Modelling User Preference for Embodied Artificial Intelligence and Appearance in Realistic Humanoid Robots

Carl Strathearn [1] and Minhua Ma [2]

[1] School of Computing and Digital Technologies, Staffordshire University, UK
[2] Falmouth University, UK. M.Ma@falmouth.ac.uk

Abstract

Realistic humanoid robots (RHRs) with embodied artificial intelligence (EAI) have numerous applications in society as the human face is the most natural interface for communication and the human body the most effective form for traversing the manmade areas of the planet. Thus, developing RHRs with high degrees of human-likeness provides a life-like vessel for humans to physically and naturally interact with technology in a manner insurmountable to any other form of non-biological human emulation. This study outlines a human-robot interaction (HRI) experiment employing two automated RHRs with a different appearance and personality. The sample group of 20 individuals categorises into age and gender for a diverse statistical analysis. Galvanic skin response, facial expression analysis and AI analytics permitted cross-analysis of biometric and AI data with participant testimonies to reify results. This study concludes that younger subjects preferred HRI with a younger-looking RHR and the more senior age group with an older looking RHR. Moreover, the female test group preferred HRI with an RHR with a younger appearance and male subjects with an older looking RHR. This research is useful for modelling the appearance and personality of RHRs with EAI for specific jobs such as care for the elderly and social companions for the young, isolated and vulnerable.

Key Words: *Embodied Artificial Intelligence, Realistic Humanoid Robots, Human-Robot Interaction, Human-Computer Interaction.*

1. Introduction

Numerous scholars suggest that emotionally responsive AI (EmoAI) in HRI reduces negative perceptual feedback as people feel an affinity towards RHRs that can simulate empathy [1], [2], [3], [4], [5] [6], [7] & [8]. The EmoAI approach is founded on behaviours in human sociology as communication, personality and comprehension help promote understanding and empathy during human-human interaction [9], [10], [11] & [12]. Thus, people empathise more with RHRs than non- anthropomorphic robots as humans feel an innate association with machines that look human, due to the innate psychological drive to socialise and form relationships with other humans [13].

In support, the RHR 'Sophia's' ability to discuss its visual and functional limitations with humans helps people empathise with the RHRs preternatural qualities [14] & [15]. This design consideration is significant to the progression of RHRs as research in robotic AI has predominantly focused on simulating human cognition and continually neglects the significance of EmoAI in promoting natural HRI, which heightens the potentiality of adverse feedback due to emotionless robotic AI [16]. Masahiro Mori's [17] uncanny valley (UV) hypothesis accounts for the negative psychological stimulus propagated by RHRs upon observation, as the more human-like RHRs appear, the higher the potential for humans to feel repulsed by their appearance.

Quality aesthetics is a foundation for reducing the UV in RHR design [18] & [19]. However, the aesthetics approach is a viable first stage for increasing the authenticity of an RHRs appearance but neglects the importance of naturalistic functionality and movement. Per the UV, poor functionality supersedes good aesthetical design in the same way that quality aesthetics become secondary to inadequate functionality [20] & [21].

Therefore, consideration of both appearance and functionality is crucial in developing greater human-like RHRs and reducing the UV [22]. However, cultural background influences the UV as observed in many studies conducted by Eastern scholars which often rate lower levels of the UV than in Western cultures [23], [24], [25], [26], [27], [28]. Furthermore, many scholars argue that children are less susceptible to the UV as they are naturally more curious and accepting of RHRs than adults due to a lack of media influence and risk perception [29], [30], [31], [32], [33]. Thus, exposing children to RHRs at a young age is a methodology for ethically reducing the UV as it builds a foundation of understanding before media influence [34] & [35].

Following these findings, developing emotionally responsive RHRs with higher degrees of visual and functional human-likeness has the potentiality to increase affinity and reduce the UV, which may prove essential in enculturating RHRs into society and developing RHRs with higher degrees of human-likeness. Nevertheless, little practical research evaluating user preference and the influence of personality and appearance in automated RHRs with EAI exists in HRI. Therefore, this study exposes and explores a significant gap in current HRI, EAI and RHR research and critically investigates this area by outlining the processes utilised in developing an EAI personality for RHRs, substantiated by the results of an HRI experiment measuring the influence and significance of emulating the human appearance and personality in HRI.

2. Embodied Artificial Intelligence and Emotional Artificial Intelligence

Until recently, the primary focus of AI research was on creating algorithms that can compute data that humans are incapable of calculating [36]. However, a shift in AI research and development towards emotionally responsive systems with human-like personalities has become a critical factor in developing AI to interact naturally with humans [37], [38], [39]. Thus, EmoAI applications are better suited for situations that involve human communication and sensitivity compared to traditional forms of robotic AI [40], [41], [42]. EmoAI systems such as Cogito monitor emotional cues in the user's speech during telephone conversations to respond to users naturally and make HCI more engaging [43]. Comparatively, an EmoAI system named 'Surrportiv' implements natural language processing to converse with users. It adjusts the tonality and empathy of the speech synthesis output to help tackle situations such as user stress and anxiety [44].

A similar EmoAI support system named 'Butterfly' utilised in business environments monitors staff stress levels, collaboration, creativity and confliction between co-workers. The system offers help and guidance to negotiate stressful work situations by sending employees supporting text messages [45]. However, restricting the interactive capabilities of an EAI in HCI to a single stimulus such as speech, or text neglects many fundamental communicative processes observed in natural human-human interaction such as FEs, attention, gesturing and eye contact [46]. Thus, embodying modes of EmoAI in RHRs has the potential to augment HRI and promote greater naturalistic HRI by emulating human emotions.

In support, the humanoid robot 'Pepper' recognises and responds to six basic human emotions; joy, sadness, anger, surprise, disgust and fear, using an emotion detection camera system integrated into its AI [47]. The worldwide success of the Pepper robot relies on the system's ability to communicate and respond to humans with empathy making the robot feel more approachable and friendly during HRI [48], [49] & [50]. However, Pepper's appearance is distinctly unnatural with a plastic shell for skin and immobile eyes and mouth aperture, which reduces the robot's ability to display human-like emotions, natural skin aesthetics and FEs visually [51], [52], [53] & [54]. Thus, replicating naturalistic aesthetics and sensor functionality of the human skin, as outlined in a study testing an artificial skin for tactile interfacing in HRI which emulates surface changes to touch and temperature to propagate synthetic goosebumps and hair raising [55].

Comparatively, a synthetic skin which systematically responds to physical manipulation such as pinching and pulling can determine the emotional state of the user. Unlike the previous examples, this prototype has a compatible mobile application which reports on user stress [56]. However, the goal of the robotic skin is for use in HRI to explore the intersection between man and machine with future adaptations to include skin conductance, embedded hair and temperature recognition features.

Although these prototypes do not currently function with AI, the possibility of configuring the artificial skins with EAI to elicit responses such as changes in speech tonality, pupil dilation, behaviour, temperature and perspiration is of significant value to RHR design. This approach is essential as it combines a wide range of physical and psychological human responses into one system, proximal to the nervous system. As highlighted in a report on an RHR named H1 with tactile skin sensors for enhancing HRI by responding to touch with EmoAI responses [57]. Moreover, developing an EAI personality depends highly on the functionality, sensory abilities, intellectual capacity and application of the system [58], [59]. For instance, chatbot personalities for emergency and rescue situations should be empathetic but also intellectually efficient in providing quick, logical and concise instructions [60].

In regard, Gardner's theory of multiple intelligences suggests that humans have different kinds of intelligence [61]. These include visual-spatial intelligence: people who are good at visual and spatial judgement. Linguistic-verbal intelligence: people with active reading and writing skills. Logical-mathematical intelligence: individuals who excel at reasoning, recognising patterns and analysing formulas. Bodily-kinaesthetic intelligence: people with advanced hand-eye coordination and dexterity. Musical intelligence: individuals who think in patterns, rhythms and sounds. Interpersonal intelligence: people who are good at understanding, relating and interacting with other people. Intrapersonal intelligence: individuals who can control their emotional states; and naturalistic intelligence: people who are in tune with nature.

However, current modes of robotic AI primarily focus on logical, linguistical and kinaesthetic intelligence and neglect the implications of interpersonal, musical, naturalistic and intrapersonal intelligence [62]. Thus, interpersonal and intrapersonal intelligence is vital in AI development as these human drives consider how people communicate, control emotions, understand and empathise with others [63], shown in Fig. 1. The application of Gardener's multiple-intelligence approach in ML, DL, statistics, database design and EAI systems is vital in creating multi-tasking robots with human-like capabilities [64], [65], [66], [67].
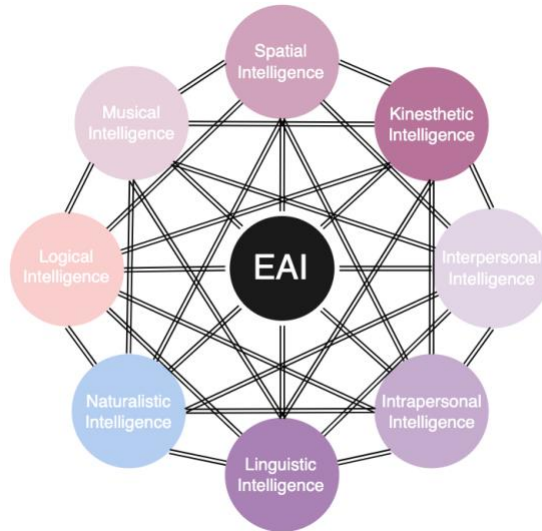
Fig. 1: Gardner's eight intelligence types combinations and application in EAI

Thus, intelligence and personality are intrinsically interlinked, and personality type is a highly influential factor in intellectual capacity [68]. Gardner's multiple intelligence approach in EAI has a significant impact on the intellectual authenticity of RHRs as human beings have a high level of intuition when examining and testing the abilities of AI systems [69]. Furthermore, EmoAI developed for RHRs is crucial in the effective and naturalistic integration of RHRs into human society [70] and [63]. This proposition is significant as, market predictions indicate a substantial rise in the manufacturing, sales and development of humanoid robots with EAI over the next five years [71].

2.1 Practical Applications of Humanoid Robots with EAI

Emulating the human body and mind is the most challenging and rewarding endeavours in science, engineering and technology as the human being is the highest functioning organism in the known universe [72] (Zlotowski et al., 2013). Although the future applications of RHRs with EAI are highly theoretical, emotionally intelligent robots with human-like personalities are gradually emerging in fields such as; Vyommitra, 2020: IND space exploration personal assistant. HAL, 2018: USA, for use in paediatric training. Jiang Lilai, 2019: China & Alex, 2019: RUS & Junko Chihira, 2016 JAP are 24 hr robotic news reporters, Sophia, 2016: USA, social researcher and conference speaker. AI-DA, 2019: UK, a robotic artist. BINA 48, 2016: USA, a robotic lecturer, Telinoid, 2006: JAP, a healthcare assistant for the elderly. CB2, 2006: JAP, a robotic child to train young parents for adulthood. Diego San, 2010, USA, an RHR to study cognitive development in children. Affecto, 2010: JAP, a child robot to study human sociology. Robocop, 2017: DXB, a robotic security guard. Furhat, 2018: SWD, a robotic personal assistant.

Thus, unlike traditional models of design and programming in robotics, the end goal of RHRs with EAI is to perform multiple tasks in different environments, to emulate the natural cognitive, emotional, communicative and physical capabilities of humans [73]. For example, industrial robots have the potentiality to generate feelings of anxiety as they cannot provide the emotional support or understanding of human co-workers [74], [75] & [76]. Therefore, a shift towards emotional industrial robots such as 'Sawyer' and 'Baxter' that emulate human FEs are reducing the anxiety between workers and machines in factory environments [77] Shell (2018). These ethical factors are crucial in the sociological integration of humans and RHRs as according [78], [79]. Assimilation is a critical aspect in human-robot integration as the fewer distinctions between humans and machines, the higher the potential for acceptance [80] & [81].

2.3 AI and Natural Language Processing System Design

Baudi and Euclid implement Amazon Lex DL AI and Amazon Polly speech synthesis software to converse with people naturally. The personalities and interests of the RHRs reflect their age to measure how participants respond to different personality types. Thus, Baudi's interests are music, food and travel, and Euclid's interests are in art, poetry and literature. The DL algorithm in amazon Lex permits operators to access Amazon's large cloud-based databank of information relating to these topics, depicted in Fig. 2. The conversational AI systems function using a series of intents: which are actions the user wants to perform. Sample utterances: examples of things people may say to convey the intent. Slots and containers: which are data banks of information relating to specific words and topics and fulfilments: providing an appropriate answer/s using the information provided by the user.

Moreover, Amazon Lex is a commercial product designed for business, and the UI is mostly inaccessible compared to other open-source AI applications. This configuration makes developing conversational AI for general-purpose communication problematic as the slots defined by the program are mainly for specific types of business such as restaurants, music shops and book shops. Therefore, mapping together these functions to create a personality depends on selecting related slots and instances, linking them together and building a personality type and interests around those features. Furthermore, Amazon Lex limits the maximum amount of intents to one hundred questions per chatbot. This limitation is also problematic as it reduces the scope of potential questioning and has a high potentiality to produce incorrect or repetitive responses. However, this issue is negotiable by increasing the number of slots, sample utterances and fulfilments per inquiry.

Moreover, configuring a separate chatbot for different elements of the robot's personality, emotions, and interests provide a broader range of interaction during HRI. This approach is significant to designing an EAI personality as Amazon Lex is restrictive due to its closed-software status and UI design. However, the quality and accuracy of the DL AI system are highly effective for developing conversational AI for HRI. The Amazon Lex conversational AI system is compatible with Amazon Rekognition, ML robotic vision software to enhance the sensory capabilities of the EAI system for RHR design. However, Amazon Rekognition is expensive and difficult to calibrate and outside of the scope of resources for this study.
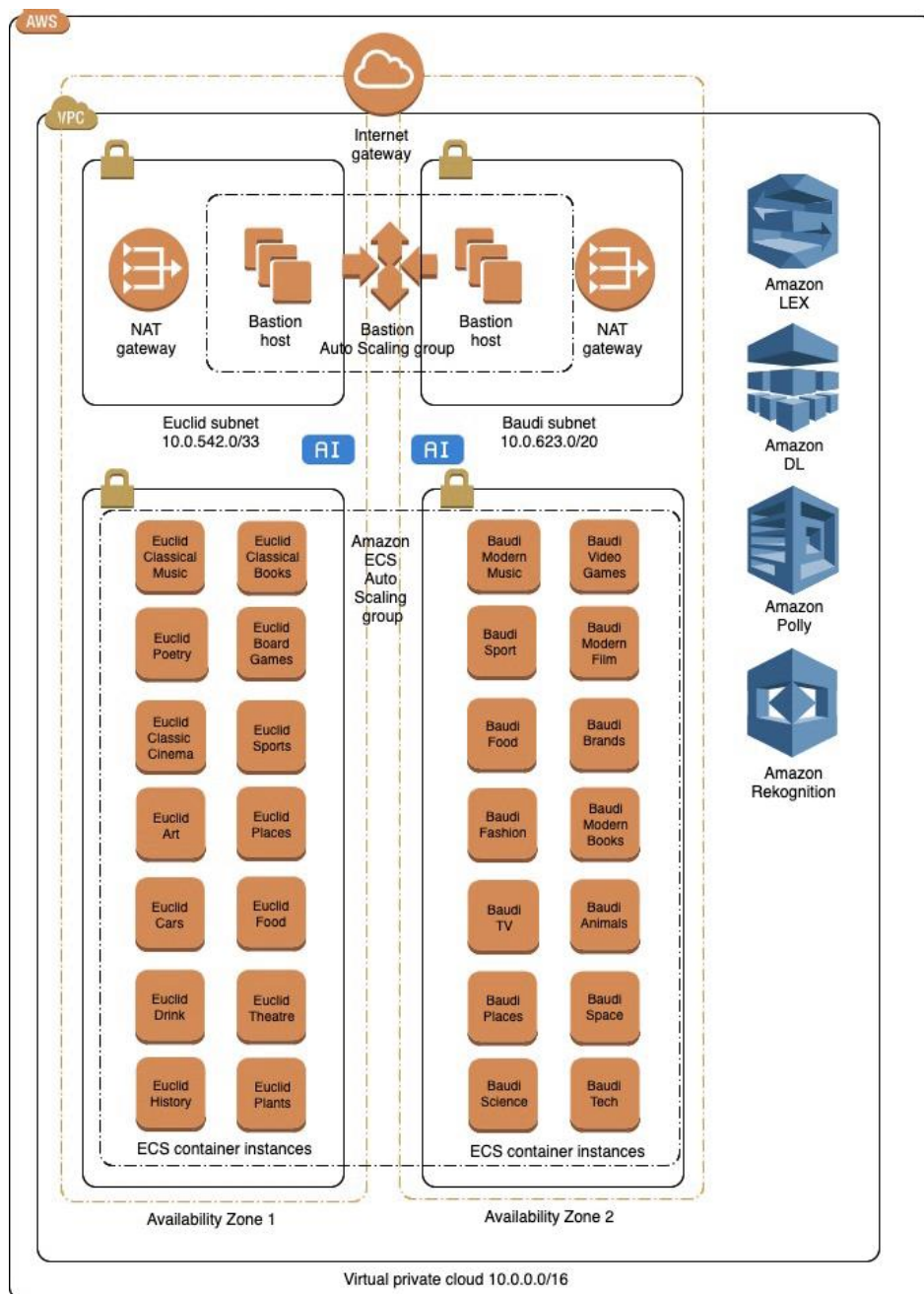
Fig. 2: Euclid and Baudi Interests and Hobbies Cloud Data

3. Human-Robot Interaction Experiment

The objective of the HRI experiment is to measure how participants perceive the authenticity, appearance, likeability and personality of the RHRs. This research is significant as there is a lack of up-to-date practical study into the UV in HRI implementing automated RHRs. The experiment starts with the participant seated in front of a raised table between two RHRs named Euclid and Baudi, shown in Fig. 3, and two computer screens.

A Galvanic skin response GSR electrode is attached to the middle and index finger of the right hand of test subjects to measure skin conductance. The Affdexme facial expression analysis (FEA) system in the eye of Baudi and a webcam in situ to Euclid can detect and track the test subjects facial expression (FEs) and rate of attention. A second camera system positioned to record the test procedure from a wide-angle is set to monitor the test environment for post-experiment analysis.

Fig. 3: RHRs with EAI 'Euclid' and 'Baudi' implemented in the HRI experiment

The RHRs activate, and the heads and eyes automatically turn towards the participants using the face tracking system. The Pololu microcontrollers activate, and the RHRs start to blink, raise eyebrows and make subtle cheek movements. The participants press the start button on the handheld controllers and say 'Hello' to initiate the AI system. The first stage of the practical HRI experiment divides into four 5-minute procedures founded on the evaluation time of the 1950 Turing Test, shown in Table. 1.

Experiment	Section. 1	Section .2	Section. 3	Section. 4
Robot	Euclid	Euclid	Baudi	Baudi
Task	Timed 5-minute spoken conversation with the robot (appearances and hobbies)	The participant plays an animal guessing game (Time capped at 5 minutes)	Timed 5-minute spoken conversation with the robot (appearances and hobbies)	The participant plays an animal guessing game (Time capped at 5 minutes)
Evaluation	The correct, incorrect and, repeat AI responses are recorded using in Lex	The correct, incorrect and, repeat AI responses are recorded using Lex	The correct, incorrect and, repeat AI responses are recorded in Lex	The correct, incorrect and, repeat AI responses are recorded using Lex

Table. 1: First Stage Evaluation Protocol

The aim of sections 1 & 3 of the first stage of the HRI experiment is to monitor and assess; how accurately the EAI systems operate and how participants physiologically and psychologically respond to the RHRs using the biometric sensors. The participants speak to the robots using their voice, and when finished, they press a button on a hand-operated controller to notify the AI system that the conversation has finished and there is a short pause to allow the AI program time to respond. Sections 2 & 4 are the gamification elements of the HRI experiment and require the participants to engage with the robots on a much more interactive and engaging level. The game implemented in this study is called the animal guessing game and participants are requested to ask the RHRs a series of questions pertaining to the identity of the animal the RHRs are 'thinking' of, for example; if Euclid is thinking of a Giraffe, the participants may ask the robot questions such as, 'does it have four legs', 'does it live in the sea', 'can it fly' and the robot responds accordingly with yes or no answers until the participant can correctly identify the animal. Throughout the first stages of the experiment, the correct, incorrect responses and 'please repeat' AI responses are recorded in Amazon Lex statistics; this data is then used to improve the robots AI after each interaction.

The second stage of the experiment is a computerised questionnaire in which participants discuss the HRI experience, personal preference and evaluate the appearance, movement, AI and speech of the robots. During the questionnaire stage, the robots continue to track the head movements of the participant, perform randomised FEs using the Pololu and speak to participants if requested to aid in answering the questionnaire. This novel interactive evaluation approach permits test subjects to engage with the RHRs while filling in the survey, compared to detaching the participant survey evaluation from the HRI test. Numerous studies across HRI implement mixed methodological approaches to help discern trends in large data sets and support and question participant taxonomies [82], [83], [3], [84]. In accordance, This study employs a mixed methodology approach to data analysis, using both quantitative and qualitative methods.

This study implements a GSR system and two camera-based FEA applications to measure the participant's emotional responsivity on a fluctuating scale during the HRI examination which exports into graphs and data fields (.CSV .TXT and .PDF). This data is quantitative and reinforced using qualitative inquiries of the HRI survey to reify outcomes. The results of quantitative dichotomous questions are statistically analysed by employing Cronbach's Alpha to examine internal consistency were lower than .60 is unacceptable, between .60 and .70 is undesirable, between .70 and .80 is respectable, and >.90 is excellent [85]. Standard deviation and the margin of error are also factored: 0.0 exact comparison, < 1.20 acceptable, >2.0 unacceptable [86], shown in Appendix 1-5.

3.1 Population Sample, Recruitment and Participant Restrictions

The population sample size for the primary survey is 20 test subjects; based on the 20 human interrogators implemented in Ishiguro's Total Turing Test (TTT) [87] for androids, Harnad's TTT [88] and Harnad & Scherzera's (2008) Robot Turing Test (RTT) [89] and Schweizer's Truly Total Turing Test (TTTT) [90] for RHRs. The population sample size of THT and HTT are incompatible with this study as these approaches do not evaluate appearance, movement, speech or AI. Similarly, the sample size of the most compatible and recent research [2] proved unsuitable for this study as the experiment neglects AI, robotic vision, tracking, speech synthesis and biometric data gathering. Although the population sample for this study is relatively small, the 50-question questionnaire and biometric devices collect large sums of data to yield valid results. Gay's [91] estimative sampling formula indicates the number of invitations required to recruit 20 participants, depicted in Eq. 6.1

(20 rr ÷ 10% er = 200a)

Equation. 1: Gay's Estimative Sampling Method: *Where: rr – response range (replies needed) er – percentile of estimated responses a – audience (amount of invitations)*

Therefore, to maximise this studies impact, recruitment was conducted through an onsite poster campaign, social media (Twitter) and targeted email for students studying science, technology, engineering and mathematical (STEM) disciplines to promote this research to a broader audience. As a condition of the HRI experiment, all participants must have 20/20 vision or corrected vision to evaluate the authenticity of the RHRs accurately and aged over 18 years to satisfy the terms of the ethics agreement. Recruitment is open to all individuals of different backgrounds, gender, cultures, abilities and nationalities. However, the ability to speak fluent English is a condition of this examination for the NLP to function accurately. Per the limitations of the Turing Test, a selective recruitment and screening procedure ensured participants had sufficient background knowledge of the field of robotics and AI for effective analysis in mixed method data research [92]. Applications from students and professionals within robotics and AI and engineering where given preference over applicants in lesser-related areas. A short discussion over e-mail regarding the participants' background, interests and prior experience of robotics and AI was conducted before acceptance.

3.2 Participants Profiles

Per the ethical requirements for approval, all test subjects were over 18 years of age. The total number of experiments conducted was 21; this is due to one participant not completing the HRI test because of a strong regional accent which was unreadable by the NLP system. The subject was made aware of this issue and continued using type-written responses. However, this approach did not generate speech synthesis responses from the RHRs which is vital to the HRI experiment as Baudi's speech synthesis accent is slow and low in tone and Euclid's intonation is higher with a faster speech rate. Thus, as the test subject was unable to evaluate the RHRs speech synthesis during HRI, the results were withdrawn from the dataset.

The number of compleated experiments is 20, equivalent to the sample of the TTT and TTTT; this includes GSR, FEA and a complete HRI questionnaire. Table. 6.5 indicates the background of each test subject, gender, age and previous experience of taking part in HRI research. As there are no similar HRI studies that implement two automated RHRs with conversational AI, all data is extracted, grouped and analysed to a high degree for future research.

The gender ratio for this study is 11 (55%) male and 9 (45%) female, as there are one more male test subjects than female, participant responses and results that are examined by gender group require significant differences in outcomes to consider viable. The age ranges group into two equitable series 18-27 (10: 50%), 28-56 (10: 50%) to provide a comparative analysis between a higher and lower age group. However, as there is a greater diverse age range in the 28-56-year-old group, only outcomes with significant differences in results will be considered viable. The analysis of the population sample indicated that a high percentage of test subjects are of Western ethnicities with English as a native language 18 (90%) compared to other ethnic backgrounds with English as a second language 2 (10%). Therefore, ethnic and cultural diversity is an issue in this data set. However, the study was open to people of all cultures and backgrounds, and no preference was given to individuals from specific cultures or ethnicities. 95% of participants indicated no previous experience in taking part in HRI experiments.

This result is not surprising per the accessibility to RHRs and the narrow recruitment factors of HRI studies. However, one participant (5%) indicated previous experience in HRI within higher education. All participants underwent a background check, as outlined in the Turing Test recruitment procedure. During recruitment, potential participants were asked to disclose their current/previous level of study and background/current experience to establish their expertise in robotics, AI and relevant fields for eligibility onto the HRI experiment.

Thus, 6/20 (30%), 1:PRO, 1:PG, 4:UG, of participants have previous or current experience with the field of robotics and AI; this percentage is high due to the related nature of the study and the targeted online recruitment of AI and robotics students and staff members from the university. 5/20 (30%), 2:UG, 3:PG: of subjects have experience in areas relating to computing and AI; this includes computer programming, application design and AI games programming. 5/20 (25%), 4:PRO, 1:UG; of participants have experience in engineering and electronics with knowledge of robotic system design. 2/20 (10%), 1:PRO, 1:UG of individuals work or study in film and media, with experience in animation and programming.

2/20 (10%), 2:PRO, of subjects, are professionals in areas relating to human psychology and physiology, to which both participants indicated an interest and personal study of AI. 8/20 (40%) of participants are current university students studying undergraduate degrees, 4/20 (20%) of subjects are undertaking postgraduate degrees at either Masters or PhD. 8/20 (40%) of candidates are professionals within their field with significant personal and industrial experience in AI, robotics and related areas, shown in Table. 2.

Subject Number	Undergrad (UG) Post-Grad (PG) Professional (PRO)	Age, Ethnicity and Gender	Background	Previous HRI Experience
1	PG	22, Male, British	Computational AI	No
2	UG	28, Male, British	Robotics and AI	No
3	UG	22, Female, British	Computing and Design	No
4	UG	20, Male, British	Robotics and AI	No
5	PRO	31, Male British	Robotics and AI	Yes
6	PRO	45, Male, British	Engineering and CAD	No
7	UG	29, Female, British	Multimedia	No
8	PG	25, Female, British	Game Design & Programming	No
9	UG	18, Female, British	Film and Media	No
10	UG	20, Male, Asian	Robotics and AI	No
11	UG	21, Male, Asian	Robotics and AI	No
12	PRO	56, Male, British	Engineering	No
13	PG	24, Female, British	Games Programming	No
14	PRO	32, Female, British	Film and Media	No
15	PG	25, Female British	Game Design & Programming	No
16	PRO	55, Male, British	Electronic Engineer	No
17	PRO	43, Female, British	Human Factors & Ergonomics	No
18	PRO	38, Male, British	Electronics	No
19	PRO	32, Female, British	Cognitive Science	No
20	UG	25, Male, British	Engineering	No

Table. 2: HRI Experiment Test Subject Profiles

3.3 Questionaire Categories The RHR and EAI user preference and design questionnaire consist of 10 open-ended questions followed by 15 quantitative questions measured on 5-point Likert scales the cover: likeability, human-likeness and competence shown in Table 3. based on previous studies [93], [94] and [87]. The qualitative and quantitative HRI questionnaire results are comparatively analysed against the GSR, FEA and AI analytic data for a comprehensive cross-analysis of all data fields.

Question 11	Like		1	2	3	4	5	Dislike
Question 12	Incompetent		1	2	3	4	5	Competent
Question 13	Human-like		1	2	3	4	5	Machine-Like
Question 14	Friendly		1	2	3	4	5	Unfriendly
Question 15	Fake		1	2	3	4	5	Natural
Question 16	Responsible		1	2	3	4	5	Irresponsible
Question 17	Unkind		1	2	3	4	5	Kind
Question 18	Aware		1	2	3	4	5	Unaware
Question 19	Moving Naturally		1	2	3	4	5	Moving Rigidly
Question 20	Pleasant		1	2	3	4	5	Unpleasant
Question 21	Unintelligent		1	2	3	4	5	Intelligent
Question 22	Unconscious		1	2	3	4	5	Conscious
Question 23	Awful		1	2	3	4	5	Nice
Question 24	Sensible		1	2	3	4	5	Foolish
Question 25	Artificial		1	2	3	4	5	Lifelike

Table. 3 Quantitative HRI Survey Questions

4. HRI Questionnaire

Question 1. Which robot do you think looks most human-like, Baudi or Euclid?

18/20 (90%) cited Euclid as the more human-like RHR. Of this dataset, 14/18 mentioned that Euclid's skin appeared more human-like, 3/18 conferred Euclid's mouth as the most human-like and 1/18 cited Euclid's eyes as more aesthetically realistic than Baudi's. 1/20 (5%) of results explained that Baudi's hair implants made the RHR look human-like. 1/20 (5%) of test subjects cited Baudi's skin tone as more human-like than Euclid's. A key theme in the results of Q1 is the aesthetical quality of the silicone skins, which many test subjects testified as having a significant impact on the RHRs visual authenticity. Both silicone skins were commissioned from the same artist to minimise visible irregularities between the RHRs and aesthetical and material quality. Euclid's skin wrinkles were a recurring theme in participant responses.

RHRs are typically young in appearance and designed without skin imperfections or blemishes [33], [82], which is not an accurate representation of the natural tonal variance and ageing of human skin. This aesthetical consideration is significant to RHR design as although Baudi's skin is not flawless; the RHR has no skin wrinkles and fewer skin imperfections. Therefore, as Euclid is much older in appearance with a greater range of facial defects, the RHRs appearance further blurs the line between human beings and RHRs. More male (M) subjects cited Euclid as human-Like than females (F), F8, M10 and Baudi: F1, M1. There was little difference between the higher and lower age groups, H= ages (18-27) and L= ages (27-56), Euclid: L9, H9 and Baudi: L1, H1

Question 2. Which robotic voice did you prefer/understand better, Baudi, or Euclid?

13/20 (65%) of test subjects preferred the voice of Baudi. Of that dataset, 8/13 explained that Baudi's speech was slower and easier to understand than Euclid's and 5/13 stipulated that Baudi's accent was easier to understand than Euclid's. 7/20 (35%) cited Euclid's speech synthesis as the most understandable. Of those results, 3/7 cited the pitch and tone of Euclid's voice as the most human-sounding. A key theme in the HRI results considered the American accents of the robots, as the speech synthesis program implemented in this study only outputs American accents. This configuration appeared to impact natural speech understanding of non-American listeners. A recurring theme in the test results was the preference of the American accent as (2/20) stated the American accents were off-putting, and 4/20 suggested that they liked the American accents.

However, 4/20 of test subjects explained that the American accents were difficult to understand at times. A higher number of male test subjects preferred the speech synthesis of Baudi due to the speech rate, depth and tone: M9, F4 compared to Euclid M2, F5. Age-related statistics indicated that the lower age group preferred the speech synthesis accent of Baudi: L9, H4, compared to Euclid: L1, H6.

Question 3. Did you feel empathy/emotion towards Euclid or Baudi?

8/20 (40%) explained that they did not feel empathy or emotions towards the RHRs because the appearance, movement, and AI did not feel entirely human. However, out of that dataset, 3/6 advocated that although they did not feel any emotion, they felt relaxed, happy and calm during HRI with the RHRs, counter to the UV. 5/20 (25%) cited that they felt no emotion or empathy because the robots were just heads without a body which made them feel inhuman.

This outcome is justifiable with the current configurations of the RHRs as time limitations prohibited the design, development and production of a robotic body and implementing a still manakin type body was difficult due to spatial constraints. 5/20 (25%) of test subjects suggested that they felt empathy towards Baudi due to his unnatural appearance, which resembled human disability. 2/5 explained that Baudi's dilating eyes played a role in instigating an emotional response. 2/20 (10%) felt that Euclid's realistic appearance made them feel awkward and intimidated, which coincides with the UV. However, this sample is not sufficient in size to support the existence of the uncanny valley.

A higher number of female subjects felt an emotional response towards the RHRs compared to male subjects F6, M1, and a greater number of males indicated no emotional responsivity F3, M10. Out of that dataset, F5 cited an emotional response when interacting with Baudi and F1/M1 an emotional response during HRI with Euclid. Similarly, there was a notable difference in the age-related statistical analysis. L8, H5 cited no emotional response and L2, H5 suggested an emotional interaction. Out of that dataset, L2, H3 mentioned feeling an emotional response when interacting with Baudi and, L0, H2 with Euclid.

Question 4. How did you feel when Baudi made eye contact with you?

8/20 (40%) of test subjects stated that they felt nothing resembling an emotional response when Baudi made eye contact with them. Of this dataset, 5/8 indicated that they did not feel the same way as when making eye contact with other humans. 3/8 suggested that the eye contact interaction was not consistent to make it feel like the RHR was looking at them naturally.

12/20 (60%) stated that they felt an emotional response when making eye contact with Baudi, 7/12 suggested that it felt as if the robot was looking back at them. Of these results, 3/12 cited Baudi's eye aesthetics as making the eyes appear creepy. 2/12 explained that the pupil dilation function made the eyes look human-like. Gender statistical analysis indicated a higher percentage of female participants felt an emotional response to Baudi's eyes than male test subjects F7, M1 compared to no emotional response F2, M10. However, the results of the age-related statistical analysis indicated little correlation between an emotional response L3, H5 and no emotion L7, H5.

Question 5. How did you feel when Euclid made eye contact with you?

13/20 (65%) suggested that they did not feel anything when Euclid made eye contact with them. 8/13 argued that making eye contact with Euclid was like looking into the glass eyes of a doll or the eyes of a painting. 3/20 suggested that Euclid's eyes were not real enough to instigate an emotional reaction. 2/13 gave no reasoning other than to state that they did not feel anything relating to an emotional response. 7/20 (35%) suggested an emotional response to making eye contact with Euclid. 4/7 cited that they felt like the robot was looking back at them, or another human was looking at them through a camera. 2/7 advocated that Euclid had dead eyes which made them feel unnerved. 1/7 explained that Euclid maintained constant eye contact with them, which made them feel unsettled. Gender statistical analysis suggested no correlation between male and female test subjects, no emotional response: F5, M8 and emotional response: F4. M3. Similarly, the results of the age-related statistical analysis indicated no discernable relationships between emotional responsivity L6, H7 and no emotion L4, H3.

Question 6. Which robot head do you think moved most naturally, Euclid or Baudi?

16/20 (80%) of candidates cited Euclid's head as moving the most naturally. Of this dataset, 9/16 explained that they felt a greater natural interaction when communicating with Euclid compared to Baudi. 4/16 stipulated that Euclid exhibited a more extensive range of FEs.

2/16 mentioned Euclid's mouth as a decisive factor in the RHRs naturalistic movement, and 1/16 of outcomes suggested that Euclid's head moved more naturally. 4/20 (20%) explained that they felt Baudi exhibited a more extensive range of natural movement. 3/4 of test subjects advocated that Baudi exhibited a broader range of FEs. 1/4 suggested that Baudi made consistent eye contact with them.

These outcomes suggest that Euclid functioned with a higher range of movement than Baudi. However, both robots are of the same build and implement the same tracking and face detection applications and operate using a single Kinect system paired between the two robots to minimise interference. Therefore, there should be few distinctions between the functionality of the RHRs apart from the eye and mouth components. Thus, it is probable that the appearance and skin movement of the RHRs influenced how test subjects perceived the movement of the RHRs, as both robots instigate eyebrow raises and cheek raises at random intervals using the Pololu microcontroller. Thus, it is probable that during individual sessions, one robot expressed a more extensive range of these functions due to the randomisation of the FEs. This configuration may account for the differences in emotive HRI as the RHRs function differently with a wide range of variables during each session. These outcomes support the findings of the literature review that suggest poor aesthetical quality supersedes quality functionality and vice versa [20] and [21]. Gender analysis indicated more male test subjects cited Euclid's movement M10, F6 as more naturalistic than Baudi F3, M1. However, age-related statistics inferred little discernible difference between the naturalistic movement of Euclid L8, H8 and Baudi L2, H2.

Question 7. Which robotic head expressed the greatest range of facial expressions?

11/20 (55%) advocated that Euclid performed the broadest range of FEs. 4/11 suggested that Euclid's eyebrow raises, and cheek lifts were prominent, 2/11 explained that Euclid appeared to respond to questions with FEs, 2/11 argued that Euclid's skin wrinkles and skin stretching made FEs appear more realistic and 3/11 stipulated that Euclid portrayed a wider range of emotive FEs than Baudi. 8/20 (40%) suggested that Baudi displayed the broadest range of FEs. 5/8 argued that Baudi appeared to synchronise FEs with AI responses more coherently. 3/8 claimed that Baudi performed the widest range of FEs. 1/20 (5%) of test subjects advocated that they could not distinguish which RHR displayed the broader range of FEs.

These outcomes suggest that natural skin aesthetics plays a significant role in displaying FEs in RHRs. This approach is vital in developing greater human-like RHRs as typical methods in RHR design neglect skin blemishes and wrinkles as these features do not fit in with the paradigm of cutting-edge technologies. Gender statistical analysis suggested that a greater number of male participants thought Euclid expressed the broadest range of FEs, F2, M9, and F6, M2 stated Baudi displayed the most FEs. 1F unclassified. Age-related analysis indicated that more subjects in the higher age group thought Euclid displayed the most FEs - EL:3, EH:8 and BL:6, BH:2, 1L unclassified.

These results correspond with existing literature in FE recognition in human psychology. For example, a comprehensive study measuring facial FE recognition [95], discovered that older adults focus on the mouth area for gauging emotive FEs and younger individuals concentrate on the eyes. This consideration is significant as the robotic mouths are greater articulated and expressive than the robotic eyes, per the results of the literature review (practical vs virtual HRI for evaluating the UV). Thus, 'own age advantage' may account for test subjects in the higher age group recognising more FEs during HRI, as their attention is on the mouths of the RHRs rather than the eyes.

Question 8. Which AI did you prefer communicating with during the conversation?

13/20 (65%) preferred to interact with Euclid's AI. Of this dataset, 4/13 indicated that they felt Euclid had more relatable interests and hobbies than Baudi. 7/13 cited that they felt Euclid gave more accurate and fully formed responses than Baudi's and 2/13 advocated that Euclid's personality was more engaging than Baudi's. 7/20 (35%) preferred Baudi's AI system. 3/7 of that dataset cited that Baudi gave more accurate and full responses, and 2/7 stated that they had more in common with the interests of Baudi. 2/7 explained that they felt Baudi was the more engaging RHR. During HRI, many individuals attempted to deceive the RHRs into giving incorrect responses by asking questions outside the scope of the AI system. For example, one participant asked Baudi, "*How many hairs do you have on your head?*" which produced an incorrect response of "*My eyebrows are made of human hair*". Male test subjects: M8, F5 preferred to converse with Euclid Baudi M3, F4, and the higher age group favoured Euclid's AI and the lower age group preferred Baudi's AI: EL:4, EH:9 and BL:6, BH:1

Question 9. Which AI system did you prefer during the guessing game?

11/20 (55%) preferred Baudi's conversational AI during the guessing game session. 8/11 cited Baudi as the more competent RHR. 2/8 of candidates advocated that Baudi had a broader knowledge base, and 1/8 stated that Baudi was more engaging. 9/20 (45%) cited that they preferred to interact with Euclid during the guessing game session. 5/9 argued that Euclid produced more correct answers, 3/9 stated that Euclid had a wider knowledge base, and 1/9 explained that Euclid was more naturalistic during HRI.

As with the previous test, several subjects asked unusual or irrelevant questions to deceive the AI system. Male subjects preferred to interact with Baudi M8, F3 and females with Euclid: M3, F6. However, there was little notable difference in the age-related factors, Baudi: L5, H6 and Euclid L5, H4.

Question 10. Did you prefer communicating with an older or younger-looking robot?

12/20 (60%) preferred to communicate with Euclid. Of this dataset, 5/12 explained that they felt more comfortable interacting with an older looking robot because they felt trust towards an RHR with an aged appearance. 2/13 suggested they prefer to interact with an older looking RHR as it was less intimidating than a younger model. 3/13 suggested that Euclid reminded them of older relatives, which made HRI less intimidating. 2/13 stated that as they felt Euclid looked more human-like than Baudi, Euclid's highly realistic appearance made the communication more naturalistic. 8/20 (40%) advocated that they prefer to communicate with Baudi, 5/8 explained that Baudi appeared less threatening because his appearance was softer.

2/8 indicated that they felt Baudi was more engaging than Euclid and 1/8 suggested that as they interact more with younger people daily that they felt a greater connection with a young-looking RHR. These results indicate that participant outcomes were closely divided between communication with a younger-looking robot and an older looking RHR. This finding is intriguing as it correlates with the age-related statistics, which infers that the higher age group favoured interaction with Euclid H8, L4, and the lower age group preferred HRI with Baudi H2, L6. This outcome is vital to this study as results suggest that younger test subjects felt a greater connection with a younger-looking RHR, and the older age group felt a greater connection with an older looking RHR. Moreover, although Baudi is designed to look younger than Euclid, neither RHR is assigned a specific age, meaning participants were left to decide how old the robots appeared to them. Interestingly, several subjects interpreted the appearance of the RHRs as threatening or intimidating, depending on how old they looked. This result may relate to how young people perceive older adults and older people see younger individuals in society. Gender factors did not influence decision making, F5, M7 prefer to interact with a robot with an elderly appearance and F4, M4 preferred HRI with a younger-looking robot.

4.1 Comparative analysis of qualitative and quantitative results (All Test Subjects)

Series. 1A. Likeability: (Q11, Q14, Q17, Q20 & Q23). On average Baudi rated 4/5 more likeable than Euclid 1/5. Internal consistency was high in this subset, 0.8 α indicating high levels of consistency in participant responses. These results confirm the outcomes of Q3, Q6-Q7 in which candidates cited a more significant emotional response to Baudi during HRI, However, these results contrast with the results of Q10 were test subjects preferred to interact with an older looking RHR (Euclid) over a younger-looking RHR (Baudi). These outcomes collectively conclude that participants found Baudi more likeable than Euclid.

Series. 2A. Human-likeness: (Q13, Q25, Q18, Q21 & Q15). Test subjects rated Euclid 4/5 more human-like than Baudi 1/5. Coefficient factors were high in the subset 0.9 α, suggesting a high level of accuracy in participant outcomes. However, 1/5 (Q15) of test subject outcomes were of an equivalent rating on the 5-point Likert scale (3). These results validate the findings of Q1 of the HRI questionnaire, where test subjects cited Euclid as appearing, functioning and moving more human-like than Baudi. Therefore, these outcomes collectively conclude that test subjects found Euclid more human-like than Baudi.

Series. 3A. Competency: (Q12, Q16, Q19, Q22 & Q24). Test subjects rated Euclid 3/5 as more competent than Baudi (2/5) However, out of these results 2/3 (Q19 & Q24) rated equivalent (3) on the 5-point Likert scale. Internal consistency was acceptable 0.7 α, suggesting some variance in the outcomes of participants responses. These results collate with the findings of Q6, Q7 & Q8, which cite Euclid as moving and responding to participants more competently than Baudi. However, these results contrast with the outcomes of Q2, and Q13, which suggested Baudi's AI functioned with a higher degree of competency during the guessing game session and clarity in speech synthesis during verbal HRI. These results collectively confirm that participants rated Euclid as the more competent RHR during HRI.

4.2 Comparative analysis of (Q1-14) and (Q21-35) results (Male & Female)

Series. 1B. Likeability: (Q11, Q14, Q17, Q20 & Q23). Male subjects liked Euclid 3/5 more than female subjects 2/5, and 3/5 (Q14 & Q20) equivalent ratings of (3). Female subjects rated Baudi as more likeable 3/5 than male subjects 1/5, coefficient factors were high in this subset 0.8 α, suggesting high consistency in the results of Series. 1B. These results correspond with the results of Q3 as more female test subjects cited feeling an emotional response to Baudi: F5/M0, and Q4-Q5 were female participants stated feeling emotional during eye contact interfacing with Baudi: F7, M1. These results collectively conclude that male subjects preferred Euclid and female test subjects favoured Baudi.

Series. 2B. Human-likeness: (Q13, Q25, Q18, Q21 & Q15). Group analysis of series. 2 indicates that both gender groups rated Euclid as more human-like 4/5 and females liked Baudi more than male subjects M0/5, F1/5. Internal consistency was high 0.9 α in the collective results of series 2. These outcomes support and validate the findings of Q1, were both gender groups rated Euclid as more human-like: Euclid: F8, M10 than Baudi: F1, M1. Q5, were both gender groups preferred Euclid's eye aesthetics: Euclid: 6F, 9M and Baudi: 2F, 2M. Q6, were both gender groups cited Euclid's eyes as moving the more human-like: Euclid: 5/F, 7M.

Series. 3B. Competency: (Q12, Q16, Q19, Q22 & Q24). Male test subjects rated Euclid as more competent: (E.M:2/5, E.F:1/5) & (B.M:1/5, B.F:1/5) 2/5 of ratings were equivalent [3] Q16 & Q24. Internal consistency was acceptable 0.7 α, suggesting some level of variance in participant results. These results correlate Q7 In which male test subjects cited Euclid as displaying a broader range of FEs than Baudi: E: F2, M9, and B: F6, M2. Q8 as male test subjects achieved higher accuracy in Euclid's 5-minute topical conversation AI test, ME:146 (51%), FE:136 (49%) and MB:88 (56%), FB:65 (44%). Q9, as male participants achieved higher accuracy scores in the 5-minute guessing game AI test. However, these results contrast with the results of Q2, were more males cited Baudi's speech synthesis as more understandable than Euclid, (B: M9, F4 and E: M2, F5). These outcomes are difficult to collectively verify due to the level of variance in results of the gender statistical analysis.

4.3 Comparative analysis of (Q1-14) and (Q21-35) results (Age Groups)

Series. 1C. Likeability: (Q11, Q14, Q17, Q20 & Q23). The higher age group (18-27) liked Euclid: EL1/5, EH3/5 and 1/5 (Q20) rating of (3) and the lower age group BL3/5, EH1/5 and 1/5 (Q23) rating of (3). Coefficients were 0.8 α, indicating a high level of consistency in participant outcomes.

These findings correlate with the results of Q3: emotions during HRI (E: L2, H3 and B: L0, H2). Q6-Q7 eye contact interaction, Euclid: L6, H7 and Baudi L3, H5. Q14 were the lower age group preferred HRI with Baudi and the higher age group with Euclid: E.H8, L4 and B.H2, L6, depicted in Fig.7.18. Therefore, these results collectively substantiate and validate the findings that the younger age group preferred HRI with Baudi, and the more mature age group preferred HRI with Euclid.

Series. 2C. Human-likeness: (Q13, Q25, Q18, Q21 & Q15). Euclid rated L4/5 and H 3/5 and Baudi L0/5 H1/5. Coefficient factors were high in the dataset, 0.9 α, suggesting a high level of consistency in participant responses. These results correspond with the outcomes of Q1, where both age groups cited Euclid as the more human-like RHR: EL9, H9 and B: L1, H1.Q4, were both age groups suggested Euclid's eye movement as the more aesthetically realistic: E.H8, 6L and B.H2, L2, 1L unclassified.

However, these results conflict with the results of Q5 as the higher age group preferred the eye movement of Euclid, E: H8, L4 and B: H2, L6. These results collectively conclude that both age groups rated Euclid as the most human-like RHR.

Series. 3C. Competency: (Q12, Q16, Q19, Q22 & Q24). The lower age group rated Euclid as more competent 3/5 than the higher age group 2/5, and (3) in Q13, Q16 & Q24 and Baudi rated 2/5 in the higher age group and 1/5 in the lower age group. Internal consistency was acceptable 0.7 α, suggesting some variance in participant outcomes. These results correlate with Q6 and Q7, where both age groups rated Euclid as moving more naturally in head movement: E: L7, H8 and B: L3, H1 and mouth movement L10, H1. However, these results conflict with Q2 speech synthesis, B: L8, H4 and E: L2, H6. Q11 FEs, EL:3, EH:7 and BL:6, BH:3. Q12 AI topical conversation session EL:129 (46%), EH:154 (54%) and BL:95 (39%), BH:145 (61%) and Q13 5-minute AI guessing game: HE: 369 (54%), LE: 316 (46%). The quantitative questionnaire results of Series. 3C conflict with the findings of the qualitative results (Q1-10), making the outcomes of this dataset difficult to cross-analyse and substantiate.

4.4 Analysis of FEA and GSR Biometric Data Feeds

The following section examines the FEA, and GSR biometric data feed extracted from test subjects during the HRI experiment. The FEA reads the following FEs from 0-100%.

1. Rest. 0-100%: this configuration is the default position of the FEA system

2. Frown, 0-100%.: the frown function measures negative FEs which examine the position of the lips, cheeks, and eyebrows.

3. Smile, 0-100%.: similar to the frown function, the FEA system measures the correlation and position of the lips, cheeks and eyebrows to determine if and to what extent the test subject expresses a positive facial expression.

4. Disengage, 0-100%.: this mode engages when the system is unable to track the user.

5. Attention, 0-100%.: the FEA system measures the frequency of the test subjects eye positions with the camera position to monitor attention rates.

The GSR data combined with the FEA data provides an overview of the type and level of stress experienced by the test subjects at any given state. This combinatory approach permits the verification of positive and negative emotional states in the GSR and FEA data to eliminate interference from other facial movements such as speech, gesturing and facial tics that may register as a positive or negative facial expression in the FEA application. This method acts as a low pass filtration system and permits greater accurate measurements of frequency, duration and range of emotive FEs for comparative analysis between the AI results and the findings of the HRI questionnaire.

4.4.1 All Test Subjects: 5-Minute Conversation. FEA and GSR Data Analysis

The results of the FEA biometric data analysis indicated that test subjects spent more time in the rested state when interacting with Euclid (Avg 18%). Compared to Baudi (Avg 15%), suggesting participants spent more time in the neutral facial position when interacting with Euclid. Negative FEA data was higher in Euclid's results, (Avg intensity (INTST), 8%, Fq:228, Avg:11) compared to Baudi (Avg INTST, 7%, Fq:213, Av:11), suggesting subjects experienced greater negative stimulus during HRI with Euclid. Positive FEA results cite a greater positive FE HRI with Baudi (Avg INTST, 10%, Fq:200, Avg:10) compared to Euclid (Avg INTST, 9%, Fq:201, Avg:10). The attention level was marginally higher in Baudi's FEA results, Avg: 73% compared to Euclid, Avg: 72%, shown in Fig. 4. However, subjects exhibited more disengaging behaviour with Baudi than Euclid, (B:53. Avg 3, E:41. Avg 2). The FEA results validate the results of Series. 1A (likeability), Series. 2A (Human-likeness) and Series. 2A (competency). These outcomes collectively suggest greater human-like RHRs emit higher levels of the UV.

Interestingly, although the FEA results were marginal, the range of internal consistency was low across the data set (0.04 α / 0.5 α), suggesting very high levels of variance. Therefore, per the results of the FE data analysis, although test subjects expressed a wide range of FE's during the 5-minute AI conversation experiment, these did not significantly change in frequency or intensity between the RHRs. However, the GSR data analysis indicated higher levels of stress during HRI with Euclid: Total Avg, 17.8μS. (Neg stimuli: Avg 5.6μS / Pos stimuli: Avg 4.2μS) compared to Baudi: Total Avg, 12.3μS. (Neg stimuli: Avg 2.1μS / Pos stimuli: Avg 4.2μS) Suggesting test subjects experienced a greater negative HRI experience with Euclid as cited in the results of Series. 1A.
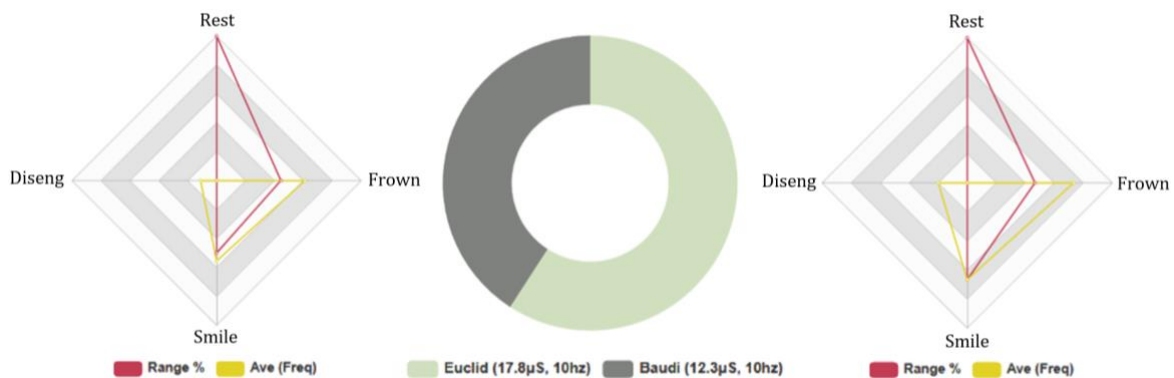
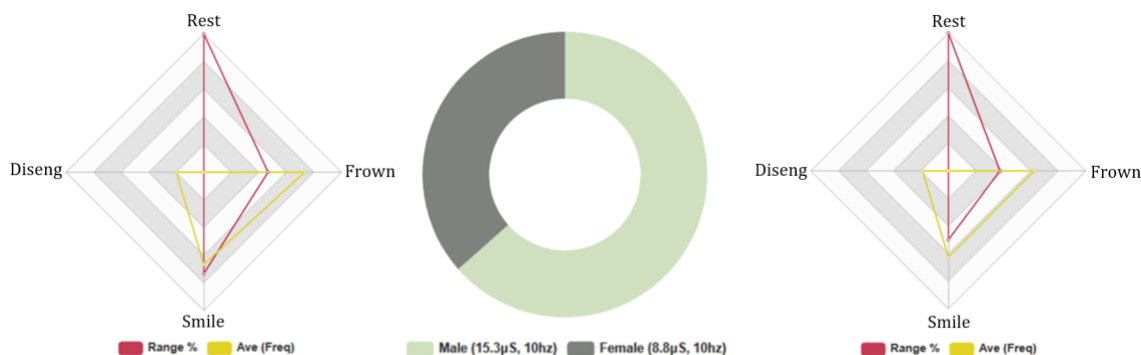**Fig. 4:** 5 Minute Topical AI Conversation Experiment: All Test Subjects
Biometric Data. Left: Euclid FEA results, Right: Baudi FEA results, Middle Euclid and Baudi GSR

4.4.2 Euclid: Females & Males, 5-Min Conversation. Biometric Data Analysis

The FEA results indicate that male test subjects spent more time in the rested FE state when interfacing with Euclid (M:20%, F:16%). Negative FEA data was higher in males subjects, M: (Avg INTST, 9%, Fq:118, Avg:11), F: (Avg INTST, 7%, Fq:110, Avg:12). Comparatively, male subjects displayed greater positive FE than female participants, M: (Avg INTST, 11%, Fq:125, Avg:11), F: (Avg INTST, 7%, Fq:76, Avg:8). These results suggest that male participants expressed the greatest range and intensity of negative and positive FE. The average attention level was similar in the results of male and female test subjects, M: Avg 73%, and, F: Avg 72%. However, male subjects exhibited more disengaging behaviour than female subjects, (M:26. Avg 2, F:19. Avg 2), suggesting greater accuracy in the results of female subjects, shown in Fig.5

Coefficient factors ranged between (0.03 α / 0.7 α), indicating a high level of variance in the test results. However, as in the previous set of results, male test subjects expressed similar FE intensities to females, yet, FE frequencies were significantly higher in male subjects (M:429, F:243), indicating that males engaged with greater emotional FE, HRI with Euclid. Conversely, the GSR biometric data analysis showed that female test subjects exhibited higher levels of stress during HRI with Euclid compared to male participants, F: Total Avg, 19.3μS. (Neg stimuli: Avg 4.9μS / Pos stimuli: Avg 4.4μS) and M: Total Avg, 16.6μS. (Neg stimuli: Avg 3.2μS / Pos stimuli: Avg 2.8μS). The GSR results contradict the FE results; therefore, it is likely that although female test subjects displayed less FE than males, female subjects responded with higher electrodermal activity 'stress'. The FEA test results support the outcomes of Series. 1B (likeability) and Series. 3B (competency), and per the UV the GSR results support the findings of Series. 2B (human-likeness) as both gender groups rated Euclid as human-like at a rate of (4/5).
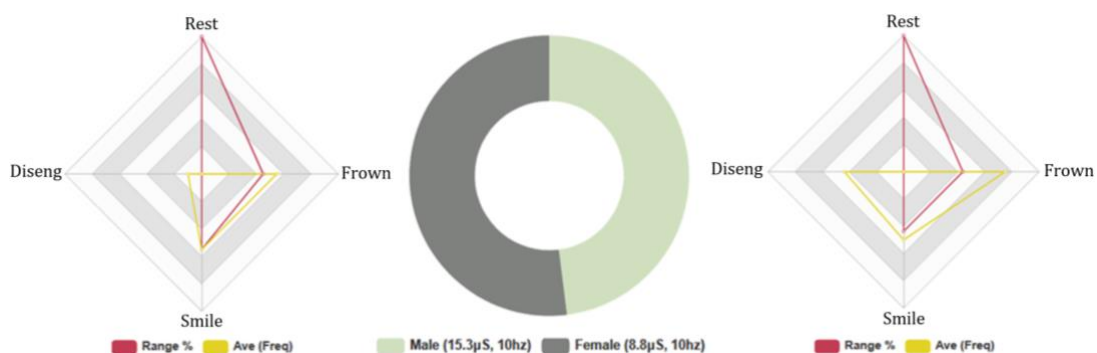
Fig. 5: Euclid: 5-Minute Topical AI Conversation, Male / Female Subjects
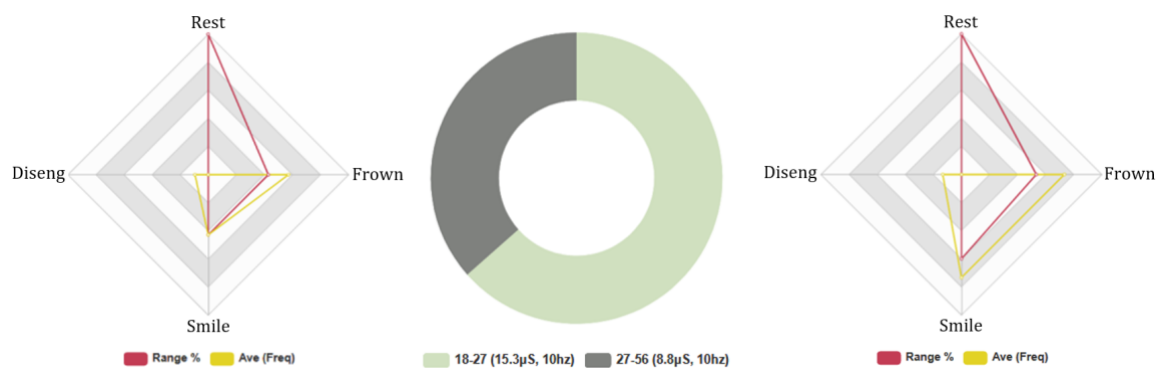
Biometric Data. Left: Male FEA results, Right: Female FEA results, Middle Male and Female GSR

4.4.3 All Test Subjects, 5-Minute Guessing Game, Biometric Data Analysis

The FEA results suggest that candidates spent more time in a neutral FE state when interacting with Baudi during the 5-Minute guessing game session (B:13%, E:14%). Negative FEA indicates that test subjects experienced higher levels of negative FE intensity when interacting with Euclid. However, negative FE frequencies were more elevated in Baudi's results, Euclid: (Avg INTST, 10%, Fq:322, Avg:16), Baudi: (Avg INTST, 9%, Fq:325, Avg:16), suggesting that although participants expressed a higher rate of negative FE during HRI with Baudi, negative FE intensity was greater during HRI with Euclid. Positive FEA results indicate higher levels of positive FE during HRI with Baudi, E: (Avg INTST, 7%, Fq:205, Avg:10), B: (Avg INTST, 8%, Fq:197, Avg:10), suggesting test subjects experienced greater positive FE during HRI with Baudi. The average level of attention was higher during HRI wit Baudi, E Avg 71%, and, B: Avg 74%. However, disengaging behaviour in test subjects was significantly higher in Baudi's results, yet similar on average, (E:20. Avg 2, B:49. Avg 2), suggesting specific individuals exhibited high levels of disengaging behaviour which was not representative of the whole dataset.

These results correlate with Series 1A (likeability), were subjects cited Baudi positive emotional responses. Internal consistency was low across the dataset ranging between (-0.06 α / 0.4 α), indicating a high level of variance in FEA data. On average GSR results suggest higher levels of positive electrodermal activity during HRI with Baudi and greater negative GSR during HRI with Euclid, which supports the FEA results and the results of Series 2A (human-likeness) and Series. 3A (competency) per the UV. Euclid: Total Avg, 10.10μS. (Neg stimuli: Avg 4.3μS / Pos stimuli: Avg 2.8μS) and Baudi: Total Avg, 12.10μS. (Neg stimuli: Avg 3.1μS / Pos stimuli: Avg 5.2μS), depicted in Fig.6.
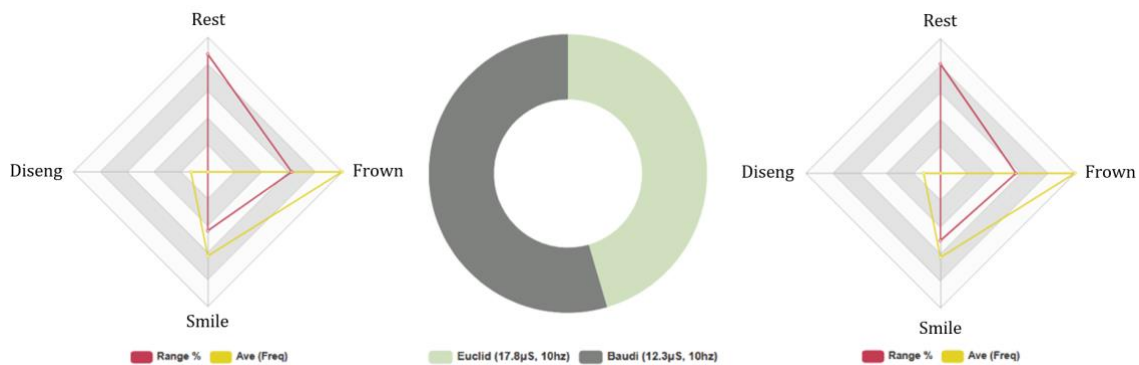
Fig. 6: 5 Minute AI Guessing Game Experiment: All Test Subjects

Biometric Data Comparison. Left: Euclid FEA results, Right: Baudi FEA results, Middle GSR results

4.4.4 Baudi: Females & Males, 5-Minute Conversation. FEA/GSR Data Analysis

The FEA results cite that female subjects spent more time in the rested FE state during HRI with Baudi, (F:16%, M:15%). Negative FEA was higher in male subjects, M: (Avg INTST, 7%, Fq:125, Avg:11), F: (Avg INTST, 6%, Fq:88, Avg:10), suggesting that although FE intensity was similar in both gender groups, the frequency of negative FE was significantly higher in male subjects. Comparatively, positive FEA was higher in male participants, M: (Avg INTST, 11%, Fq:114, Avg:10), F: (Avg INTST, 8%, Fq:86, Avg:10), indicating similar levels of intensity in both gender groups. However, positive and negative FE frequencies were considerably higher in male test subjects compared to female participants. The FEA results suggest that males subjects experienced higher levels of emotive FE during HRI with Baudi than female subjects. The average attention level was higher in the male test group, M: Avg 74%, and, F: Avg 72% and similar in disengaging behaviour, (M:29. Avg 3, F:23. Avg 3), suggesting marginally higher accuracy in the results of female participants. Internal consistency was low in the dataset ranging from (0.02 α / 0.7 α), indicating high variability in the FEA test results.

The FEA results correspond with Series. 1B (likeability), and Series. 3B (competency) as male subjects rated the RHRs higher than females. However, the FEA data conflict with the outcomes of Series 2B (human-likeness), which cited both gender groups as experiencing similar emotions during HRI. GSR data analysis reinforces the results of the FEA data and questionnaire results, as male subjects discharged higher levels of electrodermal activity than female participants, M: Total Avg, 15.3μS. (Neg stimuli: Avg 3.6μS / Pos stimuli: Avg 4.2μS) and F: Total Avg, 8.8μS. (Neg stimuli: Avg 2.1μS / Pos stimuli: Avg 3.5μS), suggesting that male subjects experienced significantly higher levels of positive and negative emotional stimulus during HRI, depicted in Fig. 7.

Fig. 7. Baudi: 5-Minute Topical AI Conversation, Male / Female Subjects

Biometric Data. Left: Male FEA results, Right: Female FEA results, Middle Male and Female GSR results

4.4.5 Euclid: Age Groups, 5-Minute Conversation. FEA/GSR Data Analysis

The FEA results indicate that the lower age group spent significantly more time in a neutral FE state, L: 21%, H: 15%, during HRI with Euclid. Negative FEA was higher in the lower age group for FE intensity and frequency, L: (Avg INTST, 9%, Fq:118, Avg:12), H: (Avg INTST, 8%, Fq:110, Avg:11). Positive FEA results indicate that both age groups experienced similar positive FE intensities and total frequencies. However, the average positive FE frequency was greater in the higher age group. L: (Avg INTST, 9%, Fq:94, Avg:9), H: (Avg INTST, 9%, Fq:107, Avg:11).

The average level of attention was higher in the lower age group, L: Avg 73%, and, H: Avg 71% and the lower age group exhibited higher levels of disengaging behaviour, (L:19. Avg 2, H:22. Avg 2), suggesting similar levels of accuracy in both gender groups. Coefficients were incredibly low across the dataset, ranging between (-0.05 α / 0.6 α), indicating a high level of variance in FEA test results. These outcomes support the findings of Series 1C (likeability), were the higher age group rated a greater positive HRI experience with Euclid. Furthermore, the FEA outcomes verify the results of; Series. 2C (human-likeness) where the lower age group rated Euclid as more human-like and Series. 3C (Competency), where both age groups cited similar positive and negative emotions during HRI.

The results of the GSR data analysis indicated greater levels of electrodermal activity positive and negative stimuli in the higher age group, L: Total Avg, 22.5μS. (Neg stimuli: Avg 4.4μS / Pos stimuli: Avg 5.1μS) and H: Total Avg, 13.2μS. (Neg stimuli: Avg 6.2μS / Pos stimuli: Avg 3.4μS). The GSR outcomes add validity to the findings of the FEA data analysis as they suggest that the lower age group exhibited greater negative emotions during HRI with Euclid compared to the senior group, shown in Fig. 8.

Fig. 8: Euclid: 5-Minute Topical AI Conversation Age Groups

Biometric Data. Left: (27-56) FEA results, Right: (18-27) FEA results, Middle: (18-27) / (27-56) GSR

4.4.6 Baudi: Age Groups, 5-Minute Conversation. FEA/GSR Analysis.

The FEA results indicate that the lower age group spent more time in a rested FE state (L:17%, H:14%). Negative FEA intensity was proximal in both age groups; however, negative FEA frequency was marginally greater in the higher age group, L: (Avg INTST, 7%, Fq:101, Avg:10), H: (Avg INTST, 7%, Fq:112, Avg:11). These results suggest that the higher age group experienced similar negative emotions during HRI with Baudi. Conversely, positive FEA was significantly greater in the higher age group, H: (Avg INTST, 9%, Fq:103, Avg:10), L: (Avg INTST, 8%, Fq:97, Avg:10), indicating that the higher age group expressed greater positive emotional HRI with Baudi.

The average attention level was higher in the lower age group, L Avg 73%, and, H: Avg 71% and the higher age group exhibited significantly greater disengaging behaviours, (H:33. Avg 3, L:19. Avg 2). These results suggest greater concentration during HRI and higher accuracy in the results of the lower age group. Coefficient factors were relatively low in the dataset ranging between (0.2 α / 0.6 α), suggesting a high level of variability in the FEA test results. GSR data analysis indicates that on average, the lower age group exhibited greater electrodermal activity than the higher age group, L: Total Avg, 16.8μS. (Neg stimuli: Avg 2.4μS / Pos stimuli: Avg 4.1μS) and H: Total Avg, 7.9μS. (Neg stimuli: Avg 3.1μS / Pos stimuli: Avg 2.5μS). However, both age groups discharged similar levels of negative electrodermal activity, and the higher age group exhibited greater positive GSR readings which support the outcomes of the FEA data analysis, shown in Fig. 9. These results collectively support the findings of Series. 2C (human-likeness) and Series. 3C (competency) were test subjects cited similar ratings and positive and negative emotions during HRI with the RHRs. However, more significantly, these results support the outcome of Series. 1C (likeability), which suggests that the lower age group, preferred HRI with a younger-looking RHR.
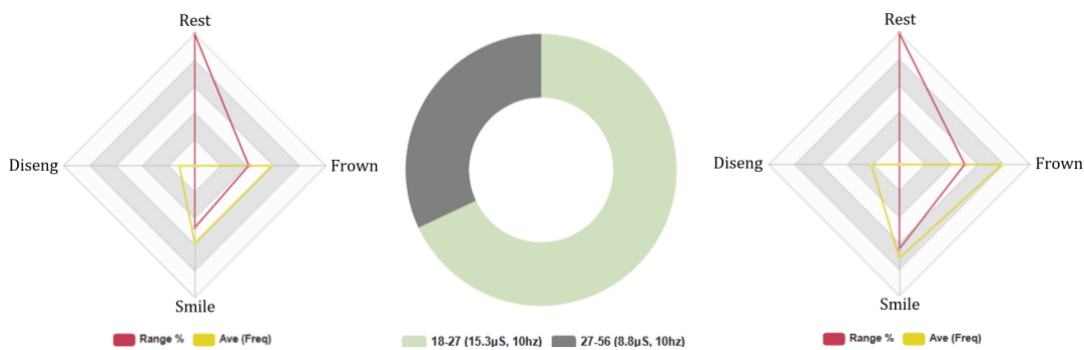
**Fig. 9:** Baudi: 5-Minute Topical AI Conversation Age Groups

Biometric Data. Left: (18-27) FEA results, Right: (27-56) FEA results, Middle: (18-27, 27-56) GSR

4.5 FEA / GSR Biometric Data Analysis: 5 Min Guessing Game HRI.

4.5.1 Euclid: Females & Males 5-Min Guessing Game, FEA / GSR, Analysis

The FEA results indicate that male subjects spent more time in a neutral FE position (M:14%, F:13%). Negative FE results were higher in male test subjects, M: (Avg INTST, 11%, Fq:179, Avg:16), F: (Avg INTST, 9%, Fq:143, Avg:16). Positive FEA was also higher in male candidates, (Avg INTST, 8%, Fq:118, Avg:11), compared to female participants (Avg INTST,7%, Fq:87, Avg:10), suggesting male test subjects exhibited greater emotive FE during the 5-minute guessing game HRI session. The FEA data support the findings of Series. 1B (likeability) and 3B (competency) were male subjects cited comparable results. Coefficient factors were low across the dataset ranging between (0.01 α / 0.5 α), indicating a high level of variance in FEA results.

However, male subjects exhibited significantly more disengaging behaviour than female subjects, (M:32. Avg 3, F:15. Avg 2), suggesting greater accuracy and attention in the results of female subjects. The results of the average attention rate support these outcomes as female participants rated higher levels of concentration than males, M: Avg 71%, and, F: Avg 72%. GSR data analysis indicated that female test subjects demonstrated higher levels of positive and negative electrodermal activity than male candidates, Males: Total Avg, 9μS. (Neg stimuli: Avg 3.6 μS / Pos stimuli: Avg 1.5 μS) and Females: Total Avg, 14 μS. (Neg stimuli: Avg 3.1μS / Pos stimuli: Avg 2.4μS), shown in Fig. 10. The FEA and GSR results add validity to the results of the 5-min topical conversation session with Euclid were male subjects expressed greater emotive FE, and female participants greater emotional GSR during HRI. These results support the findings of Series. 2B (human-likeness) were both gender groups rated comparable feelings and emotions during HRI.
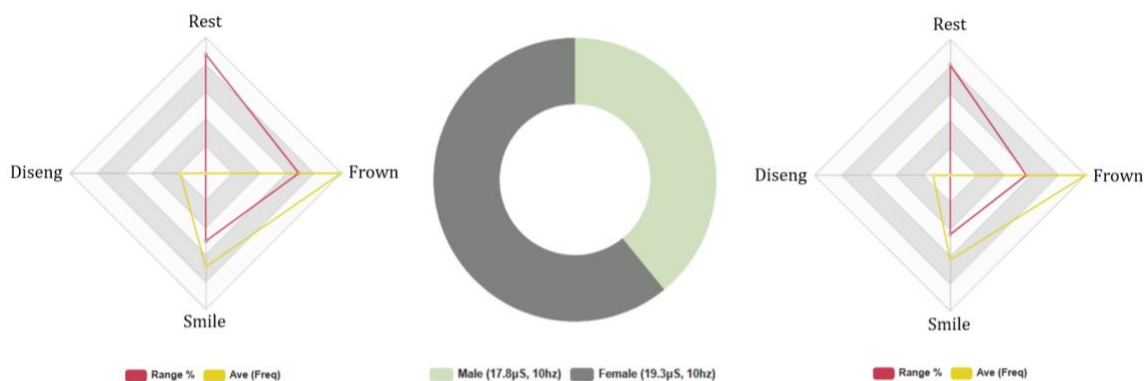
Fig. 10: Euclid: 5 Minute AI Guessing Game Test: Male / Female Subjects

Biometric Data. Left: Male FEA results, Right: Female FEA results, Middle Male and Female GSR

4.5.2 Baudi: Females & Males, 5-Min Guessing Game, FEA / GSR Analysis

The FEA results indicate that female test subjects registered greater neutral FE states during HRI with Baudi, (M:12%, F:13%). The Negative FEA outcomes suggested male subjects exhibited greater intensity and frequency of negative FE, M: (Avg INTST, 10%, Fq:187, Avg:17). F: (Avg INTST, 8%, Fq:138, Avg:15), and both gender groups expressed similar positive FE intensity M: (Avg INTST, 8%, Fq:109, Avg:10), F: (Avg INTST, 8%, Fq:88, Avg:10). However, positive FE frequencies were higher in the results of male test subjects. The FEA results support the outcomes of Series 1B (likeability), and Series 2B (human-likeness) were male subjects rated Baudi as the less human-like and competent RHR compared to female candidates M0/5, F1/5.

Attention levels were higher in female test subjects, M: Avg 72%, and, F: Avg 76% and male subjects exhibited greater disengaging behaviours, (M:29. Avg 3, F:20. Avg 2), suggesting greater immersion and accuracy in the results of female subjects. Internal consistency was low across the dataset ranging between (0.01 α / 0.4 α), indicating high levels of variance in the FEA data. GSR results suggest that female test subjects exhibited higher levels of electrodermal activity than male candidates, M: Total Avg, 8μS. (Neg stimuli: Avg 2.2μS / Pos stimuli: Avg 1.9μS) and F: Total Avg, 16μS. (Neg stimuli: Avg 4.2μS / Pos stimuli: Avg 3.9μS), as highlighted in Fig 7.11. Interestingly, these results correlate with Baudi's FEA and GSR results during the 5-minute topical conversation session. Therefore, it is highly likely that the male test subjects expressed greater positive and negative emotive FE, and female participants dispersed higher levels of positive and negative electrodermal activity during HRI with Baudi. These outcomes are significant as results cite notable differences between the gender groups in FEA and GSR data, which adds greater validity to the results of Series 3B (competency).
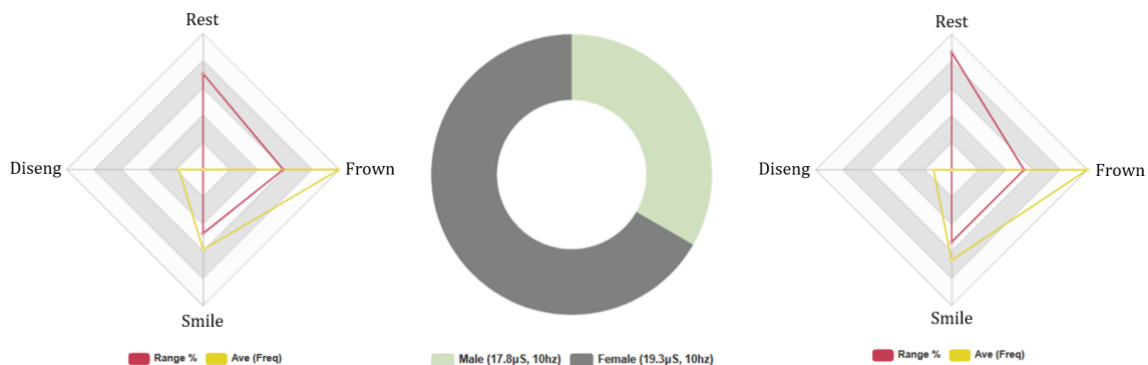
**Fig. 11:** Baudi: 5 Minute AI Guessing Game Test: Male / Female Test Subjects

Biometric Data. Left: Male FEA results, Right: Female FEA results, Middle Male and Female GSR

4.5.3 Euclid: Age Groups: 5-Minute Guessing Game, FEA/GSR, Analysis.

The FEA results suggest that the lower age group spent more time in a rested FE state (L:14%, H:13%). Negative FEA was higher in the lower age group in intensity and frequency, L: (Avg INTST, 11%, Fq:177, Avg:18), H: (Avg INTST, 8%, Fq:145, Avg:15. Conversely, positive FEA intensity and frequency were greater in the higher age group, L: (Avg INTST, 7%, Fq:101, Avg:10), H: (Avg INTST, 8%, Fq:104, Avg:10). The negative and positive FEA outcomes parallel with Euclid's previous FEA results in the 5-minute topical conversation session. The FEA results correspond with the findings of Series. 1C (likeability), were the higher age group cited having a positive HRI, and Series. 2C (human-likeness) and Series. 3C (competency) were the higher age group rated Euclid as more human-like (per the UV).

Internal consistency was low, ranging between, (0.01 α / 0.5 α), indicating high variability in the FEA data. The average attention rate was greater in the higher age group, L: Avg 69%, and, H: Avg 73%. However, the higher age group exhibited more disengaging behaviours (L:29. Avg 3, H:20. Avg 2). These results suggest higher attention levels in the lower age group and greater distraction in the higher age group. The FEA results contrast with the 5-minute topical conversation session AI data. A probable cause for this shift is the parameters of the gamification session, which requires concentration and strategy, compared to the 5-minute topical conversation session, which is relaxed and unstructured. GSR readings indicate that the lower age group exerted higher levels of positive electrodermal activity than the higher age group, and negative GSR was higher in the lower age group, H: Total Avg, 15μS. (Neg stimuli: Avg 3.6μS / Pos stimuli: Avg 2.3μS), and L: Total Avg, 8μS. (Neg stimuli: Avg 2.9μS / Pos stimuli: Avg 3.9μS). The GSR results correlate and support the results of the FEA data, shown in Fig. 7.12.
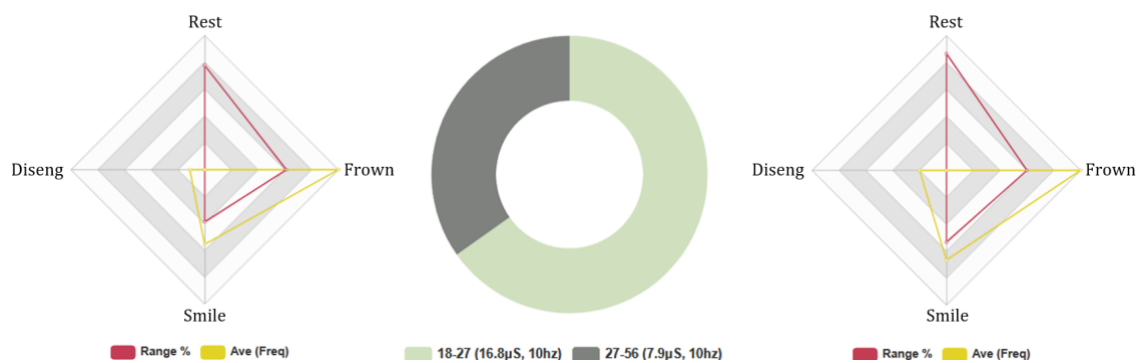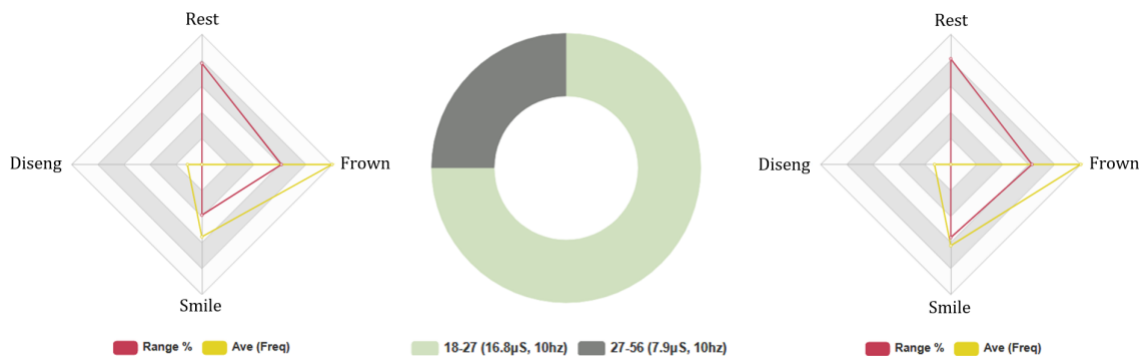

Fig. 12: Euclid: 5-Minute AI Guessing Game, Age Groups (18-27) / (27-56)

Biometric Data. Left: (18-27) FEA results, Right: (27-56) FEA results, Middle: (18-27) / (27-56) GSR

4.5.4 Baudi: Age Groups: 5-Minute Guessing Game, FEA / GSR, Analysis

The FEA results indicate higher levels of resting FE behaviour in the lower age group, (L:12%, H:13%). Negative FEA intensity was greater in the higher age group, and negative FEA frequency higher in the lower age group, L: (Avg INTST, 9%, Fq:164, Avg:16), H: (Avg INTST, 10%, Fq:161, Avg:16). Conversely, Positive FEA results indicate that the higher age group expressed greater intensive and frequent FE, L: (Avg INTST, 8%, Fq:95, Avg:10), H: (Avg INTST, 9%, Fq:102, Avg:10), as illustrated in Fig. 13. Coefficient factors were low, ranging between, (0.01 α / 06. α), indicating high levels of variance in the FEA data. Attention levels were greater in the higher age group, L: Avg 73%, and, H: Avg 75%, and disengaging behaviour was lower in the higher age group, (L:26. Avg 3, H:23. Avg 2). The FEA results conflict with the previous FEA results registered during the 5-minute topical conversation session, as observed in Euclid's results. The GSR results indicate that the lower age group exerted significantly higher levels of positive and negative electrodermal activity, H: Total Avg, 18μS. (Neg stimuli: Avg 4.1μS / Pos stimuli: Avg 3.3μS), and L: Total Avg, 6μS. (Neg stimuli: Avg 1.9μS / Pos stimuli: Avg 1.1μS). These outcomes support the findings of Series 1C (likeability) and Series. 2C (human-likeness), and conflict with the results of Series. 3C (competency) as the lower age group rated Baudi as the more competent RHR.

However, the lower age group exhibited higher GSR and FEA readings in the gamification session than the topical conversation component. These results suggest the lower age group experienced a greater immersive and visceral HRI during the guessing game session compared to the more mature age group, which exerted higher biometric feedback during the verbal communication component. These outcomes are explored and comparatively analysed against the AWS AI data in the following section.

Fig. 13: Baudi: 5-Minute AI Guessing Game, Age Groups (18-27) / (27-56)

Biometric Data. Left: (18-27) FEA results, Right: (27-56) FEA results, Middle: (18-27) / (27-56) GSR

## 4.6 Topical Conversational and Guessing Game AI Data Analysis

The following section examines the results of the 5-minute topical conversation and the guessing game AI sessions recorded in AWS during HRI. The results are categorised and cross analysed with the HRI questionnaire results and FEA and GSR data to validate findings.

4.6.1 Euclid and Baudi: All Test Subjects, Topical Conversational AI Data

Euclid delivered 283/545 (52%) of questions correctly, Fig. 14 and Baudi achieved a lower rating, 241/497 (48%), Fig.15. Euclid produced incorrect responses at a rate of 158/545 (28%) and Baudi 153/497 (31%), and Euclid executed the 'please repeat' command, 144/545 (20%) compared to Baudi, 105/497 (21%). These results do not tally with the findings of Q8 as subjects cited Euclid as providing more correct responses, E:7/13, B: 3/7. However, the AI test results support the outcomes of Series 3A as subjects rated Euclid as the more competent RHR (incorrect and repeat responses), and the FEA and GSR, as subjects rated higher levels of positive biofeedback during HRI with Baudi.
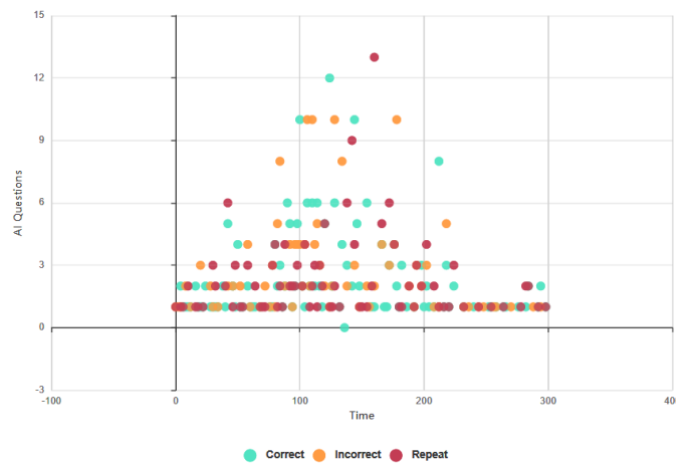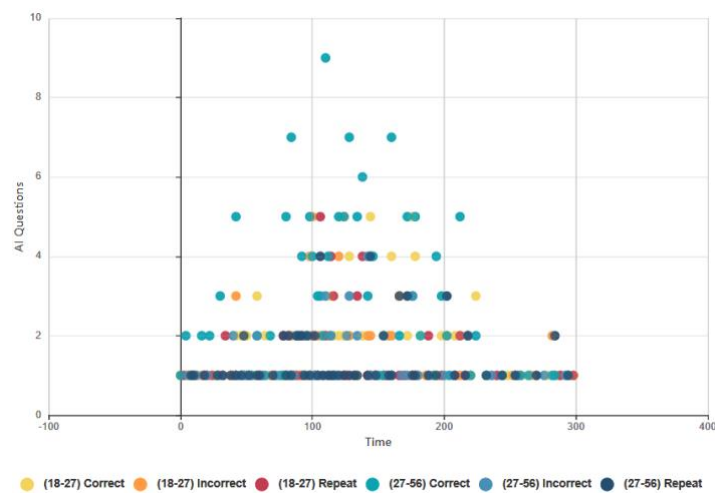

Fig. 14: Euclid: 5-Min Topical Conversation, AI Data (All Test Subjects)
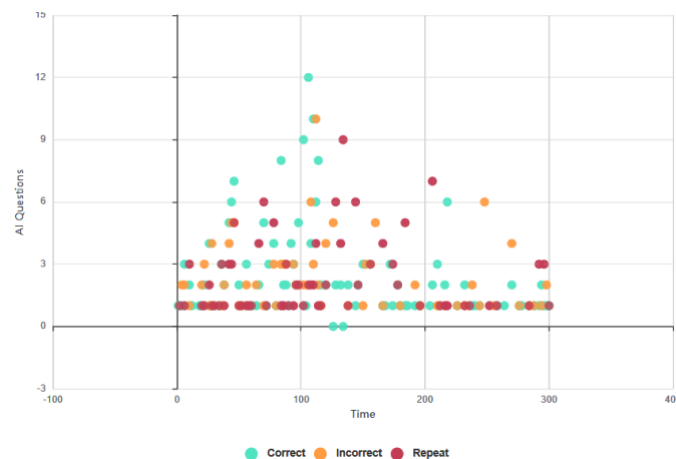
Fig. 15: Baudi: 5-Min Topical Conversation, AI Data (All Test Subjects)

4.6.2 Euclid and Baudi: Males and Females, Topical Conversation AI Data

Male test subjects achieved a greater number of correct AI responses than females during HRI with Euclid, EM:146 (47%), EF:136 (55%), depicted in Fig. 16 and similarly with Baudi BM:133 (45%), BF:108 (54%), Fig. 17. The AI test results support the outcomes of Q9 as more male test subjects: M8, F5 preferred to converse with Euclid, and similarly in the results of Baudi M3, F4, as Euclid achieved greater accuracy E:283, B:241. However, males scored a higher rate of incorrect responses than females, EM:96 (31%), EF:62 (25%) and BM:88 (30%), BF:65 (32%) and in instigating the, 'please repeat' response, EM:66 (22%), EF:48 (20%) and BM:72 (25%), BF:33 (16%) which suggests male test subjects asked a higher number of questions outside the scope of the AI system. These results support the findings of Series. 3B (competency) as male candidates rated Euclid as the more competent RHR, and correlate with the gender-based FEA and GSR biometric data results during the topical conversation session.
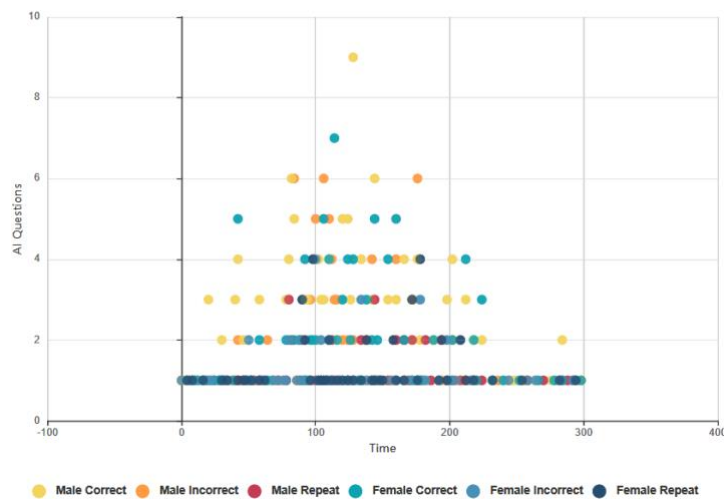
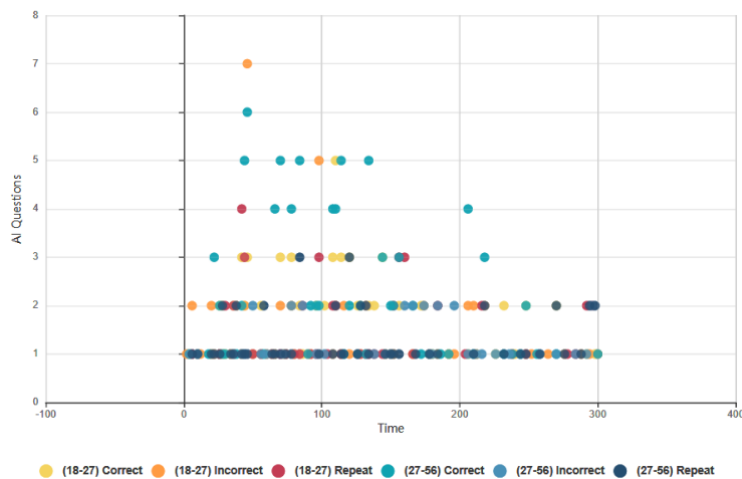**Fig. 16:** Euclid: 5-Min Topical Conversation, AI Data (Males and Females)
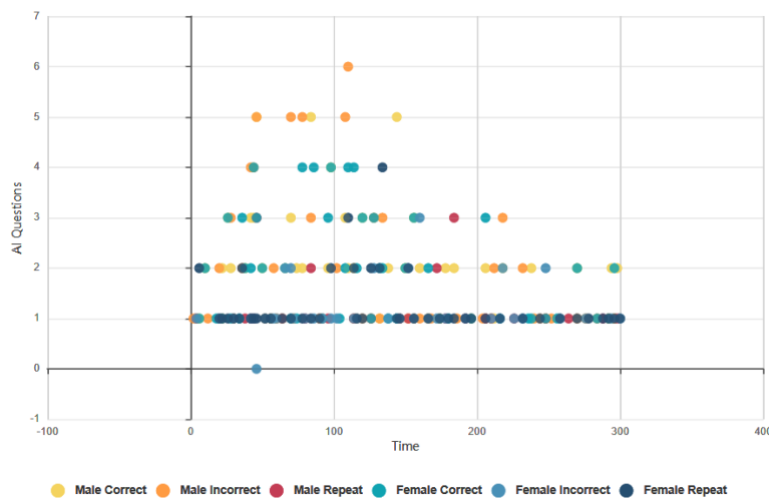
Fig. 17: Baudi: 5-Min Topical Conversation, AI Data (Males and Females)

4.6.3 Euclid and Baudi: Age Groups Topical Conversational AI Data.

The higher age group achieved a greater number of correct responses during HRI with Euclid, EL:129 (47%), EH:154 (55%) and similar results during HRI with Baudi, BL:95 (39%), BH:145 (61%) and incorrect answers EL:87 (55%), EH:71 (45%) and BL:74 (41%), BH:79 (30%), and in instigating the, please repeat response' EL:59 (21%), EH:55 (20%) and BL:61 (27%), BH:44 (16%) suggesting a greater competency in the higher age group which corresponds with the outcomes of Q8: EL:4, EH:9 and BL:6, BH:1, shown in Fig 18 & Fig. 19. However, the AWS, AI findings contradict the results of Series. 3C (competency) as the lower age group rated Euclid as the most competent RHR and the lower age group cited Baudi as the more competent RHR. Nevertheless, these outcomes support the findings of the biometric data analysis as the higher age group exhibited greater FEA, GSR and attention during the topical conversation session.


**Fig. 18:** Euclid: 5-Min Topical Conversation, AI Data (Age Groups)

Fig. 19: Baudi: 5-Min Topical Conversation, AI Data (Age Groups)

4.6.4 Euclid and Baudi: All Test Subjects, Guessing Game AI Data

The objective of the AI guessing game was to engage subjects with a greater challenging mode of HRI to examine how the DL AI functioned at a faster rate of processing. The AI results indicate that Baudi achieved a marginally higher accuracy rating than Euclid: B:705/1235, (57%), E:685/1223, (56%). However, Baudi generated a higher sum of incorrect answers than Euclid, B:257/1235, (21%), E:245/1223, (20%) and 'please repeat' replies E:292 (24%), B:273 (22%) shown in Fig. 20 & 21. These results support the findings of Series. 3A (competency) as test subjects rated the RHRs as competent, E:3/5, B:2/5 with (2/5 equivalent ratings of 3). Significantly, the results of Q9 (guessing game) conflict with the findings of Series. 3A, yet validate the AWS AI data. Thus, Euclid rated as more competent in movement and Baudi in AI interaction during the guessing game session. Furthermore, the biometric data analysis validates the AI results as Baudi rated higher in positive FEA/GSR than Euclid during this session.
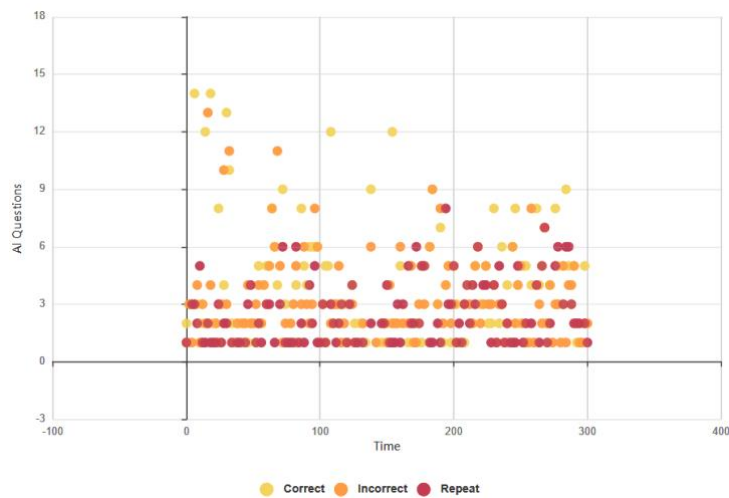

Fig. 20: Euclid: 5-Minute Guessing Game Results, AI Data (All Participants)
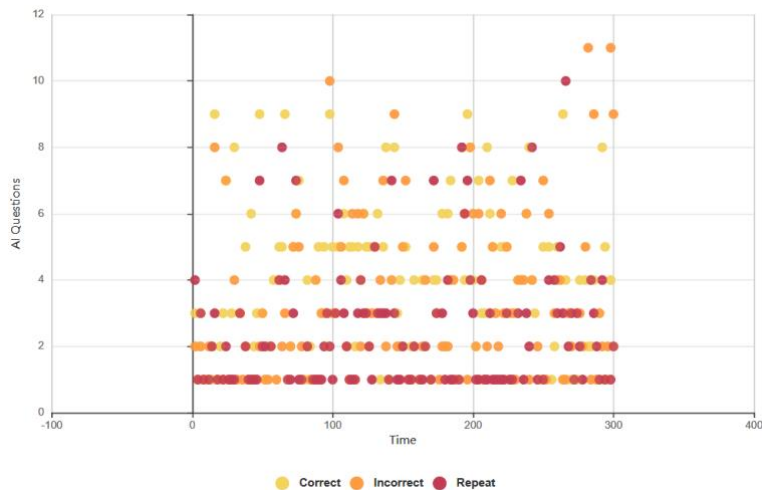

Fig. 21: Baudi: 5-Minute Guessing Game Results, AI Data (All Participants)

5.6.5 Euclid and Baudi: Male and Female Subjects Guessing Game AI Data

The AWS AI results indicate a higher number of male subjects achieved correct responses than females: EM: 394 (57%), EF:291 (55%) and BM: 415 (59%) BF: 290 (54%). However, male candidates incurred a higher sum of incorrect responses than females: EM: 128 (18%), EF: 117 (22%) and BM: 140 (20%), BF: 117 (22%), and similar in the 'please repeat' response, EM: 168 (25%), EF: 124 (23%) and BM: 147 (21%), BF: 126 (24%), shown in Fig. 22 & Fig. 23. These figures are significant as female test subjects asked fewer questions than males FE:532 (Av 27), ME:691 (Av 35) and FB:533 (Av 27), MB: 702 (Av 35). These AI test scores support the outcomes of Q9 as M8, F3 preferred to interact with Baudi and M3, F6 with Euclid, and validate the outcomes of Series 3B, as male subjects rated Euclid as the more competent RHR. Furthermore, the results of Series. 3B (competency) cited high variability in participant responses that parallel with the frequencies of incorrect and 'please repeat' responses indicated in AWS AI analytics.
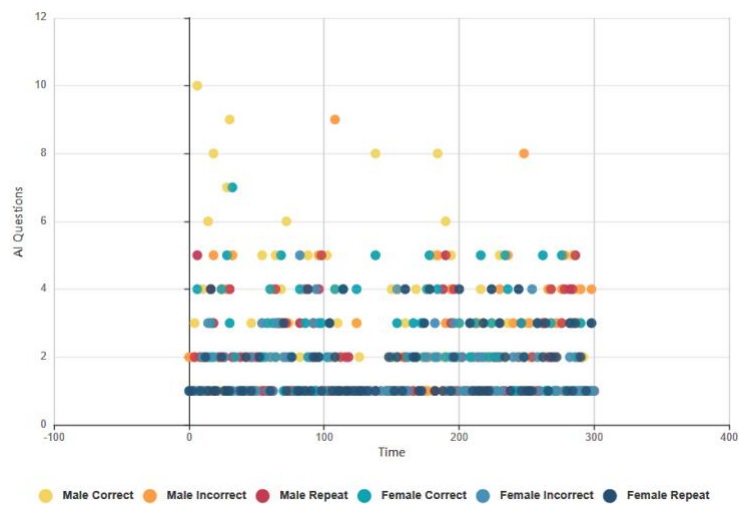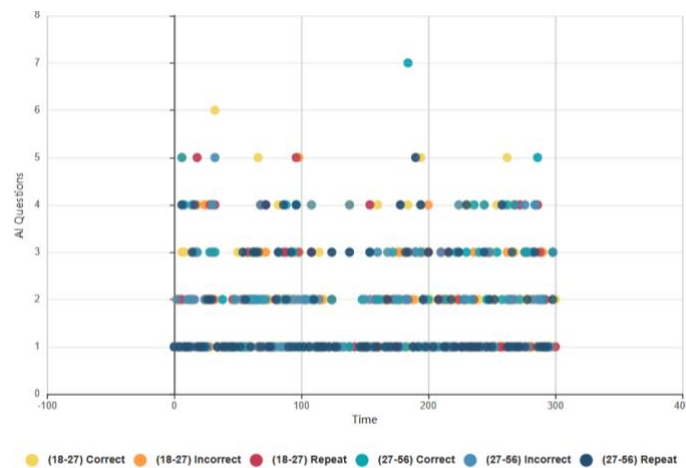

**Fig. 22:** Euclid: 5-Min Guessing Game, AI Data (Males and Females)
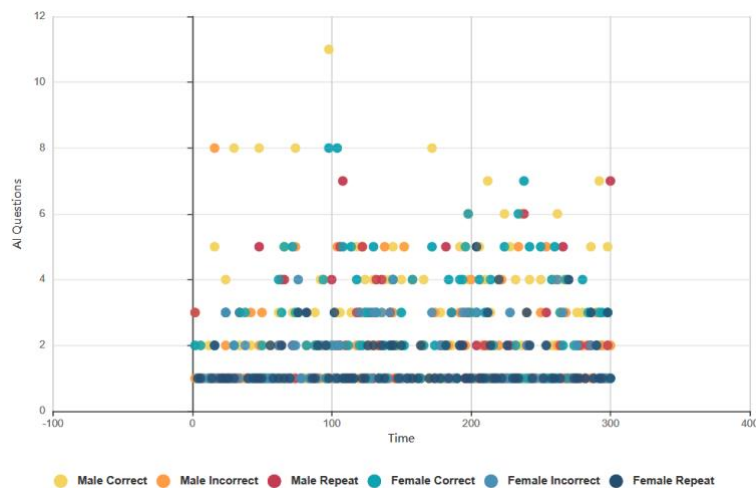
Fig. 23: Baudi: 5-Min Guessing Game, AI Data (Males and Females)

4.6.6 Euclid and Baudi: Age Groups Guessing Game AI Data

The higher age group achieved greater accurate responses during the guessing game session: correct responses: EL: 316 (54%), HE: 369 (61%), BL: 338 (57%), BH: 367 (57%), and incorrect responses EL: 119 (20%), EH: 126 (18%), BL: 126 (21%), BH: 131 (21%) and inciting the 'please repeat' response, EL: 151 (26%), EH: 141 (21%), BL: 132 (22%), BH: 141 (22%), shown in Fig 24 and. Fig 25. However, there was little statistical difference in the age-related results of Q9, Baudi: L5, H6 and Euclid L5, H4. These results conflict with the outcomes of Series. 3C (competency) due to the high variance in participant outcomes. However, the biometric data analysis indicated higher levels of positive feedback in the lower age group than in the more senior age group, which collates with the AI response data. Significantly, the level of attention was greater in the higher age group during the guessing game session, which validates the AWS AI data. Nevertheless, these outcomes are problematic to substantiate due to the high level of variance between the results of the HRI questionnaire, FEA, GSR and AWS AI data.

Fig. 24: Euclid: 5-Minute Guessing Game, AI Data (Age Groups)
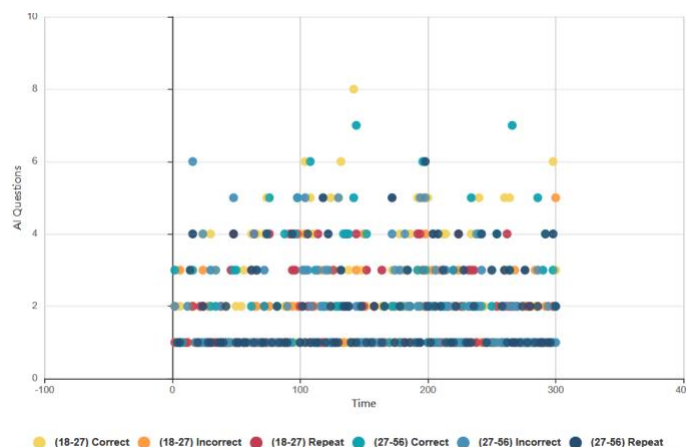
Fig. 25: Baudi: 5-Minute Guessing Game, AI Data (Age Groups)

5. Conclusion

The results of this study show that the younger age group preferred HRI with a younger-looking RHR and the more senior age group HRI with an older looking RHR, reinforced with the FEA/GSR data. These results provide a strong foundation for the appropriation of age in RHR design in fields such as older looking RHRs in care for the elderly and younger-looking RHRs as social companions for the young, vulnerable and isolated. Furthermore, test subjects stated that they felt greater trust towards an older looking RHR compared to a more youthful RHR as it reminded them of the trust and care of older relatives. Similarly, the gender-based analysis indicated that male test subjects preferred HRI with Euclid (older in appearance) and females with Baudi (younger in appearance). Although there is little scientific evidence to substantiate the reasoning behind this outcome, it is an interesting finding that requires further study.

Upon review of the AWS AI data, male subjects asked more questions outside the scope of the AI system compared to female test subjects. This anomaly was observed during HRI and identified in the outcomes of the questionnaire when test subjects purposefully attempted to deceive the AI system into giving incorrect responses. Thus, more male subjects prompted the RHRs into giving incorrect responses compared to female participants. This outcome coincides with the FEA data analysis as male subjects expressed significantly greater disengaging behaviours and reduced attention levels than female participants. Thus, it is plausible that as the RHRs are modelled on the male form, male participants felt a need to test the capacity and abilities of the system more than female subjects. This hypothesis may account for the irregularities between the HRI questionnaire and biometric data results as in the survey; many male subjects cited feeling no emotions during HRI. However, these outcomes contradicted the FEA data that firmly showed male subjects exerted significantly higher levels of positive and negative FE biometric feedback than females.

Finally, another possible method of evaluating RHRs is to compare two identical systems side by side. This method would reduce influential factors such as different ages and gender and narrow the scope of evaluation to specific variables such as static and dilating pupils. However, creating identical RHRs is incredibly difficult due to the variability in materiality, aesthetics and application of hand-made silicone skin onto the underlying exoskeleton. This visual irregularity is observable in the RHR 'Sophia' as new generations of the RHR look distinctly different from the first due to the application and reapplication of the silicone skin.

Nevertheless, as in this study, comparing RHRs with different appearances, genders and personalities permits the investigation of gender and age in terms of user preference, which yielded highly interesting and intriguing results for modelling user preference in EAI and appearance in future RHR design.

5.1 Employing FEA and GSR to support Questionnaire results

The combinatory approach to biometric data gathering, camera feed analysis and AI analytics provided critical insights into the behaviours and emotions of test subjects not cited in the HRI questionnaire. For example, male participants expressed significantly greater positive and negative emotive FE than females during HRI. Yet, the majority of male subjects cited feeling no emotions in the HRI questionnaire. Similarly, the lower age group expressed greater levels of negative FEA and GSR during the 5-minute topical conversation and guessing game AI sessions, which correlates with the results of the AI test data as the lower age group achieved a lower number of correct answers and a higher number of incorrect and 'please repeat' responses during HRI. However, these results conflict with the questionnaire outcomes as there were little discernable differences in participant results as male subjects stated feeling little to no emotions during HRI. Furthermore, the implementation of the GSR and FEA systems did not restrict the participant's freedom of movement and accessibility to the RHRs.

This outcome is significant for future HRI studies as these non-invasive methods of biometric data gathering permitted test subjects to interact and communicate with the RHRs naturally without any obstruction or interference from the biometric sensors. Employing a mixed-method biometric data gathering approach permitted a more precise analysis of the data feeds as high and low levels of FEA were confirmable by the GSR data and a drop/loss of GSR signal verifiable with the FEA data and camera feed. This method provided consistent and valid biometric data for analysis as anomalies in the data feed could be attributed to either incredibly higher or low levels of emotional stimulus or as a result of issues with the connectivity of the biometric systems. Finally, an issue with the GSR and FEA mixed-method approach was the delay in the data processing as the GSR lagged the FEA, which resulted in a time offset of approx. ±5s between the feeds. However, this issue was negatable by examining the Hz frequencies of the devices to determine the processing rate and electrode sensitivity of the systems to help synchronise the data fields. Moreover, the Esense GSR application provides the precise offset of the stimulus to the biometric readings, in seconds and milliseconds, which also helped precisely align the two datasets for comparative analysis.

5.2 Ethical and Broader Issues in RHR design and Application in Society

The outcomes of this study provided critical data regarding how the test subjects perceive the future of RHRs in society. One key area that test subjects highlighted as a suitable application for RHRs with EAI was the care and support industry as the RHRs provided useful and entertaining content in a manner proximal to human-human communication. In support, 45% of participants stated that as the RHRs cannot make personal judgements or become offended, HRI was more relaxed, honest and free-flowing than human-human communication.

This outcome correlates with previous studies [96], [97], [98] and the results of the HRI age-related statistical analysis that suggested younger test subjects preferred HRI with a younger-looking RHR and the more mature age group with an older looking RHR. This study covered two key areas of RHR and EAI ethics relating to robot rights and the fallibility of human perception in RHR design. Firstly, 75% of candidates stated that RHRs should be treated as machines regardless of human-likeness and intelligence. However, this approach is susceptible to a fundamental flaw in human perception, as if an RHR can authentically emulate a human in a manner proximal to the human condition in real-world conditions, then there is no definitive way of determining if the RHR is human or robot without an internal examination.

Secondly, 90% of participants advocated that it is unnecessary to give RHRs human / human-like rights and citizenship. Nevertheless, these results do not account for when/if RHRs become perceptually indivisible from humans. Per the argument of the previous inquiry and Turing's hypothesis, if people cannot perceptually discern the difference between an RHR and human in real-world situations, then that RHR is considered equivalent to a human. Thus, it is argued that if a person cannot determine the differences between humans and machines, then what difference does it make? Finally, 45% of participants stated that EmoAI would be useful in human society.

Interestingly, more male subjects cited applications of Emotional AI, which correlates with the outcomes of the FEA data as male subjects exhibited more emotional stress than females. Thus, it is likely male subjects felt that they could express emotions freely during HRI compared to recounting them in the HRI questionnaire. Similarly, more subjects in the higher age group cited the use of EmoAI in society. These results interlink with the biometric, AWS AI data analytics, and HRI questionnaire results as the higher age group exhibited greater levels of positive and negative emotional stimulus during HRI.

Appendix 1. All Participants Qualitative Survey Responses (Q11-25)

Euclid Q Num	Euclid Ave	Euclid Mode (Fq)	Euclid Std. Dev Conf & Error	Euclid Var	Euclid COF (α)	Baudi Q Num	Baudi Ave	Baudi Mode (Fq)	Baudi Std. Dev Con & Error	Baudi Var	Baudi COF (α)
Q11	3.8	4	s =0.61, Con 95% Err ±7.10%	S^2:0.3	0.81α	Q21	3.25	3	s =0.71, Con 95% Err ±9.66%	S^2:0.60	0.51α
Q12	3.35	4	s =0.93, Con 95% Err ±12.21%	S^2:0.87	0.78α	Q22	2.90	3	s =0.85, Con 95% Err ±12.88%	S^2:0.85	0.72α
Q13	2.6	2	s =0.94, Con 95% Err ±15.85%	S^2:0.88	0.82α	Q23	3.75	4	s =0.78, Con 95% Err ±12.88%	S^2:0.61	0.86α
Q14	2.8	3	s =0.83, Con 95% Err ±13.05%	S^2:0.69	0.80α	Q24	2.35	2	s =1.03, Con 95% Err ±19.40%	S^2:1.08	0.86α
Q15	3.15	4	s =0.98, Con 95% Err ±13.75%	S^2:0.97	0.77α	Q25	2.40	3,2	s =1.04, Con 95% Err ±19.11%	S^2:1.09	0.87α
Q16	2.9	3	s =0.85, Con 95% Err ±12.88%	S^2:0.72	0.81 α	Q26	2.9	3	s =0.64, Con 95% Err ±9.68%	S^2:0.41	0.88α
Q17	3.6	4	s =0.68, Con 95% Err ±8.29%	S^2:0.46	0.82α	Q27	3.9	4	s =0.78, Con 95% Err ±8.86%	S^2:0.62	0.86α
Q18	2.95	3	s =0.51, Con 95% Err ±7.58%	S^2:0.26	0.80α	Q28	3.4	3	s =0.82, Con 95% Err ±10.58%	S^2:0.67	0.85α
Q19	3.15	4	s =1.08, Con 95% Err ±15.16%	S^2:1.18	0.78α	Q29	3.7	5	s =1.25, Con 95% Err ±14.62%	S^2:1.56	0.87α
Q20	2.35	2,3	s =0.81, Con 95% Err ±15.16%	S^2:0.66	0.81α	Q30	2.4	2	s =1.18, Con 95% Err ±21.69%	S^2:1.41	0.85α
Q21	3.3	4,3	s =0.92, Con 95% Err ±12.26%	S^2:0.85	0.80α	Q31	3.1	3	s =1.11, Con 95% Err ±15.82%	S^2:1.25	0.86α
Q22	3.1	3	s =0.85, Con 95% Err ±12.05%	S^2:0.72	0.78α	Q32	2.7	2,3	s =1.03, Con 95% Err ±16.74%	S^2:1.06	0.86α
Q23	3.8	4	s =0.69, Con 95% Err ±8.03%	S^2:0.48	0.79α	Q33	3.6	3	s =1.94, Con 95% Err ±11.45%	S^2:0.88	0.86α
Q24	2.45	3	s =0.99, Con 95% Err ±19.86%	S^2:0.99	0.82α	Q34	2.4	3,2	s =1.04, Con 95% Err ±19.11%	S^2:1.09	0.88α
Q25	3.05	4	s =0.99, Con 95% Err ±14.35%	S^2:0.99	0.78α	Q35	2.3	2	s =0.86, Con 95% Err ±16.47%	S^2:1.74	0.85α
Total Ave	*N/A*	*N/A*	*N/A*	*N/A*	*0.81 α*	*Total Ave*	*N/A*	*N/A*	*N/A*	*N/A*	*0.87 α*

Appendix 2. Male Participants Qualitative Survey Responses (Q11-25)

Euclid Q Num	Euclid Ave	Euclid Mode (Fq)	Euclid Std. Dev Conf & Error	Euclid Var	Euclid COF (α)	Baudi Q Num	Baudi Ave	Baudi Mode (Fq)	Baudi Std. Dev Con & Error	Baudi Var	Baudi COF (α)
Q11	3.9	4	s =0.53 Con 95% Err ±8.15%	S^2:0.29	0.77α	Q21	3.1	3	s =0.75 Con 95% Err ±13.94%	S^2:0.56	0.81α
Q12	3.1	3	s =0.70 Con 95% Err ±13.94%	S^2:0.56	0.74α	Q22	3.1	3	s =0.75 Con 95% Err ±13.94%	S^2:0.56	0.85α
Q13	2.5	2	s =0.93 Con 95% Err ±21.69%	S^2:0.87	0.78α	Q23	3.7	4	s =0.90 Con 95% Err ±14.34%	S^2:0.81	0.81α
Q14	2.8	3	s =0.98 Con 95% Err ±20.58%	S^2:0.96	0.72α	Q24	2.2	2	s =1.00 Con 95% Err ±26.24%	S^2:1.01	0.84α
Q15	3.2	4	s =1.00 Con 95% Err ±18.22%	S^2:1.01	0.72α	Q25	2.2	2	s =1.00 Con 95% Err ±26.24%	S^2:1.01	0.81α
Q16	2.8	3	s =1.07 Con 95% Err ±22.62%	S^2:1.16	0.78α	Q26	2.9	3	s =0.70 Con 95% Err ±14.23%	S^2:0.49	0.84α
Q17	3.5	4	s =0.52 Con 95% Err ±8.70%	S^2:0.27	0.76α	Q27	3.9	4	s =0.53 Con 95% Err ±8.15%	S^2:0.29	0.85α
Q18	3	3	s =0.44 Con 95% Err ±8.81%	S^2:0.2	0.76α	Q28	3.5	3,4	s =0.68 Con 95% Err ±11.46%	S^2:0.47	0.82α
Q19	3.2	4	s =0.78 Con 95% Err ±14.20%	S^2:0.61	0.74α	Q29	4	4	s =1.18 Con 95% Err ±17.48%	S^2:1.4	0.83α
Q20	2.3	2,3	s =0.67 Con 95% Err ±16.86%	S^2:0.45	0.79α	Q30	2.6	3	s =1.12 Con 95% Err ±25.11%	S^2:1.24	0.82α
Q21	3.4	4	s =0.93 Con 95% Err ±15.98%	S^2:0.87	0.80α	Q31	2.9	3	s =1.13 Con 95% Err ±23.08	S^2:1.29	0.80α
Q22	3.1	3	s =0.75 Con 95% Err ±13.94%	S^2:0.56	0.75α	Q32	2.8	3	s =0.87 Con 95% Err ±18.32%	S^2:0.76	0.82α
Q23	3.9	4	s =0.70 Con 95% Err ±10.59%	S^2:0.49	0.74α	Q33	3.5	4	s =0.82 Con 95% Err ±13.67%	S^2:0.67	0.83α
Q24	2.7	3	s =0.90 Con 95% Err ±19.60%	S^2:0.81	0.80α	Q34	2.5	3	s =1.12 Con 95% Err ±26.19%	S^2:1.27	0.82α
Q25	3.1	4	s =0.87 Con 95% Err ±16.23%	S^2:0.76	0.73α	Q35	2	2	s =0.53 Con 95% Err ±15.24%	S^2:0.29	0.82α
Total Ave	*N/A*	*N/A*	*N/A*	*N/A*	*0.77α*	*Total Ave*	*N/A*	*N/A*	*N/A*	*N/A*	*0.83 α*

Appendix 3. Female Participants Qualitative Survey Responses (Q11-25)

Euclid Q Num	Euclid Ave	Euclid Mode (Fq)	Euclid Std. Dev Conf & Error	Euclid Var	Euclid COF (α)	Baudi Q Num	Baudi Ave	Baudi Mode (Fq)	Baudi Std. Dev Con & Error	Baudi Var	Baudi COF (α)
Q11	3.6	3,4	s =0.70 Con 95% Err ±12.60%	S^2:3.66	0.85α	Q21	3.3	4	s =0.70 Con 95% Err ±13.86%	S^2:3.33	0.89α
Q12	3.5	4	s =1.13 Con 95% Err ±20.77%	S^2:1.27	0.82α	Q22	2.5	2,3	s =0.88 Con 95% Err ±22.55%	S^2:0.77	0.90α
Q13	2.6	3,2	s =1 Con 95% Err ±24.50%	S^2:1	0.85α	Q23	3.7	4	s =0.66 Con 95% Err ±11.53%	S^2:0.44	0.90α
Q14	2.7	3	s =0.66 Con 95% Err ±15.68%	S^2:0.44	0.85α	Q24	2.4	2	s =1.13 Con 95% Err ±30.21%	S^2:1.27	0.86α
Q15	3	3	s =1 Con 95% Err ±21.78%	S^2:1	0.81α	Q25	2.5	2,3	s =1.13 Con 95% Err ±28.90%	S^2:1.27	0.91α
Q16	3	3	s =0.5 Con 95% Err ±10.89%	S^2:0.25	0.84α	Q26	2.8	3	s =0.60 Con 95% Err ±13.59%	S^2:0.36	0.90α
Q17	3.6	4	s =0.86 Con 95% Err ±15.43%	S^2:0.75	0.86α	Q27	3.8	4	s =1.05 Con 95% Err ±17.71%	S^2:1.11	0.88α
Q18	2.8	3	s =0.60 Con 95% Err ±13.59%	S^2:0.36	0.84α	Q28	3.2	3	s =0.97 Con 95% Err ±19.70%	S^2:0.94	0.88α
Q19	3	3	s =1.41 Con 95% Err ±30.80%	S^2:2	0.83α	Q29	3.4	4	s =1.33 Con 95% Err ±25.29%	S^2:1.77	0.90α
Q20	2.3	2	s =1 Con 95% Err ±28.00%	S^{21}:1	0.84α	Q30	2.1	2	s =1.26 Con 95% Err ±39.28%	S^2:1.61	0.88α
Q21	3.1	3	s =0.92 Con 95% Err ±19.49%	S^2:0.86	0.81α	Q31	3.3	3	s =1.11, Con 95% Err ±21.12%	S^2:	0.89α
Q22	3	3	s =1 Con 95% Err ±21.78%	S^2:1	0.82α	Q32	2.5	3,2	s =1.23 Con 95% Err ±31.60%	S^2:1.52	0.89α
Q23	3.6	4	s =0.70 Con 95% Err ±12.60%	S^2:0.5	0.83α	Q33	3.6	4	s =1.11 Con 95% Err ±19.92%	S^2:3.66	0.88α
Q24	2.1	2	s =1.05 Con 95% Err ±32.62%	S^2:1.11	0.84α	Q34	2.2	2	s =0.97 Con 95% Err ±28.57%	S^2:0.94	0.91α
Q25	2.8	3	s =1.16 Con 95% Err ±26.38%	S^2:1.36	0.83α	Q35	2.5	2,3	s =1.13 Con 95% Err ±28.90%	S^2:1.27	0.88α
Total Ave	*N/A*	*N/A*	*N/A*	*N/A*	*0.84α*	*Total Ave*	*N/A*	*N/A*	*N/A*	*N/A*	0.90 α

Appendix 4. Ages (18-27) Participants Qualitative Survey Responses (Q11-25)

Euclid Q Num	Euclid Ave	Euclid Mode (Fq)	Euclid Std. Dev Conf & Error	Euclid Var	Euclid COF (α)	Baudi Q Num	Baudi Ave	Baudi Mode (Fq)	Baudi Std. Dev Con & Error	Baudi Var	Baudi COF (α)
Q11	3.7	4	s =0.48 Con 95% Err ±8.09%	S^2:0.23	0.91α	Q21	3.2	3	s =0.62 Con 95% Err ±12.25%	S^2:0.4	0.91α
Q12	3.1	3	s =1.10 Con 95% Err ±22.00%	S^2:1.21	0.90α	Q22	2.8	2	s =0.91 Con 95% Err ±20.34%	S^2:0.84	0.92α
Q13	2.7	3	s =0.82 Con 95% Err ±18.90%	S^2:0.67	0.92α	Q23	3.9	4	s =0.99 Con 95% Err ±15.80%	S^2:0.98	0.91α
Q14	2.8	3	s =1.03 Con 95% Err ±22.86%	S^2:1.06	0.91α	Q24	2.2	2	s =1.13 Con 95% Err ±31.98%	S^2:1.28	0.91α
Q15	3.1	3	s =1.10 Con 95% Err ±22.00%	S^2:1.21	0.89α	Q25	2	2	s =0.94 Con 95% Err ±29.22%	S^2:0.88	0.91α
Q16	2.7	3	s =0.82 Con 95% Err ±18.90%	S^2:0.67	0.91α	Q26	3	3	s =0 Con 95% Err ±0%	S^2:0	0.91α
Q17	3.8	4	s =0.42 Con 95% Err ±6.88%	S^2:0.17	0.91α	Q27	4	4	s =0.94 Con 95% Err ±14.61%	S^2:0.88	0.90α
Q18	3	3	s =0.47 Con 95% Err ±7.94%	S^2:0.22	0.91α	Q28	3.5	3,4	s =0.97 Con 95% Err ±17.21%	S^2:0.94	0.91α
Q19	2.8	3	s =1.22 Con 95% Err ±27.21%	S^2:1.51	0.90α	Q29	3.9	4	s =1.19 Con 95% Err ±19.03%	S^2:1.43	0.90α
Q20	2.1	2	s =0.73 Con 95% Err ±21.78%	S^2:0.54	0.91α	Q30	2.5	2,3	s =1.26 Con 95% Err ±31.47%	S^2:1.61	0.91α
Q21	3	3	s =1.05 Con 95% Err ±21.78%	S^2:1.11	0.90α	Q31	3.2	3	s =1.39 Con 95% Err ±27.09%	S^2:1.95	0.90α
Q22	3	3	s =1.05 Con 95% Err ±21.78%	S^2:1.11	0.89α	Q32	2.5	3,2	s =0.97 Con 95% Err ±24.09%	S^2:0.94	0.90α
Q23	3.9	4	s =0.73 Con 95% Err ±11.73%	S^2:0.54	0.90α	Q33	3.8	4	s =1.03 Con 95% Err ±16.85%	S^2:1.06	0.91α
Q24	2.5	3	s =0.84 Con 95% Err ±21.07%	S^2:0.72	0.91α	Q34	2.5	3	s =1.09 Con 95% Err ±26.78%	S^2:1.16	0.90α
Q25	3	3	s =1.15 Con 95% Err ±23.86%	S^2:1.33	0.89α	Q35	2.2	2	s =0.91 Con 95% Err ±25.89%	S^2:0.84	1α
Total Ave	*N/A*		*N/A*	*N/A*	*0.91α*	*Total Ave*	*N/A*	*N/A*	*N/A*	*N/A*	0.91α

Appendix 5. Ages (27-55) Participants Qualitative Survey Responses (Q11-25)

Euclid Q Num	Euclid Ave	Euclid Mode (Fq)	Euclid Std. Dev Conf & Error	Euclid Var	Euclid COF (α)	Baudi Q Num	Baudi Ave	Baudi Mode (Fq)	Baudi Std. Dev Con & Error	Baudi Var	Baudi COF (α)
Q11	3.9	4	s =0.73 Con 95% Err ±11.73%	S^2:0.54	0.77α	Q21	3.3	3	s =0.82 Con 95% Err ±15.46%	S^2:0.67	0.82α
Q12	3.6	3,4	s =0.69 Con 95% Err ±12.04%	S^2:0.48	0.79α	Q22	3	3	s =0.81 Con 95% Err ±16.87%	S^2:0.66	0.83α
Q13	2.5	2,3	s =1/08 Con 95% Err ±26.78%	S^2:1.16	0.77α	Q23	3.6	4	s =0/51 Con 95% Err ±8.89%	S^2:0.26	0.84α
Q14	2.8	3	s =0.63 Con 95% Err ±14.00%	S^2:0.4	0.80α	Q24	2.5	3,2	s =0.97 Con 95% Err ±24.09%	S^2:0.94	0.81α
Q15	3.2	3	s =0.91 Con 95% Err ±17.80%	S^2:0.84	0.79α	Q25	2.8	3	s =1.03 Con 95% Err ±22.86%	S^2:1.06	0.83α
Q16	3.1	3	s =0.87 Con 95% Err ±17.51%	S^2:0.76	0.83α	Q26	2.8	3	s =0.91 Con 95% Err ±20.34%	S^2:0.84	0.82α
Q17	3.4	3	s =0.84 Con 95% Err ±15.37%	S^2:0.71	0.79α	Q27	3.8	3	s = 0.63 Con 95% Err ±10.32%	S^2:0.63	0.81α
Q18	2.9	3	s =0.56 Con 95% Err ±12.13%	S^2:0.32	0.81α	Q28	3.3	3	s =0.67 Con 95% Err ±12.68%	S^2:0.45	0.82α
Q19	3.5	3,4	s =0.84 Con 95% Err ±15.05%	S^2:0.72	0.82α	Q29	3.6	4	s =1.34 Con 95% Err ±23.24%	S^2:1.82	0.83α
Q20	2.6	2	s =0.84 Con 95% Err ±20.10%	S^2:0.71	0.79α	Q30	2.3	2	s =1.15 Con 95% Err ±31.25%	S^2:1.34	0.83α
Q21	3.6	3	s =0.69 Con 95% Err ±12.04%	S^2:0.48	0.82α	Q31	3	3	s = 0.81 Con 95% Err ±16.87%	S^2:0.66	0.80α
Q22	3.2	3	s =0.63 Con 95% Err ±12.25%	S^2:0.4	0.80α	Q32	2.9	3	s =1.10 Con 95% Err ±23.52%	S^2:1.21	0.82α
Q23	3.7	4	s =0.67 Con 95% Err ±11.31%	S^2:0.45	0.81α	Q33	3.4	4	s =0.94 Con 95% Err ±15.37%	S^2:0.71	0.81α
Q24	2.4	2	s =1.17 Con 95% Err ±30.31%	S^2:1.37	0.77α	Q34	2.3	2	s =1.05 Con 95% Err ±28.55%	S^2:1.12	0.84α
Q25	3.1	3	s =0.87 Con 95% Err 17.51%	S^2:0.76	0.78α	Q35	2.4	2	s =0.84 Con 95% Err ±21.78%	S^2:0.71	0.81α
Total Ave	*N/A*	*N/A*	*N/A*	*N/A*	*0.81α*	*Total Ave*	*N/A*	*N/A*	*N/A*	*N/A*	*0.83α*

References

[1] Hanson. D, Olney. A, Prilliman. S, Mathews. E, Zielke. M, Hammons. D, Fernandez. R, & Stephanou .E. (2005). Upending the Uncanny Valley. *AAAI*. Corpus ID: 34117944

[2] Nishio. T, Yoshikawa. Y. Ogawa, Ishiguro. H (2019). Development of an Effective Information Media Using Two Android Robots. *Applied Sciences*. 9. 3442. DOI:10.3390/app9173442.

[3] Ishi. T, Minato. T, Ishiguro. H (2017), Motion Analysis in Vocalized Surprise Expressions and Motion Generation in Android Robots, *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1748-1754, July 2017. DOI: 10.1109/LRA.2017.2700941

[4] Goertzel. B (2019) Why We Need Humanoid Robots Instead of Faceless Kiosks. Retrieved: https://www.forbes.com/sites/forbestechcouncil/2019/12/27/why-we-need-humanoid-robot s-instead-of-faceless-kiosks/#4dfc88192e21. Accessed: 15.02.20

[5] Todo. T (2018) SEER: simulative emotional expression robot, SIGGRAPH '18: ACM SIGGRAPH 2018 *Emerging Technologies*, 15, pp.1–2. DOI:10.1145/3214907.3214921

[6] Doering. D, Glas. H and Ishiguro. H (2019) Modelling Interaction Structure for Robot Imitation Learning of Human Social Behaviour, in *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 219-231, June 2019. DOI: 10.1109/THMS.2019.2895753

[7] Haque. Z and Nasir. J (2007) Investigating the Uncanny Valley and human desires for interactive robots, *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Sanya, 2007, pp. 2228-2233. DOI: 10.1109/ROBIO.2007.4522516

[8] Yamazaki. R, Nishio. S, Ogawa. K, Matsumura. K, Minato. T, Ishiguro. H, Fujinami. T, & Nishikawa. M. (2013). Promoting Socialization of schoolchildren using a teleoperated Android. *Int. J. Humanoid Robotics, 10*. DOI: 10.1142/S0219843613500072

[9] Camble. A, Frances. J (2018) Talking about disfigurement. Retrieved: https://www.sec-ed.co.uk/best-practice/talking-about-disfigurement-1/. Accessed: 18.04.20

[10] Waskul. D & Vannini, P. (2012). Body/Embodiment: Symbolic Interaction and the Sociology of the Body. pp. 1-297. ISBN-13: 978-0754647263

[11] Whitaker. I, Shokrollahi. K, Dickson. W (2019) Burns, Oxford Specialist Handbooks in Surgery. ISBN:0191503584, 9780191503580.

[12] Harcourt. D, Rumsey. N (2008) Psychology and Visible Difference. Retrieved: https://thepsychologist.bps.org.uk/volume-21/edition-6/psychology-and-visible-differe nce. Accessed.13.03.20

[13] Riek. T, Rabinowitch. B. Chakrabarti. C and Robinson. P (2009) How anthropomorphism affects empathy toward robots, 2009 4th ACM/*IEEE International Conference on Human-Robot Interaction* (HRI), CA, pp. 245-246. DOI: 10.1145/1514095.1514158

[14] Kleinman. Z (2018) CES 2018: A clunky chat with Sophia, the robot. Retrieved: https://www.bbc.co.uk/news/technology-42616687. Accessed: 19.04.20

[15] Chung. S (2018) Meet Sophia: The robot who laughs, smiles, and frowns just like us. Retrieved:https://edition.cnn.com/style/article/sophia-robot-artificial-intelligence-smart -creativity/index.html. Accessed. 13.03.20

[16] Pessoa. L (2019) Intelligent architectures for robotics: The merging of cognition and emotion. *Physics of Life*, 31 (2019): pp.157–170. DOI:10.1016/j.plrev.2019.04.009

[17] Mori. M (1970). The Uncanny Valley. Energy, Issue 7, pp.33-35. DOI: 10.1109/MRA.2012.2192811

[18]   Hanson. D (2015). Expanding the aesthetic possibilities for humanoid robots. In IEEE-RAS international conference on humanoid robots, Tsukuba, Japan. Retrieved: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.2518&rep=rep1&type=pdf. Accessed 27.03.19.

[19] Hanson. D (2006). Exploring the Aesthetic Range for Humanoid Robots. Corpus ID: 17275411

[20] Tromp. J, Bullock. A, Steed. A, Sadagic. A, Slater. M, Frécon. E. (1998). Small-group behaviour experiments in the COVEN Project. *IEEE Computer Graphics and Application*. 18. pp,53-63. DOI:10.1109/38.734980

[21] Garau. M, & Weirich. D. (2003). The Impact of Avatar Fidelity on Social Interaction in Virtual Environments. Corpus ID: 141182802

[22] Nass. C, Isbister. K, and Lee. J (2000). Truth is a beauty: Researching conversational agents. *Embodied conversational agents,* MIT Press, pp. 374-402. doi/10.5555/371552.371575

[23] Bar-Cohen. Y, Hanson. D, Marom. A (2009) Introduction. In: The Coming Robot Revolution. Springer, New York, NY. DOI: 10.1007/978-0-387-85349-9

[24] Schoenherr. J. R & Burleigh. T. J (2015). Uncanny sociocultural categories. *Frontiers in psychology*, 5, 1456. DOI:10.3389/fpsyg.2014.01456

[25] Browne. W & Kawamura. K (2005). Uncanny valley revisited. *IEEExplore*, pp.151 - 157. DOI: 10.1109/ROMAN.2005.1513772.

[26] Barthelmess. U, & Furbach. U (2013). Do we need Asimov's Laws? ArXiv, abs/1405.0961.

[27] Vallverdu. J (2011). The Eastern Construction of the Artificial Mind. *Enrahonar*. 47. Pp.233-253. DOI: https://philpapers.org/rec/VALTEC

[28] Cheetham. M (2017). Editorial: The Uncanny Valley Hypothesis and beyond. *Frontiers in psychology*, 8, p.1738. DOI/10.3389/fpsyg.2017.01738

[29] Brink. K & Wellman. H (2017) Hi, Robot: Adults, Children and The Uncanny Valley. Retrieved: https://www.npr.org/sections/13.7/2017/12/16/563075762/hi-robot-adults-children-and-the-uncanny-valley?t=1582650482901. Accessed: 25.02.20

[30] Brink. K & Gray. K & Wellman. H (2017). Creepiness Creeps In: Uncanny Valley Feelings Are Acquired in Childhood. *Child Development*. DOI: 90. 10.1111/cdev.12999.

[31] Feng. S, Wang. X, Wang. Q, Fang. J, Wu. Y, Yi. L, & Wei. K (2018). The uncanny valley effect in typically developing children and its absence in children with autism spectrum disorders. *PloS one*, 13(11), e0206343. DOI: 10.1371/journal.pone.0 206343

[32] Maya. M, Reichling. D, Lunardini. F, Geminiani. A, Antonietti. A, Ruijten. P, Levitan. C, Nave. G, Manfredi. D, Bessette-Symons. B, Szuts. A, Aczel. B. (2019). Uncanny but not confusing: Multisite study of perceptual category confusion in the Uncanny Valley. *Computers in Human Behaviour.* 103. DOI:10.1016/j.chb.2019.08.029.

[33] Lay. S, Brace. N, Pike. G, & Pollick. F (2016). Circling Around the Uncanny Valley: Design Principles for Research Into the Relation Between Human-likeness and Eeriness. *I-Perception*. DOI:10.1177/2041669516681309

[34] Tinwell. A & Sloan. R.J (2014). Children's perception of uncanny human-like virtual characters. *Comput. Hum. Behav*, 36, pp.286-296. DOI: 10.1016/j.chb.2014.03.073

[35] Tung. F (2016) Child Perception of Humanoid Robot Appearance and Behavior*, International Journal of Human-Computer Interaction*, 32:6, 493-502, DOI: 10.1080/10447318.2016.1172808

[36] Towers-Clark. C (2019) Keep The Robot In The Cage—How Effective (And Safe) Are Co-Bots? Retrieved:www.forbes.com/sites/charlestowersclark/2019/09/11/keep-the-robot-in-the-cagehow-effet tive--safe-are-co-bots/#3279006415cc. Accessed: 05.04.20

[37] Patel. V. N (2020) An emotionally intelligent AI could support astronauts on a trip to Mars. Retrieved:www.technologyreview.com/s/615044/an-emotionally-intelligent-ai-could-support-astronauts-on-a-trip-to-mars/. Accessed:15.02.20

[38] McDuff. D, Kapoor. A (2019) Toward Emotionally Intelligent Artificial Intelligence. Retrieved:www.microsoft.com/en-us/research/blog/toward-emotionally-intelligent-artificial-intelligence/. Accessed: 21.02.20

[39] Alasaarela. M (2017) The Rise of Emotionally Intelligent AI. Retrieved: https://machinelearnings.co/the-rise-of-emotionally-intelligent-ai-fb9a814a630e. Accessed: 22.02.20

[40] Dewan. K (2019) The Rise of AI Makes Emotional Intelligence More Important. Retrieved: https://www.entrepreneur.com/article/340637. Accessed:22.02.20

[41] Jarboe. J (2018) What Is Artificial Emotional Intelligence & How Does Emotion AI Work? Retrieved: https://www.searchenginejournal.com/what-is-artificial-emotional-intelligence/255769/. Accessed:22.02.20

[42] Fatemi. F (2018) Why EQ + AI Is A Recipe For Success. Retrieved: https://www.forbes.com/sites/falonfatemi/2018/05/30/why-eq-ai-is-a-recipe-for-success/#61 b01ae71005. Accessed:22.02.20

[43] Nordli. B (2019) Cogito, Whose AI Interprets Human Emotions, Adds $20M. Retrieved: https://www.builtinboston.com/2019/09/05/cogito-raises-20m-funding. Accessed: 21.02.20

[44] Field. H (2019) Technologists Are Creating Artificial Intelligence to Help Us Tap Into Our Humanity. Here is How (and Why). Retrieved: https://www.entrepre neur.com/article/340207. Accessed. 13.03.20

[45] Lewis. A, Guzman. L, Schmidt. T (2019) Automation, Journalism, and Human-Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News, *Digital Journalism*, 7:4, pp.409-427, DOI: 10.1080/21670811.2019.1577147

[46] Wiltz. C (2019) What is the state of Emotional Artificial Intelligence*? Drive World,* pp.234-255. Retrieved: www.designnews.com/electronics-test/whats-state-emotional-ai/20315341406148 2. Accessed:23.03.20

[47] Williams. D (2018) The Rise of Emotional Robots. Retrieved: https://www.sapiens.org/technology/emotional-intelligence-robots/. Accessed: 19.03.20

[48] Streeter. B (2019) Seriously Successful Results From HSBC Bank's Branch Robot Rollout. Retrieved:https://thefinancialbrand.com/84245/hsbc-banks-branch-robot-pepper-digital-transformation-phygital/. Accessed:21.03.20

[49] Pandey. A, Gelin. R (2018). A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind. *IEEE Robotics & Automation Magazine.* PP. 1-1. DOI:10.1109/MRA.2018.2833157.

[50] Starr. M (2015) Caring Pepper robot hits the market, sells out in a minute. Retrieved: https://www.cnet.com/news/caring-pepper-robot-hits-the-market-sells-out-in-a-minute/. Accessed: 21.02.20

[51] Goodman. J (2018) Robotics roundtable: Should robots look like humans, or like machines? Retrieved: https://internetofbusiness.com/should-robots-look-like-humans-or-like-machines/. Accessed:21.02.20

[52] Simon. M (2018a) The Creepy-Cute Robot that Picks Peppers With its Face. Retrieved: www.wired.com/story/the-creepy-cute-robot-that-picks-peppers/. Accessed:22.02.20

[53] Glaser. A (2016) Pepper, the Emotional Robot, Learns How to Feel Like an American. Retrieved: www.wired.com/2016/06/pepper-emotional-robot-learns-feel-like-american/. Accessed:21.02.20

[54] Hawes. N (2014) Pepper the emotional robot may just make you angry. Retrieved: https://theconversation.com/pepper-the-emotional-robot-may-just-make-you-angry-27655. Accessed: 21.02.20

[55] Hu, M. S, Maan, Z. N, Wu J.C, Rennert. R. C, Hong. W. X, Lai. T. S. (2014). Tissue engineering and regenerative repair in wound healing. Ann. Biomed. Eng. 42, 1494–1507. 10.1007/s10439-014-1010-z

[56] Guo. J, Xiang. C, Conn. R & Rossiter. J (2019). All-Soft Skin-Like Structures for Robotic Locomotion and Transportation. *Soft robotics*. DOI:10.1089/soro.2019.0059

[57] Williams. R (2019) Researchers create a first autonomous humanoid robot with full-body artificial skin. Retrieved: https://inews.co.uk/news/technology/researchers-autonomous-humanoid-robot-artificial-skin-sense-805511. Accessed: 15.02.20

[58] Zilnik. A (2016) Designing a Chatbot's Personality. Retrieved: chatbotsmagazine.com/designing-a-chatbots-personality-52dcf1f4df7d. Accessed: 22.02.20

[59] Pavlus. J (2016) The Next Phase Of UX: Designing Chatbot Personalities. Retrieved: https://www.fastcompany.com/3054934/the-next-phase-of-ux-designing-chatbot-personalities. Accessed. 19.03.20

[60] Marr. B (2019) Why AI And Chatbots Need Personality. Retrieved: https://www.forbes.com/sites/bernardmarr/2019/08/02/why-ai-and-chatbots-need-personality/#fc7224214f84. Accessed: 23.04.20

[61] Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. *New York: Basic Books*. pp.55-78. ISBN-13: 978-0006862901

[62] Moltzau. A (2019) AI and Nine Forms of Intelligence. Retrieved: https://towardsdatascience.com/ai-and-nine-forms-of-intelligence-5cf587547731. Accessed: 21.02.20

[63] Duffy, B.R (2001). Towards Social Intelligence in Autonomous Robotics: A Review. *International journal of advanced robotic systems*. DOI:10.5772/5632

[64] Roberts. J & Dennison. J (2015) The Photobiology of Lutein and Zeaxanthin in the Eye, *Journal of Ophthalmology,* DOI 10.1155/2015/687173

[65] Davis. K, Christodoulou. A, Seider. S, Gardner. H (2011) The Theory of Multiple Intelligences. In: Sternberg RJ, Kaufman SB Cambridge Handbook of Intelligence. New York: Cambridge University Press; 2011. pp. 485-503.

[66] Tirri. K, Nokelainen. P, & Komulainen. E. (2013). Multiple Intelligences: Can they be measured? Psychological Test and Assessment Modelling, *Apa Psycnet,* 55(4), pp.438-461. DOI: 2014-06897-007

[67] Núñez-Cárdenas. F.D, Camacho. J.H, Tomás-Mariano. V.T, & Redondo. A.M. (2015). Application of Data Mining to describe Multiple Intelligences in University Students. *IJCOPI, 6*, 20-30. Corpus ID: 27249996

[68] Kretzschmar. K, Tyroll. H, Pavarini. G, Manzini. A, & Singh. I (2019). Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents in Mental Health Support. *Biomedical Informatics Insights*. DOI:10.1177/1178222619829083

[69] Edwards .D (2019) The HeartMath coherence model: implications and challenges for artificial intelligence and robotics. *AI & Soc* 34, 899–905. DOI:10.1007/s0 0146-018-0834-8

[70] Maybury. M. T (1990), The mind matters artificial intelligence and its societal implications, *IEEE Technology and Society Magazine*, vol. 9, no. 2, pp. 7-15, June-July 1990. DOI: 10.1109/44.56350.

[71] Hiren. S (2019) Global Humanoid Robot Market Evolving Industry Trends and Key Insights by 2024. Retrieved: http://zmrnewsmagazine.com/18125/global-humanoid-robot-market-evolving-industry-trends-and-key-insights-by-2024/. Accessed 21.05.19.

[72] Zlotowski. J, Proudfoot. D, & Bartneck. C. (2013). More Human Than Human: Does the Uncanny Curve Matter? Proceedings of the HRI2013

[73] Goodrich. M, Lee. J, Kirlik. A (2013). Multitasking and Multi-Robot Management. In The *Oxford Handbook of Cognitive Engineering.* : Oxford University Press. pp.453-556. DOI: 10.1093/oxfordhb/9780199757183.013.0025.

[74] Hoskins. T (2016) Robot factories could threaten the jobs of millions of garment workers. Retrieved:https://www.theguardian.com/sustainablebusiness/2016/jul/16/robot-factor ies-threaten-jobs-millions-garment-workers-south-east-asia-women. Accessed:19.02.20

[75] Ghose. T (2016) Robots Could Hack Turing Test by Keeping Silent. Retrieved: https://www.scientificamerican.com/article/robots-could-hack-turing-test-by-keeping-silent/. Accessed: 25.02.20

[76] Solis. B (2019) Now Hiring: Robots, Please Apply Within. Retrieved: www.forbes.com/sites/briansolis/2019/09/17/now-hiring-robots-please-apply-within/. Accessed:11.02.20

[77] Shell. E. R (2018) AI and Automation Will Replace Most Human Workers Because They Do Not Have to Be Perfect, Just Better Than You. Retrieved: www.new sweek.com/2018/11/30/ai-and-automation-will-replace-most-human-workers-because-they-dont-have-be-1225552.html. Accessed: 11.02.20

[78] Burkhardt. M (2019) Living in a Robot Society: Perspective of a World-Leading Roboticist. Retrieved://towardsdatascience.com/living-in-a-robot-society-a-perspective-of-a-world-leading-robotocist-51175fb455e3. Accessed:11.02.20

[79] Yamaguchi. T (2019) Japan's robotic future, *Critical Asian Studies*, 51:1, 134-140, DOI: 10.1080/14672715.2018.1558085

[80] Chatila. R (2018) Inclusion of Humanoid Robots in Human Society: Ethical Issues. *Humanoid Robotics, Springer,* Dordrecht. DOI: 10.1007/978-94-007-7194-9_147-1

[81] Dautenhahn. K (2018) Robots and Us - Useful Roles of Robots in Human Society. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18). Association for Computing Machinery, New York, NY, USA, 1. DOI:10.1145/3171221.3171222

[82] Yan. J, Wang. Z, & Yan. Y (2014). Humanoid Robot Head Design Based on Uncanny Valley and FACS. *J. Robotics, 2014*, pp.208924:1-208924:5. DOI: 10.1155/2014/208924

[83] Sakamoto. D, Kanda. T, Ono. T, Ishiguro. H, & Hagita. N. (2007). Android as a telecommunication medium with a human-like presence. *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 193-200. DOI:10.1145/1228716.1228743

[84] Liu. Y, Shi. L. Liu. L, Zhang. Z & Jinsong. L (2008). Inflated dielectric elastomer actuator for eyeball's movements: Fabrication, Analysis and Experiments. Proceedings of *SPIE*. 6927. DOI: 10.1117/12.775946.

[85] Harkness. K & McKelvie. A (2014). The Measurement of Uncertainty in Caregivers of Patients With Heart Failure. Journal of Nursing Measurement, 21.1 (2013): 23-42. Journal of Nursing Measurement. 21. 23. DOI: 10.1891/1061-3749.21.1.23.

[86] Payakpate, Janjira. (2020). Knowledge management platform for promoting sustainable energy technologies in rural Thai communities. Knowledge management platform for promoting sustainable energy technologies in rural Thai communities. *PhD thesis, Murdoch University*. DOI:researchrepository.murdoch.edu.au/id/eprint/426/

[87] Ishiguro. H (2005). Android science: Toward a new cross-interdisciplinary framework. *Journal of Computer Science*, Corpus ID: 6105971

[88] Harnad. S (1992) The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion. *SIGART Bulletin* 3(4) (October 1992) pp. 9 - 10. [Appears preceded by an Editorial on the Turing Test by Lewis Johnson, pp. 7 - 9, and followed by another commentary by Stuart Shapiro, p. 10. DOI:10.1145/141420.141422

[89] Harnad. S, Scherzera . P (2008) First, scale up to the robotic Turing test, then worry about feeling. *Artificial Intelligence in Medicine* Volume 44, Issue 2, October 2008, Pages 83-89. DOI:10.1016/j.artmed.2008.08.008

[90] Schweizer. P (1998). The Truly Total Turing Test. Minds Mach. 8, 2 (May 1998), 263–272. DOI:10.1023/A:1008229619541

[91] Gay. L (1987) *Educational Research: Competencies for Analysis and Application, the 3rd edition.* (Publisher: Merrill Publishing Company (1987) Columbus Ohio, Chapt (2-4) the US) ISBN-13: 978-0675205061

[92] Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Ment Health*. 2015;42(5):533–544. DOI:10.1007/s10488-013-0528-y

[93] Ivaldi. S, Anzalone. S & Rousseau. W, Sigaud. O, Chetouani. M (2014). Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. *Frontiers in neurorobotics*. 8. 5. DOI:10.3389/fnbot.2014.00005.

[94] Mavridis. N, Katsaiti. Naef. M, Falasi. A, Nuaimi. A, Araifi. H, Kitbi. A (2012). Opinions and attitudes toward humanoid robots in the Middle East. AI & SOCIETY. 27. 517-534. DOI: 10.1007/s00146-011-0370-2.

[95] Birmingham. E, Svärd. J, Kanan. C, & Fischer. H. (2018). Exploring emotional expression recognition in aging adults using the Moving Window Technique. PloS one, 13(10), e0205341. https://doi.org/10.1371/journal.pone.0205341

[96] Vaidyam. A & Torous. J (2019) Chatbots: What Are They and Why Care? Retrieved: https://www.psychiatrictimes.com/telepsychiatry/chatbots-what-are-they-and-why-care/page/0/1. Accessed:11.04.20

[97] Vaidyam. A.N, Wisniewski. H, Halamka. J.D, Kashavan. M.S, & Torous. J.B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Journal of Psychiatry*, 64, pp.456 - 464. DOI: 10.1177/0706743719828977

[98] Gardiner .M, McCue .D, Negash. L, Cheng. T, White .F, Yinusa-Nyahkoon .S, Jack .W & Bickmore .W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations. *Patient education and counselling, 100 9*, pp.1720-1729. DOI: 10.1016/j.pec.2017.04.015