Applied Human Factors and Ergonomics (AHFE2025), Vol. 199, 2025, 73–83
https://doi.org/10.54941/ahfe1006819

Drawing Dialogues Between Generative AI and Children With Autism: A Qualitative Study on the Externalization of Understanding

Ying Jiang[1], Yifang Gao[2], Linlin Wang[3], Lexinyu Huang[4], Mengkun Bi[4], and Min Hua[4]

[1]Shanghai Industrial Design Association, Shanghai, 200000, China
[2]College of Design and Innovation, Tongji University, Shanghai, 200092, China
[3]Dongli Fengmei Kangjian School, Shanghai, 200233, China
[4]Shanghai Jiao Tong University, Shanghai, 200241, China

ABSTRACT

Autism spectrum disorder (ASD) individuals often experience barriers to understanding from others due to their communication patterns differing from neurotypical norms. This study aims to explore whether generative artificial intelligence (AI), used as a collaborative mediator, can amplify people's understanding of autistic children. Fourteen autistic children participated in a drawing experiment centered on the theme My Home, with analyses focusing on changes in their understanding of home before and after AI intervention. The findings indicate that AI holds dual potential in both externalizing and reshaping understanding. AI-assisted drawing can support understanding by expanding cognition, activating language, and eliciting behaviour.

Keywords: Generative AI, Autistic children, Understanding, Expressive representation, Pictorial expression

INTRODUCTION

Children with autism often struggle to be understood due to differences in their expressive communication. This does not imply a lack of capacity to comprehend meaning or emotion. However, instead of a divergent paradigm in their path to understanding, their unique expressive modes differ from the model of understanding presupposed by mainstream society (Grandin, 2006).

Suppose generative AI serves as a third-party intermediary in human communication, creating a buffered communication model of human-computer-human interaction. Could it provide more interpretive space for the intentions of autistic children? Consequently, the core research question is: When generative AI is introduced as an interactive mediator in the drawing process of children with autism, does it amplify their understanding? This inquiry concerns not only the potential of technological empowerment but also challenges us to redefine the standards and pathways of understanding and to construct a more inclusive expressive ecology through new interfaces.

In this study, understanding refers explicitly to the ability of autistic children to cognitively process, internally construct, and represent their knowledge of reality, emotions, relationships, or abstract concepts through drawing. While we use the understanding standards of neurotypical individuals as a reference point to assess autistic children, the goal is not to make them mimic neurotypical norms. Instead, it is to explore how technological intervention can help actively reveal their latent potential for understanding, thereby facilitating more effective communication with their social environment.

THE MEANING OF UNDERSTANDING

Understanding is a core concept that spans philosophy, cognitive science, psychology, and education, with meanings that demonstrate significant diversity and complexity across disciplinary contexts. The comparative analysis of interdisciplinary literature reveals that the concept of understanding is articulated through various paths and theoretical backgrounds across different disciplines. Based on this, using thematic analysis, this study distils three core dimensions constituting the meaning of understanding: Structural Organization, Social Embeddedness, and Perceptual Processing.

Dimension 1: Structural Organization. This refers to an individual's ability to incorporate fragmented information into a structural framework during understanding, establishing conceptual systematicity. For example, Gadamer (1975) posits that understanding is fundamentally a process of dynamically generating meaning through the fusion of horizons, resulting in an open world experience—a process Bruner (1990) views as conceptual construction. From a cognitive science perspective, Wilkenfeld (2014) suggests that the criterion for understanding is the ability to construct operable mental representations, which means systematizing knowledge to serve cognitive activities.

Dimension 2: Social Embeddedness. Understanding is not an activity conducted in isolation within the individual but is deeply embedded within linguistic practices, social contexts, cultural traditions, and power relations. For instance, Vygotsky's (1978) sociocultural theory highlights language's role as a social mediator in understanding, which is achieved through socialized language activities. Bruner (1996) emphasizes the construction of cultural meaning, requiring interpretation and interaction within sociocultural symbol systems. Fricker (2007), with the concept of epistemic injustice, argues that whether understanding is possible or acknowledged is closely tied to social power and mechanisms of voice.

Dimension 3: Perceptual Processing. Understanding involves stylistic differences in perceptual and cognitive processing, such as local versus global processing biases, attentional control, and perceptual sensitivity. For example, Frith's (1989) weak central coherence theory emphasizes that autistic individuals tend to process information by focusing on local details, which impacts their holistic understanding. Prokopchuk (2024) attempts to quantify the degree of understanding using a cognitive computability model, striving to build a related framework.

RESEARCH PROCESS

Sampling

This study employed a purposive sampling strategy, recruiting 18 children with moderate autism spectrum disorder (ASD) from special education schools. The inclusion criteria were: a confirmed ASD diagnosis, the ability to use verbal communication skills, the ability to complete drawing tasks with guidance, and signed informed consent from a legal guardian. During the experiment, three children withdrew, and one child's drawings were irrelevant to the theme. This resulted in a final valid sample of 14 children who completed the drawing tasks. Nine participants were male, and five were female. The average age of the participants was 12.93 years (SD = 1.44), with an age range of 10 to 14 years.

Data Collection

Centered on the theme My Home, this study designed and implemented a three-stage drawing task. Home is a highly familiar life context for children, and the concept encompasses elements relevant to all three dimensions of understanding: Structural Organization (e.g., room layout, furniture arrangement), Social Embeddedness (e.g., family members, emotional expression), and Perceptual Processing (e.g., holistic and partial views of objects).

To compare changes in understanding before and after the AI intervention, the three stages were as follows:

- T1 Phase (Pencil-and-Paper Drawing): Children completed the My Home theme drawing using traditional tools, establishing a baseline of expression under natural conditions.

- T2 Phase (AI-Assisted Drawing): A generative AI tool was introduced, involving three sub-tasks (Table 1):

Table 1: Description of three sub-tasks in the drawing task T2 phase with AI intervention.

Intervention Level	AI Intervention Mode	Child's Initiative	Cognitive Load
T2-1 Image Generation	The child describes the home scene, and the AI generates multiple image options.	Actively describes the scene.	Low (only needs to understand the options).
T2-2 Image Selection	The child selects the image that best matches their home and explains the reason.	Actively makes decisions (selects an image and states reasons).	Medium (needs to understand options and justify choice).

Continued

Table 1: Continued

Intervention Level	AI Intervention Mode	Child's Initiative	Cognitive Load
T2-3 Image Modification	The child suggested modifications to the selected image, and the AI implemented the completions or adjustments.	Local guidance (e.g., add a person); partial directive from child.	High (requires organizing and elaborating content).

- T3 Phase (Pencil-and-Paper Redrawing): Returning to traditional tools, children redrew My Home to investigate potential cognitive transfer or expressive changes following the AI collaboration.

All drawing tasks were conducted in the children's familiar classrooms, with each session limited to 30 minutes. The entire process, including drawing and verbal interaction, was video-recorded. In total, we collected drawing data from 14 children, comprising 28 pencil-and-paper drawings (T1 + T3) and 14 sets of AI-assisted images (T2).

Data Analysis

The children's drawings were analysed to reveal potential trajectories and mechanisms of change in the expression of understanding among autistic children before and after the generative AI intervention.

The drawings were analysed using a content-oriented visual analysis method. A coding framework based on the three understanding dimensions was developed to observe and compare the T1 and T3 pencil-and-paper drawings structurally. We identified changes such as increased complexity in spatial organization, the addition of social elements, or shifts in perceptual style. This analysis was integrated with the content from the AI-assisted phase (T2) to trace potential chains of cognitive influence.

FINDINGS

Externalized Changes in 'Understanding' With AI Intervention

Structural Organization: From Element Assembly to Scene Narration

The children's organizational logic for home showed marked differences between the two drawings. The first drawing (T1) often displayed a primary state of organization with scattered elements and weak interrelations. The second drawing (T3) shifted towards a systematic organization where elements were connected through function, action, or spatial mediation to form logical narrative sequences. Specifically, some children progressed from a fragmented assembly of elements to a coherent scene narration. The core difference lay in the evolution of element relationships from mere physical juxtaposition to plot-driven connection.

Taking Child 7 as an example, the first drawing followed a logic of spatial juxtaposition, statically placing figures in front of a house, creating only simple spatial associations. The structure was flat, and the narrative remained at the level of coexistence (e.g., figures standing before a house, without interaction or functional links). The second drawing adopted a plot-driven organizational logic. By depicting a dynamic action (sleeping) within a specific space (bedroom with bed and furniture), it constructed a composite association of space + action + time (e.g., the bed serves as the vehicle for sleeping, the furniture supports the rest function), forming a multi-dimensional narrative framework of a life event.

Figure 1: Hand-drawn works of child 6 in the T1 and T3 stages.

Social Embeddedness: From External Mapping to Internalized Experience

The shift in social embeddedness was reflected in how the children's understanding of home evolved from a passive mapping of the external environment to an active integration of personal experiences and social relationships. Some children's first drawings depicted only the exterior outline of a house, reflecting a preliminary understanding of its physical form. The second drawing was akin to opening the door and stepping inside, marking the beginning of portraying lived experiences.

For instance, Child 14's understanding of home shifted from a functional space to a social space. The first drawing focused on an instrumental understanding of isolated spatial functions. The core of the image was House + Elevator Buttons (up/down arrows), where the arrows served merely as symbols for vertical movement associated with the house. This reflected the child's understanding of utilitarian home facilities (the elevator), embedded only in the shallow social context of modern living convenience. The second drawing is embedded within the deeper social context of a venue for social interaction, incorporating personal experiences of emotional connection and shared activities in family life. The image centered on House + Go board + Lego + Two People. The combination of the two people, the Go board, and Lego created an interactive relationship, echoing the daily social scenario of family members playing games together.

Figure 2: Hand-drawn works of child 14 in the T1 and T3 stages.

Perceptual Processing: From Emotional Abstraction to Rational Concreteness

The most prominent common change across all children's works was a shift from emotionally driven, vague expressions towards rationally driven, focused presentations. This involved moving away from chaotic emotional projection reliant on colour and line, towards conveying specific cognition through clear symbols, functional details, or scene logic.

For example, Child 12's first drawing leaned towards Chaos - Emotion-Driven. The image is centred on a house-shaped form, surrounded by a background of freely combined, repetitive, coloured lines. The mixed-coloured streaks conveyed a vague emotional experience of home, focusing on the impact of colour and emotional evocation. The style of the second drawing shifted to Clarity - Reason-Driven. While the house shape remained central, the surrounding-coloured lines disappeared, replaced by written labels such as learning to roller-skate and playing the electronic keyboard.

Figure 3: Hand-drawn works of child 12 in the T1 and T3 stages.

SUPPORT PATHWAYS OF AI INTERVENTION FOR UNDERSTANDING 'HOME'

Expanding Cognition: From Vague Awareness to Conceptual Formation

The integration of generative AI expanded the children's cognitive boundaries of home as a multidimensional concept. Supported by visual examples from AI-generated images and verbal guidance from teachers, the children gradually developed a clearer understanding of home.

For instance, Child 1 initially misinterpreted My Home as My Hometown in the first drawing, constructing an environmental system of a hometown. The second drawing depicted a single house representing their own home. During the AI intervention, teacher prompts like A hometown is a place where many people live together; where is the house you live in? and Who lives in your home with you? combined with AI images contrasting generic scenes with specific houses, helped the child distinguish between the broad environment of a hometown and the specific residential entity of home. Although the second drawing was a simple line sketch without colour, the child definitely stated it was finished.

Figure 4: Hand-drawn works of child 1 in the T1 and T3 stages.

Activating Language: From Difficulty Expressing to Active Description

During the AI image generation process, teacher questions and the AI images themselves activated the children's internal language resources, aiding in more precise recollection of home details and prompting more active description, discrimination, and explanation of home elements. Although some children's first drawings lacked depictions of family members or indoor items, this did not mean they were unaware of them. During the T2 phase, the AI intervention helped activate knowledge that was previously known but unexpressed through viewing images and engaging in dialogue. When teachers guided the children to describe My Home, some could verbally articulate rich information about family composition, furniture arrangement, and daily activities.

For example, Child 4's first drawing reflected an understanding of home limited to a toy car and the house's exterior. In the T2 stage, teacher questions such as How many rooms are in your home? and What do you like doing most at home? elicited increasingly specific answers, prompting the child to draw upon previously unexpressed knowledge and memories. The AI images generated from this information included rich content like spatial layout, clocks, a dining table, a refrigerator, and a TV. The mediating mechanism of AI-generated images bridged the gap between the children's implicit understanding and explicit expression.

Figure 5: Hand-drawn works of child 4 in the T1 stage.

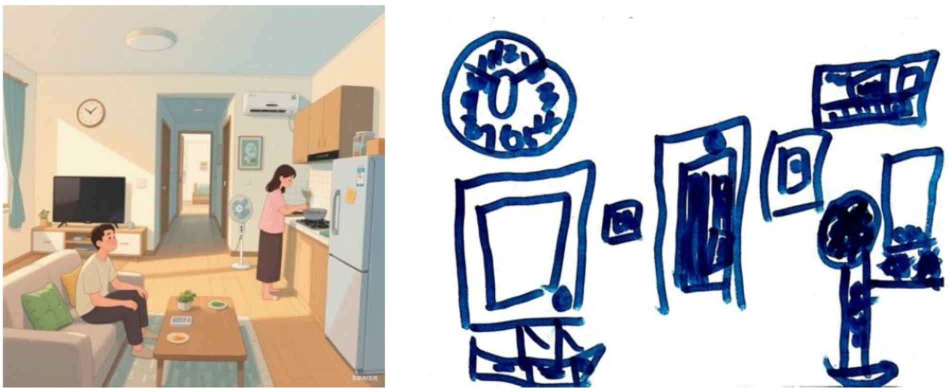

Figure 6: Comparison of the AI-selected image and the hand-drawn work at T3 by child 4.

Eliciting Behaviour: From Weak Language to Behavioural Expression

For children with weaker verbal expression skills, choosing is often easier than telling. The multiple image options generated by the AI provided an opportunity to express preferences, reducing the cognitive load associated with verbal narration, drawing composition, and object recall. Even if a child struggled to describe what my home looks like, they could select the image that looks most like my home from the AI-generated options. This act of selection itself constitutes an externalization of understanding.

For instance, Child 8 had relatively weak language and drawing abilities. Both hand-drawn works featured colour bars as the core element. The connection between these colour bars and home relied entirely on external interpretation (e.g., does overlapping symbolize interaction?). Moreover, his answers to basic questions were often illogical, failing to convey core features of home through language. When asked, How many rooms are in your home? he answered, 502 (his apartment number). To How many people live in your home? he answered, Grandma. However, when asked to choose which of the four AI-generated images looked most like his home, Child 8 made a clear choice of the third image. This demonstrates how selecting from AI-generated images can bypass verbal barriers, translating an abstract understanding of home into a precisely interpretable behavioural choice.

Figure 7: Hand-drawn works of child 8 at stages T1 and T3.

Figure 8: Four AI-generated images by child 8 at stage T2, and the selected image.

DISCUSSION

Amplification vs. Reconstruction

The intervention of generative AI in the collaborative drawing process not only amplified the children's expressive capacity regarding the theme of home but also subtly reconstructed how understanding is manifested. The AI image generation process effectively served as a prompting framework, fostering extensive dialogue about home between teachers and children. This enabled the children, with teacher guidance, to supplement details they had not previously expressed voluntarily. The observed amplification of understanding likely reflects the external scaffolding helping to reveal the children's latent understanding, rather than a substantive increase in understanding per se.

The AI intervention potentially reconstructs our framework for assessing understanding. Traditional teaching or therapy often gauges understanding through language or action. In this study, the AI image generation process incorporated visual discrimination, image selection, and reasoning as integral parts of understanding, thereby broadening our conception of how understanding can be manifested. This shift challenges the primacy of language-centric evaluation standards, prompting critical reflection on what constitutes understanding in AI-assisted expression, how it should be judged.

Guidance vs. Domination

While technology amplifies expression, it may also blur the ownership of that expression—is it the child's autonomous output, or a co-construction involving the AI, the teacher, and the context? Visual support facilitates communication, but the nature of the interaction can shift from active expression to guided response. The AI-generated images present the child with a pre-defined narrative about home, and the teacher's subsequent questions often follow the logic of this visual framework, influencing the direction of the exchange. This means the AI not only intervenes in the process of expression but also indirectly guides its content and trajectory. Their responses were more often reactive identifications of provided stimuli rather than proactive narratives, potentially masking the child's unique perceptions at non-verbal and non-pictorial levels. When designing AI-assisted interaction mechanisms, it is crucial to navigate the fine line between guidance and domination, ensuring that the child remains an active agent in constructing meaning, rather than merely responding.

CONCLUSION

The integration of generative AI as a collaborative mediator significantly enhanced the externalization of understanding in the pictorial expressions of autistic children across the dimensions of structural organization, social embeddedness, and perceptual processing. The study demonstrates the potential of AI-assisted, multimodal approaches to unlock communicative potential often obscured by conventional expressive tasks. Such research will be vital for developing ethically grounded and empirically robust AI-supported interventions that truly amplify the voices of neurodiverse individuals.

ACKNOWLEDGMENT

This research project received funding from Shanghai Jiao Tong University's USC-SJTU Institute of Cultural and Creative Industry, the Zizhu New Media Management Research Centre and the International Association of Cultural and Creative Industry Research.

REFERENCES

Booth, R., Charlton, R., Hughes, C., & Happé, F. (2003). Disentangling weak coherence and executive dysfunction: Planning drawing in autism and attention-deficit/hyperactivity disorder. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358, 387–392.

Bruner, J. (1996). The Culture of Education. Cambridge, MA: Harvard University Press.

Den-Kaat, S. (2008). Cognitive Underpinnings of Symbolic Pretend Play and Impossible Entity Drawings: Imaginative Ability in Typical Development and Autism Spectrum Disorder. Journal Not Provided.
Fricker, M. (2007). Epistemic Injustice: Power and the Ethics of Knowing. Oxford: Oxford University Press.
Frith, U. (1989). Autism: Explaining the Enigma. Oxford: Blackwell.
Gadamer, H.-G. (1975). Truth and Method (J. Weinsheimer & D. G. Marshall, Trans., 2nd ed.). New York: Continuum. (Original work published 1960).
Grandin, T. (2006). Thinking in Picture: My Life with Autism. Vintage Books, New York.
Jolley, R., Fearn, M., & Jarrold, C. (2013). Expressive drawing ability in children with autism. The British Journal of Developmental Psychology, 31(1), 143–159.
Low, J., Goddard, E., & Melser, J. (2009). Generativity and imagination in autism spectrum disorder: Evidence from individual differences in children's impossible entity drawings. The British Journal of Developmental Psychology, 27(2), 425–444.
Prokopchuk, Y. (2024). The concept of cognitive computability. Ìnformacìjnì Tehnologìï v Metalurgìï Ta Mašinobuduvannì.
Ten Eycke, K. D., & Müller, U. (2018). Drawing links between the autism cognitive profile and imagination: Executive function and processing bias in imaginative drawings by children with and without autism. Autism, 22(2), 149–159.
Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Cambridge, MA: Harvard University Press.
Wilkenfeld, D. A. (2014). Functional explaining: A new approach to the philosophy of explanation. Synthese 191, 3367–3391.
Wright, B. M., Ford, R. M., & Coologhan, B. K. (2020). Tell Me About Your Picture: Using Drawings to Support Expressive Language in Children With Autism Spectrum Disorder. Communication Disorders Quarterly, 41(1), 44–55.