata Y, Nakamura Y, Ishiguro H. Perception of emotional expression of mobile humanoid robot using gait-induced upper body motion. IEEE Access. 2021;9:124793–804.
77. Mattiassi AD, Sarrica M, Cavallo F, Fortunati L. What do humans feel with mistreated humans, animals, robots, and objects? Exploring the role of cognitive empathy; 2021.
78. Xie Y, Zhu K, Zhou P, Liang C. How does anthropomorphism improve human-AI interaction satisfaction: a dual-path model. Comput Hum Behav. 2023;148:107878.
79. Yang H, Xu H, Zhang Y, Liang Y, Lyu T. Exploring the effect of humor in robot failure. Annals Tourism Res. 2022;95:103425.
80. Gaudiello I, Zibetti E, Lefort S, Chetouani M, Ivaldi S. Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. Comput Hum Behav. 2016;61:633–55.
81. Pinney J, Carroll F, Newbury P. Human-robot interaction: the impact of robotic aesthetics on anticipated human trust. PeerJ Comput Sci. 2022;8:e837.
82. Herdel V, Kuzminykh A, Parmet Y, Cauchard JR. Anthropomorphism and affective perception: Dimensions, measurements, and interdependencies in aerial robotics. IEEE Trans Affect Comput. 2024;15(3):1463–74.
83. Bartneck C, Bleeker T, Bun J, Fens P, Riet L. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. Paladyn. 2010;1(2):109–15.
84. Binesh FN, Rippee R, Belarmino A, Syah A. Unlocking the casino of tomorrow: Analyzing human–robot interaction, trust, psychological, and utilitarian factors. Int J Social Robot. 2025;17(1):99–115.
85. Reeves B, Nass C. The media equation: How people treat computers, television, and new media like real people. Camb UK. 1996;10(10):19–36.
86. Gambino A, Fox J, Ratan RA. Building a stronger CASA: Extending the computers are social actors paradigm. Human-Machine Communication. 2020;1:71–85.
87. Xu K, Chen X, Huang L. Deep mind in social responses to technologies: A new approach to explaining the computers are social actors phenomena. Comput Hum Behav. 2022;134:107321.
88. Haslam N. Dehumanization: An integrative review. Personality social Psychol Rev. 2006;10(3):252–64.
89. Hofree G, Ruvolo P, Reinert A, Bartlett MS, Winkielman P. Behind the robot's smiles and frowns: In social context, people do not mirror android's expressions but react to their informational value. Front Neurorobotics. 2018;12:14.
90. Wieringa MS, Müller BC, Bijlstra G, Bosse T. Robots are both anthropomorphized and dehumanized when harmed intentionally. Commun Psychol. 2024;2(1):72.
91. Crowell CR, Deska JC, Villano M, Zenk J, Roddy JT Jr. (2019). Anthropomorphism of robots: study of appearance and agency. JMIR Hum factors, 6(2), e12629.
92. Küster D, Swiderska A. Seeing the mind of robots: Harm augments mind perception but benevolent intentions reduce dehumanisation of artificial entities in visual vignettes. Int J Psychol. 2021;56(3):454–65.
93. Song Y, Tao D, Luximon Y. In robot we trust? The effect of emotional expressions and contextual cues on anthropomorphic trustworthiness. Appl Ergon. 2023;109:103967.
94. Liu X, Dong J, Jeon M. Robots'woohoo and argh can enhance users' emotional and social perceptions: An exploratory study on non-lexical vocalizations and non-linguistic sounds. ACM Trans Human-Robot Interact. 2023;12(4):1–20.

95. Perugia G, Paetzel-Prüsmann M, Hupont I, Varni G, Chetouani M, Peters CE, Castellano G. Does the goal matter? emotion recognition tasks can change the social value of facial mimicry towards artificial agents. Front Rob AI. 2021;8:699090.
96. Strait MK, Floerke VA, Ju W, Maddox K, Remedios JD, Jung MF, Urry HL. Understanding the uncanny: Both atypical features and category ambiguity provoke aversion toward humanlike robots. Front Psychol. 2017;8:1366.
97. Jung S, Ahn SH, Ha J, Bahn S. A study on the effectiveness of IT application education for older adults by interaction method of humanoid robots. Int J Environ Res Public Health. 2022;19(17):10988.
98. Fu C, Alexis M, Yoshikawa Y, Ishiguro H. Enhancing the mobile humanoid robot's emotional expression with affective vertical-oscillations. Int J Social Robot. 2024;16(7):1523–40.
99. Sasser JA, McConnell DS, Smither JA. Investigation of relationships between embodiment perceptions and perceived social presence in human–robot interactions. Int J Social Robot. 2024;16(8):1735–50.
100. Dubal S, Foucher A, Jouvent R, Nadel J. Human brain spots emotion in non humanoid robots. Soc Cognit Affect Neurosci. 2011;6(1):90–7.
101. Paetzel-Prüsmann M, Perugia G, Castellano G. The influence of robot personality on the development of uncanny feelings. Comput Hum Behav. 2021;120:106756.
102. Förster F, Althoefer K. Attribution of autonomy and its role in robotic language acquisition. AI Soc. 2022;37(2):605–17.
103. Trovato G, Kishi T, Endo N, Zecca M, Hashimoto K, Takanishi A. Cross-cultural perspectives on emotion expressive humanoid robotic head: recognition of facial expressions and symbols. Int J Social Robot. 2013;5(4):515–27.
104. Ko S, Barnes J, Dong J, Park CH, Howard A, Jeon M. The effects of robot voices and appearances on users' emotion recognition and subjective perception. Int J Humanoid Rob. 2023;20(01):2350001.
105. Erden MS. Emotional postures for the humanoid-robot nao. Int J Social Robot. 2013;5(4):441–56.
106. Zinina A, Zaidelman L, Arinkin N, Kotov A. Non-verbal behavior of the robot companion: a contribution to the likeability. Procedia Comput Sci. 2020;169:800–6.
107. Sial SB, Sial MB, Ayaz Y, Shah SIA, Zivanovic A. Interaction of robot with humans by communicating simulated emotional states through expressive movements. Intel Serv Robot. 2016;9(3):231–55.
108. Delgosha MS, Hajiheydari N. How human users engage with consumer robots? A dual model of psychological ownership and trust to explain post-adoption behaviours. Comput Hum Behav. 2021;117:106660.
109. Fraune MR, Oisted BC, Sembrowski CE, Gates KA, Krupp MM, Šabanović S. Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact. Comput Hum Behav. 2020;105:106220.
110. Kang B, Jun S. User perception of social robot's emotional expression and forms of anthropomorphism. Archives Des Res. 2022;35(2):137–53.
111. Yamaguchi M. Item-level implicit affective measures reveal the uncanny valley of robot faces. Int J Hum Comput Stud. 2025;196:103443.
112. Chiang AH, Trimi S, Lo YJ. Emotion and service quality of anthropomorphic robots. Technol Forecast Soc Chang. 2022;177:121550.
113. Nijssen SR, Müller BC, Bosse T, Paulus M. You, robot? The role of anthropomorphic emotion attributions in children's sharing with a robot. Int J Child-Comput Interact. 2021;30:100319.
114. Vetter P, Badde S, Ferrè ER, Seubert J, Shinn-Cunningham B. Evaluating cognitive penetrability of perception across the senses. Nat Reviews Psychol. 2024;3(12):804–20.
115. Barrett LF. How emotions are made: The secret life of the brain. Pan Macmillan; 2017.
116. Roesler E, Manzey D, Onnasch L. A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. Sci Rob. 2021;6(58):eabj5425.
117. Ortony A, Clore GL, Collins A. The cognitive structure of emotions. Cambridge University Press; 2022.
118. Corneille O, Gawronski B. Self-reports are better measurement instruments than implicit measures. Nat Reviews Psychol. 2024;3(12):835–46.
119 Alimardani M, de Roode R, Vaitonyte J, Louwerse MM. (2024, September). Effect of a Virtual Agent's Appearance and Voice on Uncanny Valley and Trust in Human-Agent Collaboration. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents* (pp. 1–7).
120 Becker D, Rueda D, Beese F, Torres BSG, Lafdili M, Ahrens K, Wermter S. (2023, August). The emotional dilemma: influence of a human-like robot on trust and cooperation. In *2023 32nd IEEE international conference on robot and human interactive communication (RO-MAN)* (pp. 1689–1696). IEEE.
121 Breazeal C. (2001, October). Emotive qualities in robot speech. In *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)* (Vol. 3, pp. 1388–1394). IEEE.
122 Darling K, Nandy P, Breazeal C. (2015, August). Empathic concern and the effect of stories in human-robot interaction. In 2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN) (pp. 770–775). IEEE.

123 Demenescu LR, Kortekaas R, den Boer JA, Aleman A. (2010). Impaired attribution of emotion to facial expressions in anxiety and major depression. PLoS ONE, 5(12), e15058.
124 Destephe M, Brandao M, Kishi T, Zecca M, Hashimoto K, Takanishi A. (2014, August). Emotional gait: Effects on humans' perception of humanoid robots. In *The 23rd IEEE international symposium on robot and human interactive communication* (pp. 261–266). IEEE.
125 Eyssel F, Hegel F, Horstmann G, Wagner C. (2010, September). Anthropomorphic inferences from emotional nonverbal cues: A case study. In *19th international symposium in robot and human interactive communication* (pp. 646–651). IEEE.
126 Hoffman G, Breazeal C. (2004, September). Collaboration in human-robot teams. In *AIAA 1st intelligent systems technical conference* (p. 6434).
127 Ito R, de Melo CM, Gratch J, Terada K. (2024, September). Emotional Expression Help Regulate the Appropriate Level of Cooperation with Agents. In *2024 12th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 27–36). IEEE.
128 Kühnlenz B, Kühnlenz K. (2020, August). Social bonding increases unsolicited helpfulness towards a bullied robot. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 833–838). IEEE.
129 Landis JR, Koch GG. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.
130 Natarajan M, Gombolay M. (2020, March). Effects of anthropomorphism and accountability on trust in human robot interaction. In Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction (pp. 33–42).
131 Sato D, Sasagawa M, Niijima A. (2020, August). Affective touch robots with changing textures and movements. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 1–6). IEEE.
132 Takanishi A, Sato K, Segawa K, Takanobu H, Miwa H. (2000, April). An anthropomorphic head-eye robot expressing emotions based on equations of emotion. In *Proceedings 2000 ICRA. Millennium conference. IEEE international conference on robotics and automation. Symposia proceedings (cat. No. 00CH37065)* (Vol. 3, pp. 2243–2249). IEEE.
133 Tsiourti C, Weiss A, Wac K, Vincze M. (2017, October). Designing emotionally expressive robots: A comparative study on the perception of communication modalities. In *Proceedings of the 5th international conference on human agent interaction* (pp. 213–222).
134 Wellman HM, Carey S, Gleitman L, Newport EL, Spelke ES. The child's theory of mind. The MIT Press; 1990.
135 Wiltgen B, Beer JM, McGreggor K, Jiang K, Thomaz A. (2010, September). The interplay of context and emotion for non-anthropomorphic robots. In *19th International Symposium in Robot and Human Interactive Communication* (pp. 658–663). IEEE.
136 Xie Q, Luh DB. (2023). The Holistic Expression Factors of Emotional Motion for Non-humanoid Robots. Int J Adv Comput Sci Appl, *14*(10).
137 Zhang X, Mollahosseini A, Boucher E, Voyles RM, Nielsen R, Mahoor MH. (2014, August). ebear: An expressive bear-like robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (pp. 969–974). IEEE.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Cecilia Roselli[1] · Ivano Zanzarella[2] · Pasquale Grieco[2] · Edoardo Datteri[1]

✉ Cecilia Roselli
cecilia.roselli@unimib.it

[1] University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, Milan 20126, Italy

[2] University of Bari Aldo Moro, Piazza Umberto I, n. 1, Bari 70121, Italy