Integrating generative AI chatbot beyond classrooms: Exploring students' interactions with chatbot and competencies in systems thinking

Sharon Bee Wah Tan, Yuanli Zheng
Republic Polytechnic, Singapore

Novice learners often struggle to engage with the complexity inherent in systems thinking. This study explores how a customised generative artificial intelligence (GenAI) chatbot can scaffold students' systems thinking processes in a critical thinking and problem-solving skills module at an institute of higher education. Adopting a convergent mixed methods design, the study involved 52 students who interacted with the chatbot and completed an assignment addressing complex sustainability issues. Data from chatbot logs, reflection journals and assignments were analysed using rubrics that scored interaction depth and learning quality. Linear regression revealed a significant relationship between the depth of chatbot interaction and assignment quality. Building on these quantitative scores, content analysis of the chat logs was conducted to identify distinctive behaviours associated with each interaction level. For example, students with deep-level interactions often engaged in iterative questioning and elaboration, while those with moderate-level interactions sought feedback but did not always pursue further dialogue. Finally, interviews with six students provided insight into the factors shaping their interaction patterns, including how they perceived the role of the chatbot and their sense of agency.

Implications for practice or policy:

- Educators could model effective GenAI chatbot use by emphasising the intended pedagogical purpose of GenAI tools.
- Instructional designers could design complementary classroom activities to support diverse learners and extend GenAI-mediated learning.

Keywords: higher education, dialogic learning, dialogic feedback, students, generative AI, systems thinking, mixed methods

Introduction

Empowering individuals to contribute to a sustainable future is about equipping students to make informed decisions that minimise negative environmental impact (Arvai et al., 2004). Systems thinking is increasingly recognised as a vital skill in the 21st century; it helps people understand complex issues by considering all the connected parts and how they influence one another (Voulvoulis et al., 2022). Systems thinking is crucial for tackling complex problems in sustainability, which requires analysis of environmental, social and economic issues in a holistic way. However, the complexity of systems thinking and the interdisciplinary nature of sustainability problems pose significant challenges to students who are new to these concepts.

Dialogic learning is a learning method that enables a two-way communication involving making initiations and providing responses (Song et al., 2019). According to Vygotsky's (1978) sociocultural theory, learning is a cultural process facilitated through interaction and dialogue. By engaging in dialogues, students learn to consider various perspectives and systemic relationships, which is foundational to developing systems thinking. Research by Valbekmo and Bjuland (2023) has emphasised the role of interactive discussions in classrooms designed for analytical thinking, where students collaboratively refine their ideas through reasoned dialogues, thus enhancing their capacity for complex problem-solving. Hence, dialogic learning environments empower students to approach problems from multiple angles, recognise patterns and draw connections. They are essential components for mastering systems thinking and addressing complex challenges in real-world contexts.

Generative artificial intelligence (GenAI) chatbots have the potential to extend dialogic learning beyond classrooms. Powered by transformer architectures such as GPT-4, a GenAI chatbot can stimulate human-like conversations. In the context of education, an educational GenAI chatbot can enable educators to incorporate prompts to scaffold learning (Mollick & Mollick, 2024). In the context of our study, customising prompts in GenAI chatbots can assist students to break down complex systems thinking concepts into manageable components and help them to approach a complex problem gradually. In a classroom setting where short contact hours may be insufficient for the acquisition of systems thinking skills, a GenAI chatbot may complement facilitator-led classroom discussions by providing personalised and timely support (Luckin et al., 2016). However, studies have revealed that students interact differently with GenAI chatbots (Shibani et al., 2024; Winkler & Soellner, 2018). Examining these interaction patterns can reveal whether the tool supports its intended pedagogical role or whether some students engage with it in more limited ways. Such insights are important for educators when integrating GenAI into the curriculum, as they can inform targeted interventions – such as additional scaffolds or demonstrations – to support students who may not yet be using the tool productively. In this way, analysing student-GenAI interactions can help educators refine how the tool is implemented so that students entering the activity with different levels of prior experience are better supported.

Literature review

Systems thinking

Systems thinking involves understanding underlying drivers, interactions and conditions that influence decisions; it helps students to build skills in reframing problems and expanding their perspectives to anticipate unintended consequences (Voulvoulis et al., 2022). Understanding sustainability challenges requires systems thinking to foster holistic approaches to environmental and economic challenges, highlighting the importance of incorporating these perspectives into interdisciplinary curricula (Olvitt et al., 2024; Yurtseven et al., 2016). The alignment between systems thinking and sustainability is highlighted by emerging approaches that integrate technology-enhanced learning methods. Research has indicated that the experience of a flow state during information and communications technology activities can positively influence students' engagement and enhance systems thinking, particularly when clear goals and intrinsic rewards are present (Kurent & Avsec, 2024) This synergy not only supports student learning but also contributes to the achievement of Sustainable Development Goals (United Nations, n.d.) by preparing graduates who are adept at managing the interdependencies of complex systems in real-world scenarios (Olvitt et al., 2024; Yurtseven et al., 2016).

The structure-behaviour-function framework provided a conceptual structure for organising ideas about complex systems (Jordan et al., 2013) and guided the scaffolding and assessment of systems thinking in our study. However, mastering systems thinking is challenging due to its complexity and interdisciplinary nature. Students often struggle to see systems as dynamic and interconnected, instead perceiving them as isolated parts (Assaraf & Orion, 2005). Therefore, the teaching of systems thinking needs to be intentionally scaffolded.

Dialogic learning and feedback using GenAI chatbots

Dialogic learning, as conceptualised by Racionero and Valls (2007), offers a communicative and interactionist approach to achieving equitable education through egalitarian dialogue, where all participants – regardless of cultural or socio-economic background – contribute meaningfully to the construction of knowledge. At its core, dialogic learning is grounded in the idea that learning emerges through the co-creation of meaning in interaction, recognition of cultural intelligence and transformation of both learners and their contexts (Racionero & Valls, 2007). This framework is increasingly relevant in the age of GenAI, particularly when considering how students interact with GenAI chatbots in educational settings. The principle of egalitarian dialogue – central to dialogic learning – can be meaningfully applied to student-chatbot interactions, where learners engage in open-ended, reflective dialogue without fear of judgement or social bias. As Resnick et al. (2018) have noted, dialogic learning requires a shift from valuing only correct, polished answers to valuing the thinking process itself, including tentative, half-formed or emergent ideas. In chatbot-mediated learning, this shift opens space for students to engage in exploratory talk – posing questions, making mistakes and refining their reasoning in real time (Sonderegger & Seufert, 2022). Rather than passively receiving information or memorising facts, students actively engage with content, articulating their reasoning and receiving responsive feedback from the GenAI chatbot. In this way, dialogic interaction with GenAI chatbots not only supports inclusive and participatory learning but also fosters the kind of deep reasoning and systems thinking essential for addressing complex, real-world challenges.

The effectiveness of dialogic learning in fostering critical thinking has been demonstrated empirically. Frijters et al. (2008) found that a dialogic approach to teaching critical thinking – compared to a non-dialogic alternative – helped pre-vocational students develop stronger reasoning skills and a better understanding of the values behind different viewpoints. When applied to learning with GenAI chatbots, dialogic interaction can similarly support the development of complex reasoning skills, especially in domains like sustainability and systems thinking, where multiple perspectives and uncertainties must be navigated. A study by Tang et al. (2024) provided a proof of concept for a dialogic approach in moving students beyond perceiving GenAI as an information retrieval tool to a dialogic and co-constructive partner.

Beyond dialogic interaction, GenAI chatbots provide instant feedback and scaffolding, allowing students to regulate their learning and construct knowledge dynamically. A study by Pahi et al. (2024) comparing the differences between the human teachers and GenAI on quality, depth and timeliness revealed that ChatGPT performs better in presenting clarifying examples and offering encouragement although human teachers are better at identifying students' progress and challenges. A study by Zhu and Carless (2018) has suggested that dialogue activates key cognitive processes and increases the chances of students acting on the feedback. Although the study was conducted in the context of peer feedback, the process can be applied in the context of GenAI chatbot, as the receiver (who is the student) can clarify or negotiate meaning in a safe, non-judgemental environment. Yang and Carless (2013) have posited that reducing the impact of power relations and negative emotions can in turn promote learner agency and self-regulation. Therefore, the use of AI-mediated feedback is one way to circumvent issues encountered using traditional feedback, such as scalability (Carless, 2016; Er et al., 2020), one-way transmission of feedback, timeliness, specificity and clarity (Yang & Carless, 2013). In addition, GenAI is perceived as safe space with a lower stakes environment by the students, though it may not have an equal level of power, credibility and reliability as instructors when it comes to trusting the feedback (Magne et al., 2025). drawi that an implementation of GenAI mediated learning may not be well received by all students, and it is important to identify the challenges faced by these students. Our study therefore contributes to the discourse using GenAI to facilitate dialogic learning and feedback by examining how students interact and act on the feedback within GenAI mediated learning environments.

Student–chatbot interactions and learning engagement

While many studies have highlighted the cognitive benefits of GenAI chatbots, not all interactions result in meaningful learning. Research has identified different levels of interaction with chatbots, ranging from surface-level exchanges (e.g., information seeking) to deeper dialogic interactions (e.g., verifying, clarifying). Shibani et al. (2024) observed shallow levels of interaction with ChatGPT by students during an authentic writing task, such as treating it like a search engine or asking for examples. While rare, there are observations of deeper interaction, which include seeking elaboration and explanation (Shibani et al., 2024). While Shibani et al.'s study did not state what would encourage a deeper interaction with the GenAI chatbots, Wu et al. (2024) suggested that students' acceptance of technology increases the likelihood of students' tendency to explore and apply various technological resources to support their learning.

There are various factors that may shape students' interaction with GenAI chatbots. Winkler and Soellner (2018) have noted that students from different academic backgrounds exhibit varying levels of comfort and interaction with chatbot technology. For instance, students with a humanities background tend to engage in more natural dialogues, while those from technical fields may approach interactions more formally, often testing the limits of the chatbot's capabilities. This variance suggests that educators should consider students' backgrounds when integrating chatbots into learning environments to maximise their effectiveness. Another critical factor influencing learning behaviours is the metacognitive strategies employed by students during their interactions with chatbots. A study conducted by Tjahyana (2024) revealed that students often modify their questions when faced with unresponsive chatbots, demonstrating a proactive approach to learning. This behaviour reflects a deeper interaction with the learning material and indicates that students are applying metacognitive strategies such as planning and evaluation, which are essential for effective learning. Such strategies not only enhance problem-solving skills but also foster a sense of autonomy in learners.

Moreover, the design and feedback mechanisms of chatbots significantly impact student engagement with the chatbot. A well-structured chatbot design and guided instructional use can enhance student engagement by supporting the development of critical thinking and metacognitive skills (Tabib & Alrabeei, 2024). Research has indicated that providing adaptive feedback tailored to each student's needs could encourage individual participation in collaborative learning settings (Liang, 2024). This personalised approach helps to equalise participation among group members, promoting a more inclusive learning environment. Finally, the context of learning also affects how students interact with chatbots. Yin et al. (2021) have highlighted that the low stress environment provided by chatbots could lead to increased motivation and engagement compared to traditional classroom settings. Therefore, understanding the learning behaviours of students in GenAI chatbot interactions involves considering various factors, including academic backgrounds, metacognitive strategies, the chatbot affordances and the overall learning environment. By recognising these elements, educators can better integrate GenAI chatbots into learning environments, ultimately enhancing student engagement and learning outcomes.

Conceptual framework

This study drew on Yang and Carless's (2013) dialogic feedback triangle as a conceptual framework to design a GenAI-mediated dialogic learning and feedback activity to develop students' system thinking of a complex sustainability issue. The feedback triangle comprised the cognitive dimension, socio-affective dimension and structural dimension. The cognitive dimension refers to the content of the feedback which is the systems thinking concepts in this study. Our chatbot also aims to develop students' cue consciousness (Yang & Carless, 2013) by asking specific questions required by the assessment process, giving feedback on what can be done to obtain optimal results and assisting students to appraise the gap between current and desired performance. The socio-affective dimension refers to the social practice of engaging with feedback, including engaging with emotions and students' level of engagement with feedback. The GenAI-mediated feedback is aimed to reduce the impact of power relations and negative affect of the students. Lastly, the structural dimension refers to timing, sequencing and modes of feedback. Students are encouraged to engage with the chatbot independently during the e-learning to prepare for the assignment as a form of formative assessment. The chatbot affords timely, immediate feedback. Scaffolding resources, such as a digital workbook, were used to complement the dialogic learning and feedback activity.

Purpose of the study

While research has explored GenAI chatbots in various educational contexts, less attention has been given to how students' interaction with GenAI chatbots influence their ability to foster systems thinking. Studies have focused primarily on immediate learning gains rather than the cognitive processes underlying chatbot interactions. Moreover, few studies have investigated why some students interact deeply while others interact at a superficial level when using GenAI chatbots for learning. The purpose of this study was to further understand how students' interaction with a customised GenAI chatbot assisted or hindered their ability to develop systems thinking skills. This study applied dialogic learning, feedback and the structure-behaviour-function framework (Jordan et al., 2013) to facilitate systems thinking using a GenAI chatbot.

The study addressed three research questions (RQs):

- RQ1: What is the relationship between the depth of students' interaction with the GenAI chatbot and the quality of their systems thinking outcomes?
- RQ2: What patterns of interaction emerge in students' use of the GenAI chatbot for systems thinking?
- RQ3: What explains the differences in how students interacted with the GenAI chatbot in learning systems thinking?

Methods

Study design and participants

This study adopted a convergent mixed methods design (Creswell & Clark, 2018) to investigate how a GenAI chatbot supports students in developing systems thinking skills. The chatbot was integrated into a module titled Introduction to Critical Thinking and Problem-Solving Skills, in which students explore complex sustainability problems. The study was conducted with 52 first-year students enrolled in an institution of higher education in Singapore. These students were selected through convenience sampling as part of their participation in the mandatory foundation module aimed at developing critical thinking and problem-solving skills. They ranged in age from 17 to 21 years old, 24 females and 28 males. Prior to the study, students were introduced to GenAI and basic prompting practices through a lesson involving a guided interaction with a custom chatbot that was primarily used for information seeking. The purpose of this study was to move students beyond using GenAI tools for information seeking to engaging in dialogic learning.

Data were collected from four sources. First, chatbot conversation logs recorded during the intervention were coded using a three-level rubric representing the depth of chatbot interaction and used as a variable in the statistical analysis. Second, student reflection journals describing how the chatbot influenced their learning were analysed and contributed to the rubric dimension assessing reflection on AI-assisted systems thinking. Third, student assignments proposing solutions to sustainability issues were analysed as formative assessments of students' systems thinking and coded using a three-level rubric. Chatbot interaction scores and assignment performance were analysed using simple linear regression to examine the relationship between interaction depth and learning outcomes. Finally, semi-structured interviews were conducted with six students selected through purposive sampling based on interaction depth (low, moderate, deep), with two students from each category.

By integrating these strands, the study not only identified statistical relationships between the depth of chatbot interaction and systems thinking outcomes but also provided qualitative findings to explain why students interacted that way. This mixed methods approach enabled a more holistic understanding of how GenAI chatbot can scaffold complex cognitive processes like systems thinking in educational contexts.

Procedure

The chatbot used in this study was implemented through a commercially available educational chatbot platform that allows instructors to configure custom prompts to guide student interactions. The prompts were designed based on an adapted structure-behaviour-function systems thinking framework (Jordan et al., 2013). While the interaction followed a scaffolded process, students could navigate the prompts flexibly. The prompts were calibrated to support learning across three key stages: (a) understanding the problem, (b) analysing interrelationships and impacts within the system and (c) proposing potential solutions. A screenshot of the chatbot interface is shown in Figure 1.

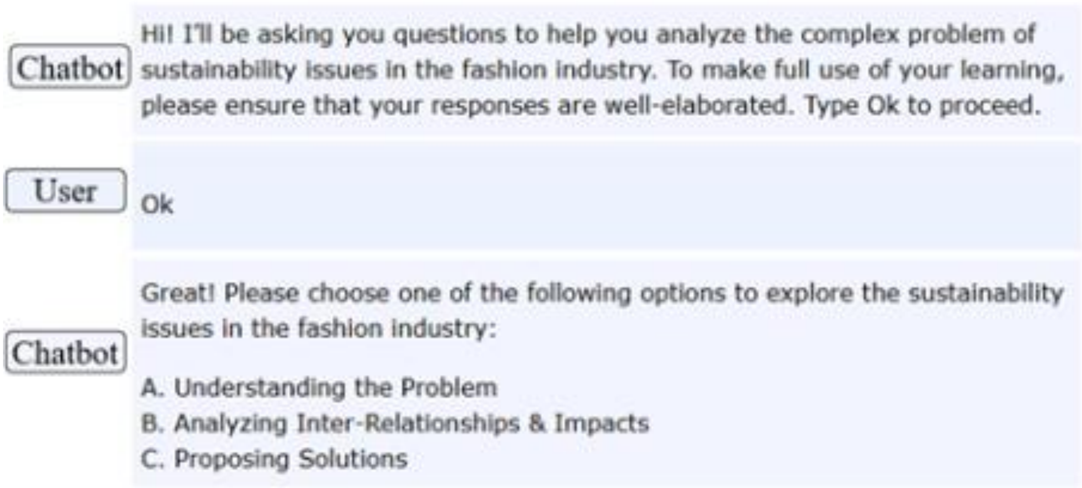

Figure 1. Screenshot of the GenAI chatbot interface

Students interacted with the chatbot through a text-based interface by typing responses to the scaffolded prompts. To respond meaningfully, students often needed to conduct brief research outside the platform (e.g., consulting online sources) before entering their responses. The chatbot was prompted to provide adaptive feedback based on the length and quality of students' replies. For instance, when students provided overly brief responses, the chatbot prompted them to elaborate further and encouraged them to support their ideas with relevant evidence or sources. This interaction structure aimed to guide students towards more thoughtful and evidence-informed responses. The prompts were iteratively refined based on observations of student interactions. For instance, when some students attempted to prompt the chatbot to provide direct answers instead of articulating their own reasoning, additional prompt constraints were introduced to discourage answer generation and redirect students towards elaborating their responses.

The chatbot was accessed through a QR code or web link, allowing students to interact with it using their personal devices. Access to the chatbot was limited to three specific lessons within the module. In the first face-to-face lesson, the instructor demonstrated how the chatbot worked, and students were given approximately 30 minutes to interact with it during class. During this activity, the instructor monitored students' chat logs in real time in order to identify misconceptions related to both the systems thinking concepts and the use of the chatbot. In the second lesson, which was conducted asynchronously as an e-learning session, students were encouraged to continue interacting with the chatbot independently as part of their preparation for an assignment draft. In the third face-to-face lesson, the instructor provided consolidated feedback on students' work, and students were given time to refine their assignments further. Outside of these lessons, the chatbot was not available for student use. Figure 2 outlines the study procedure.

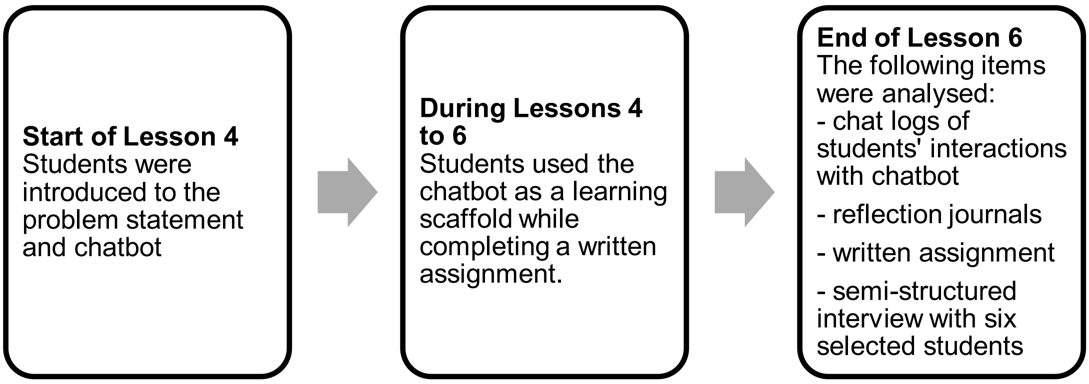

Figure 2. Study procedure conducted across three lessons

Ethical considerations

The study was conducted in accordance with ethical research standards. Informed consent was obtained from all participants, and anonymity was ensured by giving students pseudonyms during data collection, analysis and reporting. Data were securely stored, and only authorised researchers had access to identifiable information.

Findings

RQ1: What is the relationship between the depth of students' interaction with the GenAI chatbot and the quality of their systems thinking outcomes?

The conversation logs and student reflections were analysed using an adapted framework from Shibani et al. (2024), which categorised interactions into dimensions such as critical interaction, exploration of interrelationships and reflection (see Table 1).

Table 1
Coding framework for analysing the depth of student interaction with the GenAI chatbot

Dimension	Code (score)	Description
C1: Critical interaction for understanding and identifying factors and/or stakeholders	Deep (3)	Student demonstrates critical, thoughtful interaction with the chatbot by thoroughly engaging with prompts to identify a broad range of factors and stakeholders involved in sustainability issues in the fashion industry. This includes improving on own responses upon the chatbot's feedback and may include probing for deeper explanations about each factor's influence on the system.
	Moderate (2)	Student engages in surface-level interaction with the chatbot to identify some stakeholders or factors without deep exploration of their roles or impact. They may use the chatbot's suggestions without further questioning or rely on basic responses to complete the identification task.
	Low (1)	Student demonstrates limited or no interaction with the chatbot to identify or understand stakeholders in the sustainability issue.
C2: Critical interaction for analysing interrelationships and impacts	Deep (3)	Student critically interacts with the chatbot to explore complex interrelationships and impacts among stakeholders. They may improve on own responses upon the chatbot's feedback, ask follow-up questions, seek explanations about feedback loops and examine how each stakeholder influences others and the broader system.
	Moderate (2)	Student interacts at a surface level with the chatbot to identify basic relationships among stakeholders without exploring deeper systemic connections. They may only recognise direct impacts or rely heavily on the chatbot's suggestions without probing further.

Dimension	Code (score)	Description
	Low (1)	Student demonstrates no interaction with the chatbot for analysing interrelationships among stakeholders.
C3: Critical interaction for solution exploration and evaluation	Deep (3)	Student thoroughly interacts with the chatbot to explore a range of sustainable solutions, questioning feasibility, potential impacts and ethical considerations. They may improve on own responses upon the chatbot's feedback, critically evaluating each solution's implications for stakeholders and the system.
	Moderate (2)	Student uses the chatbot to gather basic ideas or a limited range of solutions without in-depth evaluation or consideration of broader impacts. They may follow the chatbot's prompts to generate simple solutions but do not critically assess them.
	Low (1)	Student demonstrates no interaction with the chatbot to explore or evaluate sustainable solutions.
C4: Personal reflection on GenAI-assisted systems thinking	Deep (3)	Student reflects critically on their use of the chatbot in learning systems thinking. They consider how the chatbot-supported process helped or hindered their understanding, identifying strengths and limitations of using the chatbot to analyse complex issues. Reflections may include insights on how the chatbot shaped their approach to identifying, analysing or solving sustainability problems.
	Moderate (2)	Student demonstrates basic reflection on the chatbot use, acknowledging the chatbot's role without critical evaluation of its impact on their learning. Reflections are general and lack specific examples or deep insights.
	Low (1)	Student demonstrates no reflection on their interaction with the chatbot and its role in learning systems thinking.
C5: Conversational engagement with GenAI	Deep (3)	Student engages in dynamic, iterative conversations with the chatbot, using prompts to refine their understanding or to challenge initial responses. They expand on or critique the chatbot-generated suggestions, exploring alternatives and building deeper understanding through back-and-forth interactions.
	Moderate (2)	Student uses the chatbot in a transactional manner, following prompts in a linear way without engaging in deeper conversation. They accept the chatbot's responses with minimal further questioning.
	Low (1)	Student interacts minimally with the chatbot, only inputting required responses or skipping conversations entirely.

We coded each log independently. Inter-rater reliability was assessed using quadratic weighted Cohen's kappa, appropriate for the ordinal rubric scale. Quadratic weights were used to account for the magnitude of disagreement between rubric scores. The resulting agreement was $\kappa = .53$, indicating moderate inter-rater reliability. This level of agreement is not uncommon in ordinal, performance-based assessments where raters must interpret varying degrees of quality rather than make simple categorical judgements, and where partial disagreements across adjacent categories are expected (Sim & Wright, 2005). Discrepancies were resolved subsequently through consensus discussions to produce the final coded data set. Final scores for all 52 participants were calculated by summing the ratings across dimensions (see Table 2).

Table 2
Frequency table of the depth of interaction of the GenAI chatbot (based on 52 participants' final scores)

Depth	Frequency
Deep	20
Moderate	11
Low	21

The assignments that evaluated the quality of students' systems thinking outcomes were analysed using a systems thinking rubric aligned to the requirements of the module. The rubric assessed elements such as identification of factors and stakeholders that relate to the complex problem, explanation of inter-relationships and the quality of proposed solutions (see Table 3). The scores were summed to derive the final score. As with the conversation logs, we independently scored the assignments, and the final scores were determined through alignment meetings (see Table 4).

Table 3
Systems thinking rubric (aligned to requirements of the module)

Dimension	Code (score)	Description
Identification and explanation of factors and stakeholders	Emerging (1)	Identifies a few factors or stakeholders with minimal explanation. Shows limited understanding of each stakeholder's role in the sustainability issue.
	Developing (2)	Identifies relevant factors or stakeholders and provides explanations with some depth. Shows partial understanding of each stakeholder's role.
	Proficient (3)	Identifies most relevant factors and stakeholders with clear, detailed explanations. Demonstrates solid understanding of each stakeholder's role and relevance.
Explanation of interrelationships and impacts	Emerging (1)	Provides minimal or unclear explanation of interrelationships between factors and stakeholders. Little to no recognition of impacts or feedback loops.
	Developing (2)	Identifies some interrelationships and explains impacts with moderate depth. Focuses on direct connections but lacks broader exploration of systemic effects.
	Proficient (3)	Explains key interrelationships and impacts in detail, showing understanding of systemic connections, indirect impacts and feedback loops.
Explanation of proposed solution (specific to institution context)	Emerging (1)	Proposes a general solution with limited connection to the institution's context.

		Considers few stakeholders and demonstrates little understanding of community impact.
	Developing (2)	Proposes a solution relevant to the institution's context with moderate consideration of multiple stakeholders. Shows some understanding of impacts on the community.
	Proficient (3)	Proposes a well-suited solution for the institution's context, with thoughtful consideration of multiple stakeholders' needs and impacts. Demonstrates understanding of feasibility and outcomes.
Use of credible sources	Emerging (1)	Uses few credible sources, relying on general or non-academic sources. Sources lack relevance or depth.
	Developing (2)	Uses some credible sources with moderate relevance and depth. Sources provide support but lack diversity or rigour.
	Proficient (3)	Uses a variety of credible, relevant sources, including academic research, to support explanations and claims. Demonstrates careful selection of sources to enhance depth and reliability.

Table 4
Frequency table of the quality of systems thinking outcomes (based on 52 participants' final scores)

Quality	Frequency
Proficient	16
Developing	19
Emerging	17

A simple linear regression was conducted to examine the relationship between the depth of GenAI chatbot interaction and the quality of students' systems thinking outcomes. Prior to analysis, assumptions of normality, linearity and homoscedasticity were checked and satisfied. Pearson correlation indicated a significant positive association between the variables, $r = .573$, $p < .001$.

The regression model was statistically significant, $F(1,50) = 24.439$, $p < .001$ and explained 32.8% of the variance in students' systems thinking outcomes ($R^2 = .328$, Adjusted $R^2 = .315$). Depth of chatbot interaction significantly predicted systems thinking outcome, $b = 0.607$, SE = 0.123, $t = 4.944$, $p < .001$, 95% CI [0.361, 0.854], indicating that for every one-unit increase in chatbot interaction depth, systems thinking scores increased by 0.607 units on average. The standardised coefficient ($\beta = .573$) further indicates a moderate-to-strong positive effect.

RQ2: What patterns of interaction emerge in students' use of the GenAI chatbot for systems thinking

Drawing on the final scores derived from the analysis of students' interaction depth with the chatbot (as discussed in RQ1), we categorised the participants into three groups of interaction levels: deep (score 10–15), moderate (score 5–9) and low (score 0–4). We conducted a follow-up analysis to examine the distinctive characteristics of chatbot interactions within each group. A codebook was developed for this purpose. We independently selected random samples from each category and analysed in detail the interaction patterns observed. Initial codes were derived from these characteristics and iteratively refined to ensure consistency and alignment with the data. We then independently coded the remaining samples, with discrepancies resolved through multiple rounds of consensus discussions. This analysis aimed to uncover patterns in how students interacted with the chatbot and what these patterns revealed about their learning behaviours. Refer to Table 5 for the levels and characteristics of chatbot interactions and illustrative examples.

Table 5
Patterns of interaction with the GenAI chatbot and examples

Interaction levels	Codes of characteristics of chatbot interaction	Sample chatbot prompts from students
Deep	Iterative revision of user response based on chatbot's feedback Initiate practice Request new information or ideas Seek clarification on content Seek clarification on tasks Request for feedback Organise ideas or content	"I want to practice on the factors again" (Elmo) "What factors am I missing?" (Elmo) "What do you think of a space where students and staff can donate used clothes and do a swap if they want an item" (Sya) "Can you show me my answers for a, b and c" (Andy) "Am I supposed to be using the workbook?" (Elmo)
Moderate	One time revision of user response based on chatbot's feedback Seek clarification on content Seek clarification on task Request for new information or ideas	"Can you elaborate on the levels in the nested system?" (Umi) "Do I need to answer this in a particular format?" (Spencer) "Can you give me an example?" (Spencer)
Low	Seek clarification on task Seek clarification on content Seek direct answers	"Can you remind me what are the different sections and what we have to do for this assignment due in Lesson 06" (Iris) "Interrelated? How so?" (Mung) "How does individuals contribute to the sustainability issues in the fashion industry" (Stanly)

The deep interaction level reflects the most sophisticated interactions with the chatbot. Students demonstrating deep interaction iteratively revised their own responses, clarifying both task and content with persistence. This level was also characterised by proactive efforts to request the chatbot to summarise and organise ideas, seek additional feedback and request new information to deepen their responses. The consistent pattern of back-and-forth refinement and active feedback use suggested a high degree of critical thinking and reflective learning, where students not only followed the chatbot's guidance but expanded on it to generate more thoughtful contributions.

In contrast, students interacting at the moderate interaction level sought clarification of tasks and content but often stopped short of deeper inquiry. Although they occasionally revised their work, they did so with less frequency and depth compared to the deep interaction group. Feedback use was more surface level, with limited evidence of multiple cycles of revision or significant refinement of ideas. While students did interact with the chatbot in meaningful ways, their interactions were more transactional, using the feedback primarily for clarifying rather than expanding their understanding. The focus remained on task completion rather than ongoing exploration or self-driven improvement.

At the low interaction level, interactions with the chatbot were minimal and surface level. Students frequently clarified on the task but without significant contribution to the dialogue. Clarification efforts were limited, and when feedback was requested, it was often not acted upon meaningfully. In several instances, students ignored prompts or provided superficial responses. The interaction pattern in this category also indicated a strong reliance on the chatbot to provide answers rather than using it as a tool for deeper learning. Critical thinking was largely absent, with little evidence of self-directed exploration or iterative refinement of ideas.

RQ3: What explained the differences in how students interacted with the GenAI chatbot in learning systems thinking?

Content analysis of the semi-structured interviews was conducted to address RQ3. Six students were purposefully selected for the interview based on their variability in interaction levels with the chatbot. Each student's transcription was coded by one of us and verified by the other. We held several meetings until consensus was achieved. The purpose of the interview was to uncover the reasons why different students interacted differently with the chatbot. A detailed description of the students' interaction characteristics and reasons are provided below.

Deep level

Sally and Kelvin demonstrated a deep level of interaction, distinguished by iterative revisions based on the chatbot's feedback which was not present in the moderate and low interactions. The reasons cited include Kelvin's motivation to achieve a good grade, leading him to use the chatbot to check the accuracy of his answers, and Sally's desire for a sounding board, with the chatbot being treated as a conversational partner to explore ideas. Even though both fall into the category of deep interaction, the way they interacted was different. Kelvin was intent on iteratively refining his answers to the point when the chatbot deemed them to be satisfactory. As a result, he fed the chatbot with many different iterations, constantly seeking feedback. On the other hand, Sally came from an angle of exploration, using the chatbot to seek clarifications when questions were vague and politely thanking the chatbot after the interactions. While Kelvin recognised the usefulness of the chatbot in detecting his blind spots, Sally appreciated its assistance in helping to broaden her perspectives:

> I understood a bit more, because I'm not really exposed to this kind of thing, especially when it comes to sustainability. So, when I asked the chatbot to give me some solutions, it opened my eyes a bit more. I wasn't sure, because I am quite narrow-minded at times. So, when it gave me certain examples, I realised that, oh, not everything is one-sided. (Sally)
>
> I want assurance that my explanation, the evidence, and whatever I did, was correct … When I don't understand the feedback … I asked it, and it gives me steps on how to do it. Even if the feedback is very vague, you can ask it to help you narrow down and develop your answer. (Kelvin)

Moderate level

Kelly and Joy demonstrated a moderate level of interaction. One distinguishing characteristic different from other levels is a once off revision based on chatbot's feedback. Both Kelly and Joy perceived the chatbot as a tool to validate their answers. Both reported working out the answers on a separate word processing document before copying and pasting into the chatbot to verify whether their answers were aligned with expectations. The reasons cited include external reasons such as time constraints due to impending deadline to submit the assignment and slow chatbot response speed from Kelly, as well as internal reasons such as feeling overwhelmed by the chatbot's questions from Joy. The reasons explained that these two students may interact further if the external and internal constraints are removed. Although the interactions were not as iterative as those of Kelvin and Sally, they recognised that the chatbot provided specific and constructive feedback and chose to engage with the chatbot as best as they could, given the constraints they faced:

> I attempted to use the feedback, but then I realised the submission time was going to close, so I just submitted the work without fully finishing. (Kelly)

> I formed my initial answers outside the chatbot. Then once I get a general idea of what I want to write with the explanations, I put them into the chatbot, then it will tell me how to expand. (Joy)

Low level

Zully and Mung demonstrated a low level of interaction with the chatbot, primarily focused on seeking clarification on task and content rather than on seeking feedback or exploring new ideas compared to the other students above. The reasons cited included the lack of direct answers from the chatbot hindered learning and confusion about the assignment task from Mung, as well as limited AI literacy of the affordance of the chatbot from Zully. Mung exited the chatbot after responding to its first question as he was looking for more concrete suggestions but the chatbot returned questions instead. Meanwhile, Zully saw the chatbot as a tool for structuring and organising answers for clarity. She did not recognise its potential to go beyond structuring answers. In fact, the chatbot can assist in ideas generation, explain concepts and provide feedback:

> Because I only started using ChatGPT last year. I mainly use it (the chatbot) to help me reword some phrases, rephrase some of the content that I needed, or to generate more ideas. (Zully)

> If the student knows what they're doing, the AI will help a lot. If the student doesn't know what they're doing, it won't help … Though it's interactive, it's not humanly interactive ... At least the lecturer will know where you're lost and stuff; AI won't know where you are lost. (Mung)

These findings revealed that not all students interacted with the chatbot the same way. Reasons for higher levels of interactions included internal motivation and curiosity to explore the topic more. Low-level interactions were categorised by the presence of external and internal barriers. External barriers included time constraints and user-friendliness of the chatbot. Internal barriers that influence chatbot interaction included students' limited AI and feedback literacy.

Discussions and recommendations

Overall, the quantitative results indicate a significant relationship between the depth of chatbot interaction and the quality of the assignment assessing students' systems thinking and complex problem-solving skills. The qualitative findings suggest a variability in interaction levels, which indicate that not all students naturally interacted with GenAI tools in a meaningful way. As echoed by Magne et al. (2025), using GenAI as a dialogic partner may not yield universal benefits for all students. The differences in interaction levels suggest varying levels of self-regulated learning and feedback literacy (Carless & Boud, 2018). Differentiated support may be necessary to move students from lower to higher interaction. The low interaction category revealed the tendency for students to seek quick answers without genuine attempts to engage in revision or deeper inquiry. This highlights the risk of passive learning when using GenAI tools, where students may perceive the chatbot as a shortcut rather than a partner in the learning process. Winstone et al. (2017) argued that engaging well with feedback is challenging and multi-factorial. A few reasons offered to explain why students may be reluctant to engage with feedback included not knowing how to use feedback effectively and unrealistic expectations about the feedback specifying what they should do exactly. According to Carless and Boud (2018), educators play an important role in providing guidance, coaching and modelling. This includes explaining the rationale and potential benefits, as well as addressing barriers the students may face. To increase action on feedback by students, educators play an important role in a collaborative manner with GenAI through encouraging practices that increase self-regulation. For example, educators can encourage students to share responsibility for their own learning trajectory through demonstrating participation in dialogue with the chatbot and giving students opportunity to clarify the feedback with the human educators.

Moreover, students' perception of the chatbot directly influenced the depth of their cognitive engagement. Sally, who viewed the chatbot as a conversational partner, demonstrated the deepest interaction, using the tool for critical thinking, clarification and expansion of her ideas. Conversely, Kelly, who perceived the chatbot as an answer validation tool, limited her use to checking correctness without meaningful exploration, while Zully, who saw it as a structuring aid, focused on surface-level organisation and information retrieval rather than deeper conceptual understanding. These findings echo a study from Antony and Ramnath (2023), who highlighted that students' preferences for using AI chatbots were influenced by their perceptions of the chatbot's role in enhancing engagement and support. Students who viewed chatbots as effective and useful were more likely to engage meaningfully, while those with reservations about the chatbot's role exhibited limited interaction This suggests that how students conceptualise the tool can either amplify or constrain their learning potential. Therefore, educators should emphasise the intended pedagogical purpose of GenAI tools, positioning them as interactive learning aids that improve educational outcomes. This could be achieved by explicitly modelling productive ways to interact with the tool during instruction, such as showing students how to use the chatbot as a space for critical dialogue.

Limitations and future directions

Several limitations should be noted. First, the analysis of conversation logs and assignments is primarily qualitative, which may introduce coder bias; however, efforts were made to establish inter-rater reliability through collaborative coding and discussion. Second, the relatively small sample size for interviews may limit generalisability, although the study provides rich illustrations of interaction types that may be relevant to other educational contexts. Third, although the chatbot prompts and feedback were iteratively refined during implementation based on observations of student interactions in the chat logs, designing AI-assisted feedback that consistently supports deeper reasoning remains a challenge. For example, some students attempted to circumvent the intended interaction by prompting the chatbot to provide direct answers rather than articulating their own responses. In response, additional guardrails were incorporated into the prompts to discourage answer generation and redirect students towards explaining their reasoning. Despite these refinements, ensuring that AI-mediated prompts consistently promote productive learning behaviours remains an ongoing design challenge for future research. Lastly, systems thinking competencies may not be fully captured through written assignment quality alone. Future research may therefore consider moving beyond written assignments to include more authentic project presentations and reflections on the problem-solving process for complex problems, while continuing to investigate how students’ interactions with GenAI chatbots relate to their systems thinking and reasoning processes.

Conclusion

To date, literature has largely focused on the outcomes of GenAI-mediated learning or the technical design of chatbots. This study instead examined students’ interaction processes with a customised GenAI chatbot within a specific instructional context. Drawing on chatbot conversation logs, assignment data and student reflections, the findings indicate a relationship between the depth of chatbot interaction and the quality of students’ assignment work and reveal distinctive interaction behaviours associated with different levels of engagement. While the small sample size limits generalisability, the study provides a nuanced, situated account of how a GenAI chatbot can function as an adjunct to learning under educator guidance, and suggests directions for future research into how students’ prior knowledge of GenAI and their perceptions of the role of GenAI chatbots may influence their patterns of use and dialogue.

Author contributions

Sharon Bee Wah Tan: Conceptualisation, Investigation, Data analysis, Writing – first draft, Writing – review and editing; **Yuanli Zheng:** Conceptualisation, Investigation, Data analysis, Writing – review and editing.

Acknowledgements

The authors would like to acknowledge Dr Poh Kiaw Yeong, who provided consultation during the conceptualisation stage of the project.

References

Antony, S., & Ramnath, R. (2023). A phenomenological exploration of students' perceptions of AI chatbots in higher education. *IAFOR Journal of Education*, *11*(2), 7–38. https://doi.org/10.22492/ije.11.2.01

Arvai, J. L., Campbell, V. E. A., Baird, A., & Rivers, L. (2004). Teaching students to make better decisions about the environment: Lessons from the decision sciences. *The Journal of Environmental Education*, *36*(1), 33–44. https://doi.org/10.3200/JOEE.36.1.33-44

Assaraf, O. B., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, *42*(5), 518–560. https://doi.org/10.1002/tea.20061

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Er, E., Dimitriadis, Y., & Gašević, D. (2021). A collaborative learning approach to dialogic peer feedback: A theoretical framework. *Assessment & Evaluation in Higher Education*, *46*(4), 586–600. https://doi.org/10.1080/02602938.2020.1786497

Frijters, S., Ten Dam, G., & Rijlaarsdam, G. (2008). Effects of dialogic learning on value-loaded critical thinking. *Learning and Instructio*n, *18*(1), 66–82. https://doi.org/10.1016/j.learninstruc.2006.11.001

Jordan, R. C., Hmelo-Silver, C., Liu, L., & Gray, S. A. (2013). Fostering reasoning about complex systems: Using the aquarium to teach systems thinking. *Applied Environmental Education & Communication*, *12*(1), 55–64. https://doi.org/10.1080/1533015X.2013.797860

Kurent, B., & Avsec, S. (2024). Synergizing systems thinking and technology-enhanced learning for sustainable education using the flow theory framework. *Sustainability*, *16*(21), Article 9319. https://doi.org/10.3390/su16219319

Liang, W. (2024). Exploring a framework for supporting self-directed English learning through ChatGPT: A meta-cognitive strategy. In C. Bonk & G. Marks (Eds.), *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 8–11). Association for the Advancement of Computing in Education. https://www.learntechlib.org/primary/p/224976/

Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson. https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf

Magne, V., Vince, S., Mace, R., & Hooper, S. (2025). AI as a dialogic partner: Rethinking feedback in higher education. *Assessment & Evaluation in Higher Education*, 1–15. https://doi.org/10.1080/02602938.2025.2587247

Mollick, E., & Mollick, L. (2024). *Instructors as innovators: A future-focused approach to new AI learning opportunities, with prompts*. arXiv. https://doi.org/10.48550/arXiv.2407.05181

Olvitt, L., Davies, M., Ebrahim, A., & Cockburn, J. (2024). Transdisciplinary curriculum design for sustainability transitions: A reflective dialogue. *South African Journal of Science*, *120*(9/10). https://doi.org/10.17159/sajs.2024/17884

Pahi, K., Hawlader, S., Hicks, E., Zaman, A., & Phan, V. (2024). Enhancing active learning through collaboration between human teachers and generative AI. *Computers and Education Open*, *6*, Article 100183. https://doi.org/10.1016/j.caeo.2024.100183

Racionero, S., & Valls, R. (2007). Dialogic learning: A communicative approach to teaching and learning. In J. L. Kincheloe & R. A. Horn, Jr. (Eds.), *The Praeger handbook of education and psychology* (pp. 548–557. Praeger. https://doi.org/10.5040/9798216000198.ch-065

Resnick, L. B., Asterhan, C. S., Clarke, S. N., & Schantz, F. (2018). Next generation research in dialogic learning. In G. E. Hall, L. F. Quinn, & D. M. Gollnick (Eds.), *The Wiley handbook of teaching and learning* (pp. 323–338). Wiley. https://doi.org/10.1002/9781118955901.ch13

Shibani, A., Knight, S., Kitto, K., Karunanayake, A., & Shum, S. B. (2024). Untangling critical interaction with AI in students' written assessment. In F. F. Mueller, P. Kyburz, J. R. Williamson, & C. Sas (Eds.), *Extended Abstracts of the CHI conference on Human Factors In Computing Systems* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3613905.3651083

Sim, J., & Wright, C. C. (2005). The Kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257–268. https://doi.org/10.1093/ptj/85.3.257

Sonderegger, S., & Seufert, S. (2022). Chatbot-mediated learning: Conceptual framework for the design of chatbot use cases in education. In M. Cukurova, N. Rummel, D. Gillet, B. McLaren, & J. Uhomoibhi (Eds.), *Proceedings of the 14th International Conference on Computer Supported Education* (Vol. 1, pp. 207–215). Science and Technology Publications. https://doi.org/10.5220/0010999200003182

Song, Y., Huang, L., Zheng, L., Fan, M., & Liu, Z. (2025). Interactions with generative AI chatbots: Unveiling dialogic dynamics, students' perceptions, and practical competencies in creative problem-solving. *International Journal of Educational Technology in Higher Education, 22*, Article 12. https://doi.org/10.1186/s41239-025-00508-2

Tabib, F. M., & Alrabeei, M. M. (2024). Can guided ChatGPT use enhance students' cognitive and metacognitive skills? In A. Al-Marzouqi, S. A. Salloum, M. Al-Saidat, A. Aburayya, & B. Gupta (Eds.), *Artificial intelligence in education: The power and dangers of ChatGPT in the classroom* (Vol. 144, pp. 143–154). Springer. https://doi.org/10.1007/978-3-031-52280-2_10

Tang, K.-S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024). A dialogic approach to transform teaching, learning & assessment with generative AI in secondary education: A proof of concept. *Pedagogies: An International Journal*, *19*(3), 493–503. https://doi.org/10.1080/1554480X.2024.2379774

Tjahyana, L. J. (2024). Exploring AI chatbot development for Gen-Z: A study on first-time and experienced voters in Pemilu. *Jurnal Ilmu Komunikasi Dan Bisnis*, *9*(2), 224–236. https://doi.org/10.36914/1q9cb862

United Nations. (n.d.). *Sustainable development goals*. https://sdgs.un.org/goals

Valbekmo, I., & Bjuland, R. (2023). Students' reasoned dialogs during problem solving in a Norwegian thinking classroom. *NOMAD Nordic Studies in Mathematics Education*, *28*(1–2), 56–78. https://doi.org/10.7146/nomad.v28i1-2.149255

Voulvoulis, N., Giakoumis, T., Hunt, C., Kioupi, V., Petrou, N., Souliotis, I., Vaghela, C., & Binti Wan Rosely, W. (2022). Systems thinking as a paradigm shift for sustainability transformation. *Global Environmental Change, 75*, Article 102544. https://doi.org/10.1016/j.gloenvcha.2022.102544

Vygotsky, L. S. (1980). *Mind in society: Development of higher psychological processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Winkler, R., & Soellner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Proceedings*, *2018*(1), Article 15903. https://doi.org/10.5465/AMBPP.2018.15903abstract

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). 'It'd be useful, but I wouldn't use it': Barriers to university students' feedback seeking and recipience. *Studies in Higher Education*, *42*(11), 2026–2041. https://doi.org/10.1080/03075079.2015.1130032

Wu, D., Zhang, S., Ma, Z., Yue, X.-G., & Dong, R. K. (2024). Unlocking potential: Key factors shaping undergraduate self-directed learning in AI-enhanced educational environments. *Systems*, *12*(9), Article 332. https://doi.org/10.3390/systems12090332

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education, 18*(3), 285–297. https://doi.org/10.1080/13562517.2012.719154

Yin, J., Goh, T.-T., Yang, B., & Xiaobin, Y. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research*, *59*(1), 154–177. https://doi.org/10.1177/0735633120952067

Yurtseven, M. K., Buchanan, W. W., & Pilant, M. S. (2016). Introducing sustainability concepts into university curricula through systems thinking. *Sociology Study, 6*(9), 583–590. https://doi.org/10.17265/2159-5526/2016.09.004

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development*, *37*(4), 883–897. https://doi.org/10.1080/07294360.2018.1446417

Corresponding author: Sharon Tan, sharon_tan_bee_wah@rp.edu.sg

Please cite as: Tan, S. B. W., & Zheng, Y. (2026). Integrating generative AI chatbot beyond classrooms: Exploring students' interactions with chatbot and competencies in systems thinking. *Australasian Journal of Educational Technology*, *42*(4), 96–112. https://doi.org/10.14742/ajet.10834