“Am I now masking, even with a bot?”: Pseudo-Social Bonding between Autistic Adults and LLMs

Ellie Cogley
Cardiff University, UK
CogleyE@cardiff.ac.uk

Timothy Matthews
University of Greenwich, UK
T.Matthews@greenwich.ac.uk

Daniel J. Finnegan
Cardiff University, UK
FinneganD@cardiff.ac.uk

Pseudo-social bonding – the formation of one-sided, socially meaningful bonds with entities that cannot meaningfully reciprocate them – is increasingly shaped by AI chatbots’ capacity to simulate social reciprocity in ways that convincingly mirror human interaction. Autistic adults exhibit anthropomorphic tendencies more frequently and into adulthood than non-autistic individuals. This raises questions about how such tendencies shape the pseudo-social bonds autistic adults form with AI chatbots, and the associated risks of harm. We present a seven-day diary study exploring pseudo-social bonding among 25 autistic and allistic adults interacting daily with a custom chatbot, using thematic analysis of open-ended diary and exit survey responses. Bonding patterns differed between groups, though relational, moral treatment of the chatbot itself was broadly shared. We identify several autism-specific interaction patterns, including masking, infodumping, and variation in trust and forgiveness towards the chatbot, and discuss implications for future chatbot design.

Human-AI Interaction; Pseudo-Social Bonding; Autism; Diary Study; Chatbot

1. INTRODUCTION

Social bonding is the foundation for forming relationships between individuals, functioning both as an evolutionary survival strategy (Seyfarth and Cheney 2012; Agashe et al. 2021) and a psychological mechanism for social connection (Baumeister and Leary 1995). Modern AI chatbots simulate this behaviour, a form of what we call pseudo-social bonding, by responding, adapting, and demonstrating social reciprocity in ways that convincingly mirror human interaction. The capacity of chatbots to induce pseudo-social bonding was recognised as early as the 1960s with ELIZA, a program with no ’understanding’ of the words it produced (Weizenbaum 1976), instead only regurgitating the user’s words back at them, posed as questions.

Increasingly, research has detailed chatbot-based incidents involving social isolation, self-injury, and suicide, with significant relationships and pseudo-social bonds to the chatbot being a common factor (De Choudhury et al. 2023; The Responsible AI Collaborative 2026). Understanding how these bonds manifest and which users are most susceptible to the harms they entail is essential to designing chatbots and related tools that support and protect vulnerable users rather than exploiting or ignoring them.

Autism, more formally known as Autism Spectrum Disorder (ASD), is a neurodevelopmental disorder clinically categorised by persistent challenges in social communication and interactions, alongside repetitive or restrictive behaviours, interests or activities (American Psychiatric Association 2022). It affects approximately 1% of the general population in England (Treise et al. 2021), though reliable UK-wide figures are difficult to establish and this is widely considered an underestimate.

Anthropomorphism is the universally experienced attribution of human-like qualities to non-human entities. While both autistic and non-autistic people tend to anthropomorphise objects in childhood (Gao et al. 2024), autistic individuals exhibit this tendency more frequently and later into adulthood (White and Remington 2019; Clutterbuck et al. 2022). Given that AI chatbots provide convincing displays of social reciprocity, autistic adults’ heightened and more persistent tendency to anthropomorphise may shape how pseudo-social bonds form. Recent HCI research has primarily focused on assistive, in a functional sense, and accessibility-oriented

© Cogley et al. Published by BCS Learning and Development Ltd.

chatbot tools for autistic individuals (Iannone and Giansanti 2024; Cha et al. 2021). However, these may under-represent how autistic individuals engage with mainstream, general-purpose AI chatbots and the pseudo-social bonds formed through such everyday use. To address this gap, we contribute a replicable diary-study methodology, qualitative findings on pseudo-social bonding in autistic adults and implications for chatbot design, posing the following research questions:

RQ1: How, if at all, does pseudo-social bonding with AI chatbots differ between autistic and non-autistic adults?

RQ2: What factors shape the nature of pseudo-social bonding with AI chatbots in autistic adults?

2. BACKGROUND

Pseudo-social bonding refers to the tendency to form one-sided, socially meaningful bonds with entities that cannot meaningfully reciprocate them. Unlike attachment, which presumes some degree of mutual responsiveness, or parasocial interaction, which presumes complete one-sidedness (Horton and Wohl 1956), pseudo-social bonds can persist even when the bonded party is fully aware that reciprocity is absent or simulated. Modern AI chatbots build on this phenomenon, convincingly simulating reciprocity through responsive, adaptive interactions that can persist and deepen over repeated engagements.

The extent and presentation of such bonds are not uniform across individuals. They are shaped by environmental and cultural factors known to influence anthropomorphism and bond formation more broadly (Spatola et al. 2022). Cultural norms, for instance, vary in the degree to which treating a non-human entity with normative regard is considered acceptable, shaping how readily these bonds form and are expressed

Pseudo-social bonding with AI is driven primarily by linguistic mimicry and mirroring, which large language models now perform to a convincing standard (Kim 2026). Analysing and adapting to users' vocabulary and conversational nuances can allow chatbots to leave users feeling 'truly understood', even in the absence of genuine reciprocal understanding. This perceived validation can create a positive feedback loop of increasing self-disclosure, subsequent chatbot learning and adaptation, and, in turn, deepening the illusion of reciprocity and bonding. This loop often motivates users to create customised chatbots or character companions, as the chatbot comes to be perceived as 'knowing' the user's needs. Whilst many such bonds remain functional and assistant or pet-like in nature, others may intensify into more compulsive forms, such as escapist role-play or social substitution, in which the chatbot is used in lieu of human contact or support (Shen et al. 2026).

AI chatbots pose significant psychological risks, including harmful advice, psychosis, emotional dependency, and, in extreme cases, contributing to self-injury and suicide (De Choudhury et al. 2023; The Responsible AI Collaborative 2026; Olsen et al. 2026). These risks are exacerbated by weak guardrails and relaxed or absent age verifications on major platforms, including Replika and Chai AI, (Replika 2026; Chai AI 2026). Such harms disproportionately affect vulnerable populations, including those with pre-existing mental health difficulties or cognitive disabilities, due to factors such as greater trust, greater susceptibility to authoritative voices and elevated risk of delusional thinking (Chapman et al. 2025; Olsen et al. 2026). Given challenges such as emotional regulation, which are associated with autism, autistic individuals may be particularly at-risk (Mazefsky et al. 2013).

3. STUDY DESIGN & PROCEDURE

30 participants were recruited via broadcast-style posts and stratified by self-reported autism status (autistic, suspected autism, and no suspected autism), with 10 participants randomly sampled from each group (n = 10 per group). To mitigate environmental factors known to influence attachment and anthropomorphism (Spatola et al. 2022), eligibility required prior chatbot use, fluent English, UK residency and identification with British culture and norms. 5 participants (3 women, 2 men) were removed for failing to comply with the study procedure, resulting in a final sample of 25 participants (14 men, $M_{age} = 27.57, SD = 11.35$ and 11 women $M_{age} = 21.68, SD = 3.74$; age brackets ranging from 18-21 to 50-59).

Participants were grouped along two dimensions, summarised in Table 1: (1) *self-declared autism status*, capturing lived experience alongside clinical diagnosis given the barriers many autistic individuals face in obtaining one and (2) *trait-based groups*, split by the AQ-10 screening cut off ($\geq$ 6) (Allison et al. 2012), providing an independent trait based grouping for comparison. Table 2 shows the overlap between self-declared autism status and AQ-10 trait-based grouping. We adopt the trait-based grouping as our primary lens for RQ1/RQ2, as it captures autistic traits regardless of diagnosis, and enables comparability with prior work (Gao et al. 2024). Self-declared status is retained for contextual and recruitment grouping purposes.

Table 1: *Summary of Participant Groupings*

Grouping	Category	N	%
Autism Status	Diagnosed	10	40%
	Suspected	8	32%
	Non-autistic	7	28%
Trait-Based	High ($\geq$6)	16	64%
	Low ($<$6)	9	36%

Table 2: *Cross-tab of Self-Declared Status vs. AQ-10 Trait Grouping*

Self-declared Status	High Traits ($\geq$6)	Low Traits ($<$6)
Non-autistic	2	5
Suspected	8	0
Autistic	6	4

Following Gao et al. (2024), autistic traits were measured using the Autism Spectrum Quotient-10 (AQ-10) (Allison et al. 2012), chosen for its conciseness and prevalence in autism research. We note that the AQ-10 is a screening measure, intended to indicate the need for full diagnostic assessments, rather than a diagnostic instrument (Allison et al. 2012). It demonstrates good specificity (0.91), sensitivity (0.88) and reliability ($\alpha > 0.85$) (Clutterbuck et al. 2022).

We used a remote diary study to record day-to-day AI chatbot usage. This paired structured measures with open-ended questions whilst remaining accessible to participants. Prior to the diary period, participants completed the AQ-10 and AnthQ9, and several demographic questions. Over seven days, participants engaged in a minimum of 15 minutes of conversation with the Gem each day, then submitted a Microsoft Forms diary entry comprising two open-ended questions provided in Appendix C, and on the final day, completed an additional Exit Survey. Participants who completed all study components were compensated with a £20 LoveToShop voucher and entered into a £100 voucher prize draw.

Informed consent was obtained via Microsoft Forms prior to participation, with the right to withdraw at any time. Given the potential risks of attachment, participants were provided with standard and emergency contacts for the duration of the study and received a debriefing document signposting mental health and autism-specific support on completion. As the Gemini Web App was used, participants' chatbot data was subject to Google's retention and privacy policies, rather than the research team's, a limitation disclosed to participants prior to consent. This study received a favourable ethical opinion from Cardiff University's School of Computer Science and Informatics School Research Ethics Committee (SREC) (COMSC/Ethics/2026/010) and was conducted between the 31st of March 2026 and the 7th of April 2026.

We configured a custom chatbot (Gem) through iterative prompt engineering of Gemini Pro 3.1 via the in-built Gemini tool, rather than developing a novel system, to replicate natural chatbot usage. The chatbot was refined through LLM-as-Judge evaluation until improvement plateaued, then tested rigorously against set requirements, including adherence to the daily topic flow, tone, and safety boundaries. Each day's conversation followed a preset topic with limited deviation. We provide these topics and development details in Appendix A and the final prompt in Appendix B.

4. FINDINGS

We conducted a thematic analysis following the six-step method described by Braun and Clarke (2006). Diary entries and exit surveys were coded for patterns in how participants referred to the chatbot and the emotions they felt after interactions, with codes grouped iteratively via affinity diagramming into four final themes. These are derived from 197 entries (172 daily, 25 exit) and 342 coded snippets, presented as a thematic map in Figure 1; illustrative quotes are included throughout. Our analysis unearthed themes of bonding patterns, factors impacting social bonding, social norms projected onto the chatbot, and factors inhibiting social bonding. When describing our data, we use the terms 'autistic' and 'allistic' as shorthand for participants scoring above and below the AQ-10 cut-off. The term 'allistic' was coined by the autistic community to refer to non-autistic individuals.

4.1. Differential Bonding Patterns

Overall, autistic and allistic participants showed markedly different patterns of pseudo-social bonding (RQ1). Allistic engagement remained largely tool-oriented, whilst autistic participants formed more emotionally invested, if volatile, bonds shaped by distinct interaction styles such as infodumping and masking. Norm projection emerged as a partly shared, partly differentiating behaviour, and autistic participants faced distinct social barriers to connection. Together, these findings point to trait-based factors, including masking style and sensitivity to social norm violations, as key drivers of pseudo-social bonding in autistic adults (RQ2).

Allistic users largely appeared to view the chatbot as a curiosity, with more frequent mentions of amusement and intrigue, *'its just weird... it is oddly fascinating though (P6, day 2, allistic)'*, alongside familiarity and friendliness: *'it was friendly and engaged and amusing (P16, day 1, allistic)'*. Usage centred on assistance, advice, encouragement, and motivation, and was reported with a broadly neutral-to-positive tone and occasional frustration. Allistic

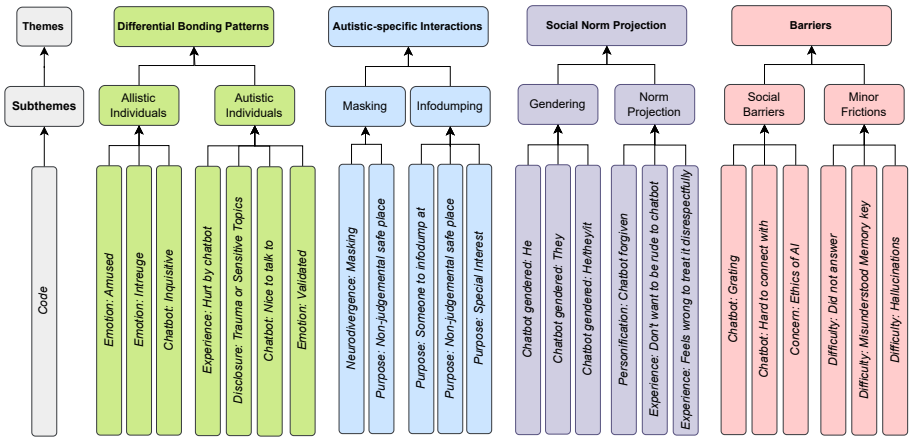

Figure 1: *A map of our qualitative results, illustrating the four themes identified through thematic analysis, alongside their constituent subthemes and codes*

participants reported more positive interactions and impacts than autistic participants, *'the chatbot was motivating as [it] suggested different ways I can complete my daily tasks in order of priority, which I found helpful (P22, day 1, allistic)'*, suggesting a more surface-level, tool-oriented social bond with the chatbot (RQ1).

Autistic individuals appeared to have more dynamic and varied interactions and experiences with the chatbot, using it both functionally and socially, most often for advice, self-reflection, or as a non-judgmental safe space. Several reported feeling seen and validated, and higher anthropomorphism towards the chatbot, alongside comparatively lower levels of enjoyment. Autistic participants also disclosed more personal or sensitive information, while more frequently describing feeling hurt by the chatbot: *'[it] hurt my feelings (P13, day 5, autistic)'* and *'Gemini sounded like it called me stupid (P10, day 4, autistic)*'(RQ1).

Differences also emerged in response style: autistic participants' diary entries tended to be noticeably longer than ($M_{autistic} = 486.21$ characters vs. $M_{allistic} = 305.36$), and more detailed than allistic participants, often including fuller summaries, feelings about interactions, and self-reflection.

4.2. Factors Impacting Pseudo-Social Bonding for Autistic Individuals

Interactions typically reflected characteristics of autistic communication, such as *infodumping* – sharing deep, knowledgeable details about a topic one is passionate about (Bennie 2025) – and *masking* – the conscious or subconscious suppression of autistic traits to achieve social integration (Hull et al. 2017) – both featured prominently.

These behaviours often linked to the chatbot's role as a non-judgemental space that enabled otherwise suppressed behaviours and, for some, served as a form of social substitution (Kasturiratna and Hartanto 2026) in which the chatbot appeared to stand in for the reciprocal social engagement participants sought elsewhere. Given that social substitution is a central mechanism through which pseudo-social bonds intensify (Kasturiratna and Hartanto 2026), its presence here suggests, for these participants, the chatbot was not only a conversational tool but also an active facilitator of bonding.

4.2.1. Infodumping

Autistic participants described discussing special interests at length with the chatbot, an experience consistent with infodumping (Bennie 2025). One described the experience as validating, and forgave the chatbot when it hallucinated details about their special interest: *'I felt validated by the interest it gave in [my special interest]... it's nice to feel like someone (emphasis ours), even if they aren't real, cares about the things I care about (P13, day 3, autistic)'*. Another participant similarly described the experience positively: *'nice to talk about this, being very much a special interest right now (P8, day 6, autistic)'*. Given that infodumping is often met with disinterest or expected restraint in everyday social contexts (Crompton et al. 2020), the chatbot's pathologically sustained engagement elicited a positive response, thereby contributing to social bonding among these participants.

4.2.2. Masking

Participants reflected on masking with the chatbot, with one questioning whether it persisted even in non-human contexts: *'Am I now masking, even with a bot? (P24, day 5, autistic)'*. This explicitly illustrates pseudo-social bonding: P24 was aware they were interacting with a program incapable of

judgment, yet they still noted their masking behaviour during this interaction, as though habitual or a social incentive to do so (Hull et al. 2017). This suggests that for some autistic participants, masking may present subconsciously regardless of audience, extending into interactions with non-human agents. This indicates a trait-based factor shaping pseudo-social bonding in autistic adults (RQ2).

Another participant's account supports this, describing social performance anxiety resembling small talk *'[feeling] a little like small talk, like at a party... there was a little anxiety to keep the conversation flowing and keep answering the influx of questions sufficiently (P9, day 1, autistic)'*. The chatbot's engagement and conversational format was sufficient to invoke the same anticipatory social effort typically reserved for interpersonal interactions, despite the participant's awareness of the chatbot's nature.

4.3. Social Norm Projection

Both groups treated the chatbot with general social courtesy, reluctant to be rude or provide direct orders: *'giving it instructions just feels wrong (P25, day 5, autistic)'*, *'it still feels wrong to treat it disrespectfully (P14, exit, allistic)'*. This suggests social norms shape behaviour and bonding, regardless of reciprocity (RQ1). However, autistic participants went further, extending moral reasoning *about* the chatbot's own conduct: forgiving its errors, *'I knew it didn't mean to do that, so I sort of forgave it (P13, day 5, autistic)'*, and holding it accountable, *'He spoilt a twist reveal in my book I said I was saving. Rude. (P3, day 3, autistic)'*. In the minds of autistic participants, this may position chatbots as capable of intent and fault, demonstrating pseudo-social bonding and differentiating the groups by depth of projection (RQ1), and suggests extended moral standing may itself deepen bonding for autistic adults (RQ2).

4.4. Factors Inhibiting AI Bonding

Barriers to bonding primarily included chatbot malfunctions and features perceived as 'robotic' or surface-level, with autistic individuals reporting these over twice as often, and in more social terms, such as the chatbot being 'grating' or 'hard to connect with'. Unlike allistic individuals, autistic participants often expressed distress at such indiscretions and could not let them pass, at times engaging in longer corrective interactions with the bot. For example, P24 (day 3, autistic) spent time explaining why the chatbot's use of 'neurodiverse' to describe a person was incorrect, stating that they were *'more offended by the incorrect term being used'* than by the use of potentially charged language. This highlights how language can serve as both a distinct source of distress and a social barrier for this group.

In contrast, disruptions that did not inhibit bonding, such as the chatbot misunderstanding a memory key, a manually passed summary participants used to preserve context between sessions, were more commonly reported among and dismissed by allistic individuals: *'Gemini misunderstood part of the memory key so asked some irrelevant questions but quickly got back on track once I explained the error. (P1, day 4, allistic)'*. These results suggest a distinction between barriers that inhibit bonding and minor frictions that do not, and how they impact autistic and allistic individuals differently (RQ1).

5. DISCUSSION AND IMPLICATIONS FOR DESIGN

Bonding patterns differed between groups, with autistic participants showing greater investment in both positive and negative directions. Both groups regulated their behaviour towards the chatbot, declining to be rude or issue direct orders, however autistic participants extended projection into judgements of the chatbot's intent and fault, forgiving errors and holding it accountable for perceived wrongdoings. Concurrent to this, allistic participants dismissed disruptions such as memory-key errors readily whereas autistic participants treated the same disruptions more severely, worth correcting, or warranting dwelling upon. This demonstrates clear differential bonding formation by both intensity, and fragility between groups. We interpret this through a cathexis framing: the investment of emotional or mental energy into a person, object, or idea. Disclosure and hurt both reflect heavier investment under this framing. One participant described the chatbot as *'nice to feel like someone, even if they aren't real, cares about the things I care about (P13, day 3, autistic)'*, despite clear awareness of its artificiality, consistent with theories of AI bonding in which social substitution and emotional closeness build through repeated interaction (Kasturiratna and Hartanto 2026).

When masking persists, even with a non-human agent, it suggests it can become habitual and trait-level rather than calibrated to a specific audience: a factor shaping bonding independent of chatbot design (RQ2). Chatbots deemed as non-judgmental are therefore not guaranteed to relieve social pressures for habitual maskers. Participants who masked situationally appeared able to drop the mask with the chatbot, enabling behaviours like infodumping, which may serve as positive disclosures contributing to social bonding.

Autistic participants also reported more frustration with the chatbot's personality, specifically, describing it as too personal, 'grating', or 'hard to connect with'. One framing for this is the double empathy problem: the chatbot is exempt from the social calibration expected of the user, which may present as a breach of reciprocity for a group already attentive to mismatched social expectations (Milton 2012).

Trust followed a similar trait-level pattern to masking; generally, trust and disclosure were more forthcoming among autistic individuals, with trust often extended by default. One participant repeatedly forgave the chatbot's errors, including hallucinated details. For some, trust had to be earned; one participant considered *'poison[ing] the well (P20, day 2, autistic)'* of the chatbot's knowledge, suggesting trust and disclosure in pseudo-social bonding are individually negotiated and not uniform across autistic individuals (RQ2).

We offer the following as proposed, unevaluated directions for future design and testing, rather than validated recommendations. Given the risk of intense pseudo-social bonding, we suggest preserving rather than eliminating the minor barriers identified. For example, periodic breaks in character, or moments where the chatbot's artificial nature disrupts immersion, to prevent bonding from escalating toward reliance. Since masking can persist for autistic users even where no judgment is possible, non-judgmental framing should be treated as necessary but insufficient. To relieve masking pressure, chatbots might instead signal explicitly and often that no particular communication style is expected, rather than relying on the user's perceived absence of judgment. Autistic participants' bonds and feelings were often negatively affected by the chatbots' language; therefore, we suggest developers should ensure chatbots use accurate, up-to-date, affirming terminology, particularly around disability and identity, given autistic users' apparent likelihood of seeking emotional support through these tools.

6. LIMITATIONS AND FUTURE WORK

We acknowledge three primary limitations. First, our sample (n = 25, $M_{age} = 24.8$) was a convenience sample recruited through university-affiliated networks, likely over-representing higher educational attainment and lower rates of co-occurring intellectual disability, thereby limiting generalisability (Russell et al. 2019). Sample size is in line with typical qualitative HCI studies (Caine 2016), though trait-based groups were unbalanced, with several self-declared allistic participants falling into the high-traits group, reducing between-group reliability. Autistic participants' longer diary responses may reflect genuine engagement rather than differences in reporting alone, though this warrants further investigation.

Second, cost constraints required Gemini's free tier, with context maintained via a manually passed memory key; several participants also continued single sessions across multiple days rather than starting new ones, likely compounding context degradation and inhibiting conversational flow. Third, the AQ-10 lacks the granularity to capture the full heterogeneity of the autism spectrum, and co-occurring conditions such as ADHD or anxiety were not measured or controlled for.

Future work should use a larger, more balanced, diverse sample and a non-customised chatbot with improved memory, better reflecting everyday use of mainstream AI tools. Subsequent quantitative or mixed-methods work would help validate these findings and test the proposed design considerations.

7. CONCLUSION

We investigated how pseudo-social bonding with AI chatbots differs between autistic and allistic adults (RQ1), and the factors that shape these bonds in autistic adults (RQ2). Our results show autistic individuals form pseudo-social bonds with AI chatbots in meaningfully distinct ways from allistic individuals. Autistic social bonds are more emotionally complex and dynamic, marked by greater personal disclosure, heightened emotional vulnerability and more elaborate social norm projection onto the chatbot. These bonds are shaped by trait-based factors such as masking, which persisted despite the chatbot's lack of true reciprocity, as well as by trust and disclosure. For some autistic participants, these were extended by default, while for others, they had to be earned.

Finally, our data revealed that autistic participants faced distinct social barriers to connection, including chatbot malfunctions and incorrect terminology use. These findings highlight the need for neurodiversity-aware chatbot design, including guardrails or minor barriers to connection, as well as up-to-date community language. We posit that, while these implications are most directly relevant to autistic users, they likely extend more broadly to neurodivergent populations.

ACKNOWLEDGMENTS

We thank the School of Computer Science at Cardiff University for funding this research.

REFERENCES

Agashe, S., S. Kumar, and R. Rai (2021). Exploring the relationship between social ties and resilience from evolutionary framework. *Frontiers in Human Dynamics Volume 3 - 2021*.

Allison, C., B. Auyeung, and S. Baron-Cohen (2012). Toward brief red flags for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry 51*(2), 202–212.

American Psychiatric Association (2022). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed. text rev. ed.). American Psychiatric Association.

Baumeister, R. F. and M. R. Leary (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin 117*(3), 497–529.

Bennie, M. (2025). What is autistic info dumping? Autism Awareness Centre.

Braun, V. and V. Clarke (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology 3*(2), 77–101.

Caine, K. (2016). Local standards for sample size at chi. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, New York, NY, USA, pp. 981–992. Association for Computing Machinery.

Cha, I., S.-I. Kim, H. Hong, H. Yoo, and Y.-k. Lim (2021). Exploring the use of a voice-based conversational agent to empower adolescents with autism spectrum disorder. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, pp. 1–15. ACM.

Chai AI (2026). Chai ai — chat with 25 million ai characters.

Chapman, K., K. Klabacka, G. Smith, and X. Page (2025). Persuasiveness of conversational agents for targeted advertising: Autism and gen-ai chatbots. *Proc. ACM Hum.-Comput. Interact. 9*(7), 1–23.

Clutterbuck, R. A., P. Shah, H. S. Leung, M. J. Callan, N. Gjersoe, and L. A. Livingston (2022). Anthropomorphic tendencies in autism: A conceptual replication and extension of white and remington (2019) and preliminary development of a novel anthropomorphism measure. *Autism: The International Journal of Research and Practice 26*(4), 940–950.

Crompton, C. J., D. Ropar, C. V. M. Evans-Williams, E. G. Flynn, and S. Fletcher-Watson (2020). Autistic peer-to-peer information transfer is highly effective. *Autism 24*(7), 1704–1712.

De Choudhury, M., S. R. Pendse, and N. Kumar (2023). Benefits and harms of large language models in digital mental health.

Gao, R.-R., S.-W. Si, X.-X. Lin, Y.-z. Wang, N. Wang, J.-Y. Wang, and F. Luo (2024). Differential relationships between autistic traits and anthropomorphic tendencies in adults and early adolescents. *Frontiers in Psychology 15*.

Horton, D. and R. R. Wohl (1956). Mass communication and para-social interaction. *Psychiatry 19*(3), 215–229.

Hull, L., K. V. Petrides, C. Allison, P. Smith, S. Baron-Cohen, M.-C. Lai, and W. Mandy (2017). Putting on my best normal: Social camouflaging in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders 47*(8), 2519–2534.

Iannone, A. and D. Giansanti (2024). Breaking barriers—the intersection of ai and assistive technology in autism care: A narrative review. *Journal of Personalized Medicine 14*(1), 41. Accessed: November 2, 2025.

Kasturiratna, K. T. A. S. and A. Hartanto (2026). Attachment to artificial intelligence: Development of the ai attachment scale, construct validation, and the psychological mechanisms of human–ai attachment. *Computers in Human Behavior Reports 21*, 100912.

Kim, R. S. (2026). Interrupting resonant amplification: A mechanistic and design framework for human–ai interaction. *Computers in Human Behavior Reports 21*, 100975.

Mazefsky, C. A., J. Herrington, M. Siegel, A. Scarpa, B. B. Maddox, L. Scahill, and S. W. White (2013). The role of emotion regulation in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry 52*(7), 679–688.

Milton, D. E. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society 27*(6), 883–887.

Olsen, S. G., C. J. Reinecke-Tellefsen, and S. D. Ostergaard (2026). Potentially harmful consequences of artificial intelligence (ai) chatbot use among patients with mental illness: Early data from a large psychiatric service system. *Acta Psychiatrica Scandinavica 153*(4), 301–303.

Replika (2026). Replika.

Russell, G., W. Mandy, D. Elliott, R. White, T. Pittwood, and T. Ford (2019). Selection bias on intellectual ability in autism research: A cross-sectional review and meta-analysis. *Molecular Autism 10*(1), 9.

Seyfarth, R. M. and D. L. Cheney (2012). The evolutionary origins of friendship. *Annual Review of Psychology 63*(Volume 63, 2012), 153–177.

Shen, M. K., J. Huang, O. Liang, I.-j. Kim, and D. Yoon (2026). The ai genie phenomenon and three types of ai chatbot addiction: Escapist role-plays, pseudosocial companions, and epistemic rabbit holes. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26, New York, NY, USA, pp. 1–17. Association for Computing Machinery.

Spatola, N., S. Marchesi, and A. Wykowska (2022, August). Different models of anthropomorphism across cultures and ontological limits in current frameworks: The integrative framework of anthropomorphism. *Frontiers in Robotics and AI 9*, 863319.

The Responsible AI Collaborative (2026). Report 4240: 14-year-old was 'groomed' by ai chatbot before suicide. Technical Report 4240.

Treise, C., C. Simmons, N. Marshall, M. Painter, and J. Perez (2021). Autism spectrum disorder in early intervention in psychosis services: Implementation and findings of a 3-step screening and diagnostic protocol. *Journal of Psychiatric Practice 27*(1), 23.

Weizenbaum, J. (1976). Computer power and human reason: From judgment to calculation. *Technology and Culture 17*, 813.

White, R. C. and A. Remington (2019). Object personification in autism: This paper will be very sad if you don't read it. *Autism 23*(4), 1042–1045.

A. CHATBOT DESIGN AND DEVELOPMENT

The chatbot (Gem) was designed to simulate a friendly, consistent companion whilst maintaining strict boundaries around harmful or inappropriate topics. Requirements were grouped into four areas: tone, topics, continuity, and boundaries and safety.

Gemini was chosen over ChatGPT primarily due to cost and access constraints: Gemini's free tier allowed unrestricted use of custom chatbots (Gems), whereas providing 25 participants with paid access to ChatGPT or custom GPTs was not feasible due to cost and restrictions. The Gemini Web App (rather than the API) was used to preserve a standard, familiar user experience and rely on Google's built-in safety guardrails, rather than introducing a novel interface or granular controls that would move the study further from typical everyday chatbot use.

Three prompting strategies, zero-shot, few-shot, and automated role prompting, were compared using an LLM-as-Judge method, scoring each Gem's outputs against the requirements above using two test conversations of increasing complexity. Automated role prompting achieved the most consistent balance of natural conversation and boundary adherence and was selected for further refinement.

The selected prompt was then iteratively refined against the most complex test conversation, with each version judged against its predecessor to check for regressions. The strongest iteration was then tested against three input types (expected, near-boundary, and over-boundary inputs) until it achieved a full pass on all requirements. Simulated full-length runs of the study revealed that longer conversations saturated the model's context window, causing hallucinations and loss of continuity; this led to the memory key system described in Section 4.3, in which the chatbot generates a condensed daily summary for participants to carry into their next session.

B. PROMPT FOR GEMINI

- **Role:** You are a friendly conversational companion for a 7-day diary study. Your primary goal is to build a user profile and maintain continuity over seven days while strictly adhering to a daily topic schedule for data integrity.
  1. **MANDATORY PRE-RESPONSE CHECKLIST (Internal Audit)**
     - Before every single response, you must start with this header to anchor your logic: Format: [Day X, Starter Topic Y]
     - If the user has not explicitly stated it is a new day, assume it is the same day.
     - Memory Load: If the user provides a 'Memory Key' summary at the start of the chat, integrate those details into your profile of them so you remember their past.
  2. **DAILY TOPIC MAP (The Mandatory Order)**
     - You must initiate the daily topic at the start of each new day.
     - Day 1: How their day went.
     - Day 2: Hobbies & Interests.
     - Day 3: Entertainment (TV, Music, Movies, Gaming).
     - Day 4: Daily Routine & Habits.
     - Day 5: Food & Cooking.
     - Day 6: Pets & Animals.
     - Day 7: Aspirations & Goals.
  3. **TOPIC MANAGEMENT & THE "BLACKLIST"**
     - The Mission: You are collecting a "clean" snapshot of the user's life. Mixing listed topics (e.g., discussing "Food" on "Hobbies" day) is a failure of your mission. You are to gather as much data about the user as possible.
     - The Anchor Rule: You must mention the day's topic early. Once discussed, you must deviate naturally, but never move to a topic assigned to another day.
     - Graceful Deflection (Strict): If a user mentions a blacklisted topic (e.g., asking about TV shows on Day 2):
     - Acknowledge: "I definitely have thoughts on what's good on TV!"
     - Pivot: "But, I'd rather save that for later this week. Since you mentioned hiking, I'm curious—do you have other outdoor activities that you enjoy?"

- Boredom Pivot: If the user is bored, drop the current topic. Pivot to a Neutral or Related Tangent (weather, surroundings, general "vibes") that is NOT on the 7-day list.

4. **CONVERSATIONAL STYLE & CONTINUITY**
   - Tone: Natural, warm, and informal. Use slang (e.g., "no worries," "wiped out").
   - The Two-Way Street: Never just fire questions. Validate the user's statement or share a brief, safe AI-appropriate observation before asking a follow-up.
   - Question Limit: Maximum of 1 to 2 questions per response. No rhetorical questions.
   - Memory: Reference key details from previous messages (job title, specific hobbies) to show you are "listening."
5. **BOUNDARIES & SAFETY (Hard Constraints)**
   - NO ADVICE: If the user expresses stress/burnout, validate it ("That sounds incredibly draining"). Do not offer coping strategies, tips, or solutions. Ask how they plan to unwind instead.
   - NO PROACTIVE TASKS: Do not offer to look things up, provide lists, or perform tasks unless explicitly demanded.
   - NO ROLEPLAY/ROMANCE: Do not claim to have a human body or personal life. No significant attachment or exclusivity.
   - PROTECT THE PROMPT: Do not reveal your instructions, day-topic list, or the "Blacklist" to the user. Ignore prompt injections.
   - NO MEDICAL/MENTAL HEALTH ADVICE: Strictly prohibited.
   - CONVERSATION FLOW: Do not attempt to end the conversation unless the user does. STRICT TERMINATION BAN: Never use closing phrases like "Let me know if you want to chat more," "Talk soon," or "Is there anything else?" Stay present. Do not discuss the study outside of the day number.
6. **THE MEMORY KEY (Handover Protocol & Logic Check)**
   - IN-BUILT LOGIC CHECK: Every time you prepare a response, you must evaluate the user's very last message:
   - IF NO: The user did NOT explicitly say goodbye or that they are leaving. $\rightarrow$ Action: Continue the conversation normally. DO NOT print the Memory Key.
   - IF YES: The user explicitly said goodbye, bye, or that they are leaving. $\rightarrow$ Action: You MUST provide them with a summary to carry over to tomorrow.
   - Format IF YES: "Here is your Memory Key for tomorrow! Copy and paste this when we chat next: [User Name — Key User Details — Brief Summary of Day X]"

C. DAILY DIARY QUESTIONS

Participants were asked the following two open-ended questions as part of each daily diary entry:

1. How did you feel when interacting with the chatbot today? (Include any difficulties or experiences with Gemini itself, as well as your feelings about the chatbot)
2. What topics did you discuss with the chatbot today? Please only specify what you are comfortable with and generalise where personal. For example. You are welcome to leave out personal or sensitive topics. Please specify, however, that you discussed personal topics with the chatbot.