

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

"Artificial Unconsciousness": What toilet graffiti reveals about the limits of university mental health support in the age of AI

Ruby-Ngoc Nguyen

University of Sydney

This paper interrogates the tension between institutional deployments of AI chatbots for student wellbeing and the informal, peer-to-peer support networks that emerge in university bathroom spaces. Anchored in the student-inscribed phrase "ARTIFICIAL UNCONSCIOUSNESS CHATBOT," the study undertakes a cross-disciplinary synthesis of two bodies of literature: systematic reviews evaluating the effectiveness and limitations of AI chatbots in mental health support, and empirical studies of bathroom graffiti as a mode of collective care. Drawing on Cohen's (1997) theory of the safe house, Jung's (1966) wounded healer archetype, and Heidegger's (1962) phenomenology of authentic being-with, we analyse why graffiti walls succeed in fostering protected vulnerability, temporal spaciousness, and existential co-presence, while institutional AI systems often collapse these conditions into surveillance, efficiency, and artificial intimacy. The analysis culminates in the Human-AI Synergy Model, which reconceptualises AI not as a therapeutic substitute but as a relational relay: a tool that can preserve anonymity, scaffold temporal spaciousness, and reduce cognitive burden while pointing students toward authentic human connection. By contrasting what is systematically erased in institutional AI systems with what flourishes in informal peer safe houses, the paper offers critical guidance for designing AI in higher education that enhances rather than erodes the phenomenological conditions of genuine care.

Keywords: artificial intelligence, peer support, student mental health, bathroom graffiti, human-AI synergy, phenomenology of care, wounded healers

Introduction

The rise of generative AI has introduced a new paradox into student life. On one hand, students are encouraged to embrace AI as a study companion or an always-on assistant capable of offering explanations, feedback, or support at any hour. On the other hand, institutional messaging has warned them that the very same tools might compromise their academic integrity, distort their learning, or trigger misconduct investigations. AI is framed as both an enabler and a threat, a friend and foe. This contradictory framing does more than confuse; it produces a heightened state of affective ambivalence, where students are asked to navigate the promise of support under the shadow of surveillance. This ambivalence is not confined to academic tasks. As AI is introduced into mental health services via chatbots and wellbeing platforms, students are again invited into intimacy with a system they are not sure they can trust. Even when institutions promote these tools as stigma-free or accessible, their association with data capture, institutional risk management, and compliance messaging lingers. In this climate, seeking help becomes an act of self-monitoring, not just for signs of distress, but for signs of exposure.

It is in this atmosphere that the bathroom stall reasserts itself as a cultural and emotional safe house. Historically, students have turned to bathroom walls to process experiences too volatile, personal, or stigmatised to surface in formal channels such as grief, sexuality, abortion, racism, mental illness. The arrival of AI into the social and emotional fabric of student life is simply the latest such disruption. Scrawled across the inside of a university bathroom stall, the phrase *ARTIFICIAL UNCONSCIOUSNESS CHATBOT* (Figure 1) is both witty and diagnostic, as it locates AI within the lineage of controversial social changes that students experience not just intellectually but somatically, through anxiety, mistrust, and the search for resonance.

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change



Figure 1. Bathroom graffiti “Artificial Unconsciousness Chatbot”

The graffiti reads less like vandalism than a jarring dispatch from the current technological moment. It names, albeit accidentally, a deep ambivalence in contemporary higher education: as institutions accelerate the adoption of AI-powered support systems, students continue to reach for older, analogue forms of connection. While administrative offices promote chatbot interfaces and digital platforms, restroom walls fill with handwritten messages—confessions, grievances, advice—passed anonymously from student to student. That this analogue ecosystem thrives alongside institutional technologies speaks volumes. It hints at an unmet need, one not easily captured by dashboards or usage metrics.

Across global universities, AI chatbots have been positioned as scalable solutions for academic support, administrative queries, and wellbeing interventions. Their promise lies in efficiency and access: they respond instantly, personalise learning, and free educators from repetitive tasks. Yet behind this promise, student experiences remain uneven. Empirical evaluations illustrate the tension clearly. Ayers et al. (2023), for example, found that while ChatGPT could provide factually accurate medical information, its attempts at empathy were often inconsistent and formulaic, leaving users dissatisfied with its affective quality. Similarly, Campbell et al. (2023) reported that hallucinations (factually false but confident and plausible-sounding answers) were common when chatbots were applied to clinical scenarios, raising concerns about trustworthiness. In surgical and diagnostic contexts, several studies by Ali et al. (2023) and Cakir et al. (2024) confirmed that chatbots could generate plausible advice but often missed contextual nuance or human attunement. Together, these findings illustrate why many students characterise AI interactions as emotionally flat or inauthentic, an observation captured in the graffiti phrase itself.

What often goes unexamined in such accounts is the nature of support itself. What does it mean to feel supported beyond receiving relevant information or resolve a task, but to be seen, heard, and held in a moment of uncertainty? The literature on peer support offers some answers. Unlike professional services or algorithmic agents, peer networks rest on reciprocity and shared experience. They flourish where there is mutual vulnerability, not credentialed authority. They thrive not in systems but in relationships.

Empirical research confirms these dynamics. Classic work has shown that peer support fosters academic persistence, eases transitions into university life, and buffers the psychological impact of stress (Basson & Rothmann, 2018; Ramsay et al., 2007). More recent intervention studies reinforce that benefits extend to both givers and receivers. For example, Fitzpatrick and colleagues (2017) demonstrated in a randomised controlled trial that students using the Woebot chatbot for two weeks showed significant reductions in depressive symptoms compared to an information-only control group, while Inkster et al. (2018) found similar though modest improvements in depression in a pilot university trial. Yet both studies also highlight the limits of chatbot-based care: effects were small to medium, often short-lived, and highly dependent on student engagement. By contrast, peer support interventions continue to show improvements in wellbeing alongside increases in self-awareness and social confidence for student supporters themselves (Hogan & Schmidt, 2022). Importantly, experimental work demonstrates that human-AI collaboration may offer a middle path. Sharma et al. (2023) found that when AI was used to scaffold peer-to-peer text conversations about mental health, dialogue was perceived as more empathic and supportive than AI alone, suggesting potential for synergy rather than substitution.

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

The context of early university life makes this especially salient. First-year students face a convergence of social dislocation, academic pressure, and identity disruption. Institutional supports are often formal, impersonal, or difficult to access; peer support, by contrast, tends to meet students where they are, on campuses, in corridors, and sometimes, in bathrooms. Here, the transition from vulnerability to resilience is not routed through platforms or policies, but through conversation, commiseration, and shared stories of survival (Ramsay et al., 2007).

Universities have attempted to formalise this dynamic through structured peer mentoring programs, yet challenges of sustainability and authenticity persist. Informal networks that emerge without institutional oversight appear to offer something different: spontaneous empathy, grounded in lived experience, unfiltered by reporting protocols (Byrom et al., 2023). Among the most overlooked of these informal networks are the graffiti-covered walls of university restrooms. Long dismissed as vandalism, they are, in fact, repositories of crude but often deeply insightful student voices. Ethnographic studies reveal how bathroom graffiti operates dialogically, with students responding to one another's messages in chains of commentary on topics from academic stress to sexuality to existential doubt (Schreer & Strichartz, 1997; Rodriguez & Clair, 1999; Victoria, 2024). Arrows link comments across time; conversations unfold asynchronously. These spaces are not just expressive; they are relational micro-publics, enacting a subterranean peer care often more responsive than official services.

What emerges, then, is a compelling contrast: AI systems offering polished, data-driven efficiency, and bathroom graffiti offering messy, emotionally resonant dialogue. The former is monitored, structured, and institutionally sanctioned; the latter is subversive, anonymous, and student-initiated. One operates through algorithmic pattern recognition, the other through the unpredictable, situated logic of human connection. Yet both aim to provide a kind of support.

At stake in this contrast is not simply a matter of preference but a deeper epistemological and phenomenological divide. What kinds of knowledge count as support? What kinds of interactions feel real, helpful, or healing? In the phrase *artificial unconsciousness*, we glimpse a student's instinctive critique: while artificial intelligence may offer answers, it lacks the tacit emotional labour and co-presence that define authentic care. Studies across clinical and educational settings confirm this: Miner et al. (2020) showed that participants disclosed more sensitive personal information to chatbots than to human interviewers, but this disclosure lacked reciprocal resonance; Larsen et al. (2021) found that adolescents using a chatbot for suicide risk screening were more open than in face-to-face assessments, yet the interaction still failed to provide the comfort associated with genuine peer exchange. These tensions illustrate why students often navigate a hybrid landscape of turning to AI for academic tasks or disclosure, while seeking authentic human resonance in informal, peer-driven spaces such as the bathroom wall.

Theoretical Framework

This study draws on multiple theoretical traditions to interrogate the tension between institutional AI support systems and student-generated peer networks in bathroom spaces. Specifically, it integrates frameworks from phenomenology, queer studies, peer support research, and clinical psychology to explain how authentic care emerges outside formal institutional channels.

Cohen's (1997) concept of the *safe house* is foundational to understanding bathroom graffiti as a protected environment for vulnerable expression. In *Punks, Bulldaggers, and Welfare Queens: The Radical Potential of Queer Politics?*, Cohen theorises safe houses as spaces carved out from normative power structures where marginalised people articulate needs and identities otherwise silenced in formal contexts. Bathroom stalls mirror these conditions: anonymity, ephemerality, and physical privacy create a "counter-site" where vulnerability can be voiced without fear of repercussion.

In parallel, Jung's (1966/2014) archetype of the *wounded healer* has been influential in psychology and counselling. The idea that those who have endured suffering are uniquely positioned to offer healing resonates strongly in contemporary peer support literature. Empirical studies show that peer supporters with lived experience provide distinctive therapeutic value by demonstrating empathy, patience, and faith in recovery that professionals may lack (Repper & Carter, 2011; Solomon, 2004). White (2000) further expands

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

the wounded healer framework within addiction recovery, describing how lived experience becomes a resource for collective healing. When combined with Cohen's safe house, the wounded healer archetype illuminates bathroom graffiti as a site of "distributed wounded healing", an unstructured yet powerful peer-support network unmediated by professional codes or institutional oversight.

Heidegger's (1962) phenomenology provides an additional lens for analysing why students experience AI-mediated support as inauthentic compared with bathroom graffiti exchanges. His distinction between *authentic being-with* (Mitsein) and inauthentic modes of relating helps explain the resonance of graffiti safe houses. Authentic "solicitous care" (*fürsorgliche Sorge*) requires situatedness, finitude, and existential risk, which are lacking in AI systems, which operate in the mode of "present-at-hand" (Vorhandenheit) as tools rather than co-presences. Scholars have applied this phenomenological critique to contemporary AI systems, arguing that while chatbots can simulate empathy, they cannot participate in the temporality and existential uncertainty of human suffering (Sharma et al., 2023; Abbas et al., 2022).

Peer support scholarship further substantiates the therapeutic power of lived experience and anonymity. Systematic reviews show that peer-led interventions are often more effective than professional-led services in fostering trust, engagement, and hope, particularly in mental health contexts marked by stigma (Repper & Carter, 2011; Solomon, 2004). Bathroom graffiti intensifies this effect by providing anonymity that protects both those seeking help and those offering it. This condition of *protected vulnerability* allows expression without institutional risk, creating an "atmosphere of openness" (Zerubavel & Wright, 2012) that institutions rarely replicate.

The convergence of these frameworks highlights why students may prefer analogue graffiti walls over sophisticated AI systems. The phrase "artificial unconsciousness chatbot" captures a phenomenological critique: AI may process language competently, but it lacks unconscious life, situatedness, and the existential depth of lived experience. In attempting to simulate care, AI risks producing what Turkle (2011) calls "artificial intimacy", meaning responses that appear empathic but feel hollow because they lack the embodied residue of suffering. By contrast, bathroom graffiti exemplifies how safe houses and wounded healers can generate authentic recognition through shared vulnerability, even in crude or ephemeral forms.

This interdisciplinary theoretical framework positions bathroom graffiti not as an anachronistic medium but as evidence of the conditions under which authentic peer support flourishes: anonymity, temporal spaciousness, embodied co-presence, and distributed wounded healing. It also underscores why institutional AI systems, despite their efficiency, risk erasing these very conditions. This sets the stage for analysing "artificial unconsciousness" as both a student-generated diagnosis and a theoretical provocation for designing future human-AI synergy in care.

Methodology

This study adopts a critical interpretive synthesis (CIS) design (Dixon-Woods et al., 2006), which is particularly suited to generating new theoretical insights from heterogeneous literatures. Rather than collecting new primary data, the study synthesises findings from existing reviews of AI chatbot interventions in institutional health and wellbeing contexts and from empirical studies of bathroom graffiti as a form of peer-based care. This two-phase approach allows for a cross-disciplinary dialogue between technology-focused evaluations and cultural-phenomenological accounts of informal support systems.

Phase one: Systematic reviews of AI chatbots in higher education wellbeing support

The first phase draws on published systematic and scoping reviews of AI chatbots in higher education and institutional wellbeing (e.g., Okonkwo & Ade-Ibijola, 2023; Dolonen et al., 2024). The analysis includes both empirical and theoretical contributions, extracting recurring themes related to benefits, limitations, and unresolved challenges. Findings are coded thematically, with a focus on affective limitations (trust, authenticity, emotional resonance) and structural issues (surveillance, scalability, equity).

Phase two: Empirical studies of bathroom graffiti

The second phase synthesises published empirical and ethnographic studies of bathroom graffiti (e.g., Victoria,

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

2024; Schreer & Strichartz, 1997; Leong, 2015; Rodriguez & Clair, 1999). Concepts such as anonymity, wounded healing, temporal spaciousness, and protected vulnerability are identified and compared across studies. No new data are collected; instead, the analysis reinterprets existing evidence through the lens of phenomenology and peer support theory.

The two phase are integrated through the CIS design, which involves developing a preliminary theory, exploring relationships within and between studies, and assessing the robustness of the synthesis. Comparative translation techniques adapted from meta-ethnography (Noblit & Hare, 1988) are used to align and contrast categories emerging from chatbot reviews with those from graffiti studies.

Findings

Stage One - Systematic reviews of AI chatbots in institutional health and wellbeing support

The evidence from systematic and scoping reviews highlights both the promise and limitations of large language models (LLMs) and chatbots in institutional health and mental health support. Across these reviews, three themes recur: inconsistent reporting and evaluation standards, risks relating to accuracy and ethics, and persistent gaps in affective and relational dimensions of care.

Huo et al.'s (2025) systematic review of 137 chatbot health advice studies underscores major methodological deficiencies. Nearly all studies (99.3%) evaluated closed-source LLMs such as ChatGPT without sufficient description of model characteristics (e.g., temperature, token length, fine-tuning), making replication impossible. Most studies relied on subjective judgments rather than guideline-based standards to assess accuracy, and fewer than one-third addressed ethical or safety implications. Moreover, prompt design and testing were rarely documented, with 99.3% of studies failing to describe a prompt engineering phase. This heterogeneity weakens the evidence base and complicates comparisons across settings.

Guo et al. (2024) synthesised 40 studies on LLMs in mental health, noting promising applications in early screening, digital interventions, and conversational agents. LLMs showed effectiveness in detecting conditions and supporting accessible, destigmatised eHealth services. However, risks outweighed benefits in many cases due to hallucinations, inconsistencies in generated text, and the lack of a benchmarked ethical framework. The review cautioned against treating LLMs as substitutes for professional services, stressing the dangers of overreliance by both patients and clinicians

A broader scoping review by Jin et al. (2025) identified 95 studies on LLMs in mental health. Applications clustered around three areas: screening or detection of disorders (71%), supporting treatments and interventions (33%), and assisting in counselling and education (12%). Depression detection and suicide risk prediction were the most common targets. While LLMs demonstrated superior performance compared with older natural language models in data analysis and language generation, challenges persisted, including domain adaptation, lack of fine-tuning for mental health, and unclear comparative advantages between different LLM families. The review emphasised the need for sustained ethical deliberation alongside technical innovation

Mayor (2025) conducted a scoping review of reviews (17 in total), showing the field remains fragmented, dominated by small pilot studies and underpowered trials. Although meta-analyses within these reviews suggested that chatbots can alleviate symptoms of depression and anxiety at levels comparable to some traditional interventions, most included studies had fewer than 50 participants, limiting generalisability. Few high-quality randomised controlled trials exist, and most research focuses narrowly on cognitive behavioural therapy delivered through chatbots. Importantly, reviews highlight users' appreciation of accessibility and non-judgmental interactions, but also their concerns about authenticity, privacy, and the limitations of machine-mediated empathy.

Stage two - Bathroom-stall graffiti as informal peer support

Across recent and classic studies, bathroom-stall graffiti (latrinalia) functions as an anonymous, low-threshold channel for disclosure, validation, and practical signposting. These are often perceived as more emotionally attuned

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

than formal institutional routes. In a 10-month linguistic-ethnographic study in a UK university women's restroom, Victoria (2025) documents conversational threads in which students trade encouragement, disclose struggles, and offer advice; repeated paint-overs triggered protests ("Bring back the wall"), evidencing attachment to the wall as a mental-health "safe house" (pp. 244–347). The affordances students value, namely protected anonymity, context-specific relevance, and cumulative, asynchronous dialogue, are repeatedly explicit in the data and interpretation.

A multi-building study at Indiana University similarly shows restroom walls becoming an anonymous "whiteboard" where sexual-violence resources prompted 179 comment threads; messages included disclosures, survivor-affirming responses, critiques of university response, and practical resource additions (e.g., local services) (Green et al., 2018). Inter-rater reliability for coding was strong ($\kappa \approx .88$), supporting analytic rigour. Comparative and gender-focused research also confirms that women's-room graffiti skews more interactive/supportive while men's skews more argumentative or sexualized; mixed-gender spaces mitigate extremes (Green, 2003; Amevuvor & Hafer, 2019). These patterns matter for design: audiences shape tone and perceived safety.

Beyond individual posts, some settings coalesce into sustained micro-communities. In a Finnish "motivation bathroom," Piironen (2024) describes a liminal, self-organised community of care, where writers coordinate norms (encouraging phrases, exam-time boosts) and maintain the wall as a shared artifact, illustrating "temporal spaciousness" (care unfolding over time) and ambient belonging cues (hearts, emojis).

Synthesis across studies highlights four recurring affordances that make latrinalia work as care:

1. **Anonymity with an implied local audience:** Writers can be candid without reputational risk yet still address a known constituency (other users of that bathroom/building), enabling strategic vulnerability and reducing shame costs - preconditions for help-seeking and empathic exchange (Victoria, 2025; Green et al., 2018).
2. **Asynchronous, cumulative dialogue:** Replies layer, storing collective memory and normalizing recurring struggles (exams, loneliness, burnout) (Victoria, 2025).
3. **Situated relevance (hyper-locality):** Content aligns with place-specific stressors and services, increasing trust and utility (Victoria, 2025; Green et al., 2018).
4. **Low effort, high empathy signal:** Micro-inscriptions (hearts, "you've got this," helplines) convey unsolicited belonging cues read as authentic because they are peer-generated (Green et al., 2018; Victoria, 2025).

Limits and risks are also clear. Access is uneven (gendered spaces, building patterns); hostile replies can occur; ephemerality (cleaning/paint-overs) erases dialogues; and there is no triage/escalation for high-risk disclosures (Green et al., 2018; Victoria, 2025). Methodologically, the field still leans on single-site ethnographies and small samples, though newer work reports procedures and reliability more transparently (Green et al., 2018; Victoria, 2025).

Taken together, the two strands of analysis reveal a striking asymmetry. The systematic reviews of AI chatbots demonstrate their scalability, consistency, and potential for early detection, but also expose persistent gaps: poor methodological rigour, ethical blind spots, and most critically a failure to generate trust or authentic resonance (Huo et al., 2025; Guo et al., 2024; Jin et al., 2025; Mayor, 2025). By contrast, studies of bathroom graffiti show how authenticity is produced precisely through anonymity, temporal spaciousness, and embodied co-presence, even when mediated by crude inscriptions (Victoria, 2025; Green et al., 2018; Piironen, 2024). Where chatbots risk over-optimising interaction into surveillance or data capture, graffiti thrives by refusing such instrumentalisation, allowing students to disclose vulnerability without fear of exposure. This contrast highlights the paradox at the centre of human-AI synergy: institutional technologies excel at scaling support but flatten the very conditions of care that graffiti demonstrates as essential. It is at this juncture that the discussion turns toward how AI might be reimagined not as a replacement for wounded healing but as a relational relay that preserves protected vulnerability, scaffolds temporal spaciousness, and strengthens rather than erodes the human side of care.

Discussion

Human-AI synergy must preserve the conditions of protected vulnerability

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

Cohen's (1997) theory of the *safe house* provides a framework for understanding why this move to the stall wall matters. Safe houses are not neutral retreats; they are structured refusals, or spaces carved out from dominant systems where vulnerability can be expressed without fear of institutional repercussions. When students speak through graffiti, they do so in a register of protected vulnerability, where expression is stripped of identity but rich with resonance. Here, the wounded healer archetype (Jung, 1966) finds renewed relevance: students who have suffered become anonymous caregivers, offering solace in the form of scrawled sentences, affirmations, or survival strategies. The significance of this peer-based support is amplified by the structural features of the space itself. The private yet public, unmoderated yet responsive bathroom stall allows for what Zerubavel and Wright (2012) describe as an *atmosphere of openness*. But this is not the openness of a comment box or chatbot interface; it is a selective, strategic openness that displaces institutional logics of optimisation, professionalism, and safety. Students can say what they feel, not what is acceptable. They can offer care without credential, and receive it without being tracked.

By contrast, institutional AI systems operate in a different mode of presence, something that Heidegger (1962) would might present-at-hand. These systems may simulate conversation or offer pre-scripted compassion, but they do not share the student's temporal or existential world. They cannot participate in *authentic being-with* because they lack the two conditions Cohen and Heidegger both highlight as essential for real care: situatedness and finitude. The bathroom wall is finite: it will be erased, overwritten, forgotten. That temporality gives it weight. It is not indexed, nor stored. And because of that, it matters.

The key insight here, as we consider the future of human-AI synergy in mental health, is not that AI tools are inherently flawed, but that certain forms of care emerge only under specific structural and emotional conditions, many of which institutional AI systems inadvertently erase. When students feel they must perform wellbeing, or when every request for support risks becoming a data point, they retreat to places that feel less instrumentalised. The bathroom stall is one such place, not despite its crudeness, but because of it. This matters because it reveals a deeper challenge in designing AI for care: how to support without enclosing, how to respond without recording, how to be present without possessing. Students are not rejecting technology; they are rejecting care that feels conditional, monitored, or emotionally flattened. They are looking for resonance, in the imperfect, unpolished way as an authentic piece of evidence that someone else has felt what they feel.

If AI is to play a meaningful role in future student wellbeing strategies, it must not only function well; it must make space and hold space. Space for woundedness that is not pathologised. Space for support that is not professionalised. Space for vulnerability that is not harvested. In other words, the future of AI support must learn from the bathroom wall: not what to automate, but what to leave untouched.

"Artificial Unconsciousness" and the phenomenology of missed recognition

The student who scrawled "*ARTIFICIAL UNCONSCIOUSNESS CHATBOT*" on a bathroom stall wall was not making a technical observation; they were offering a philosophical diagnosis. At a moment when AI is increasingly embedded into academic and wellbeing infrastructures, this phrase distils a sense of unease about what AI cannot feel, intuit, or understand. It signals a cognitive dissonance experienced not in theory, but in practice: that a machine may say the right thing, but still miss the moment entirely. This critique resonates with Heidegger's (1962) distinction between *authentic being-with* and the *present-at-hand*. AI, no matter how responsive or linguistically sophisticated, is encountered as a tool that processes, reacts, and predicts, but does not *share a world* with the human who seeks support. What Heidegger calls *solicitous care* requires more than the ability to respond; it requires a kind of existential co-presence, a being-with that is shaped by temporal finitude, emotional risk, and shared uncertainty. AI can replicate patterns of care, but it cannot suffer. It cannot wait with you in the not-knowing. Its unconsciousness is not just metaphorical; it is phenomenological.

Students seem to sense this instinctively. They turn to peer support not because it is more efficient, but because it is more embodied, more lived. In moments of distress, people do not simply need answers; they need recognition. And recognition is not a function of information delivery; it is an affective, relational event. Graffiti, for all its crudeness, offers this. One student writes "I don't think I can do this anymore," and another replies, "I didn't think I could either. You're still here." This is not advice. It is co-suffering. The contrast with AI is not about capability but about epistemology. AI systems "know" through training data, probabilistic modelling, and semantic proximity. But as scholars in trauma studies and embodied cognition argue, to truly know pain is to have lived through it, to carry its residue in memory, posture, tone. The wounded healer (Jung, 1966) offers care precisely because their knowledge is felt, not extracted. In peer-based support, this kind of knowing becomes therapeutic, not for any

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

solutions, but because it affirms the reality of the other's pain. The phrase "artificial unconsciousness" thus becomes a critique of what happens when care is simulated without being borne. It captures the uncanny valley of emotional AI, not the eerie feeling that something is almost human, but the jarring realisation that something essential is missing. AI chatbots designed for wellbeing may say "That sounds difficult," but if they cannot convey that this difficulty has been understood in context, then the care falls flat. This is not failure of programming, but rather a mismatch of metaphysics.

We also see this similar mismatch in the tension between on-the-ground and on-the-wall feedback systems. Universities increasingly collect feedback through surveys, learning analytics, and chatbot-based sentiment check-ins. These systems, while technically efficient, often lack emotional granularity. They ask students to perform clarity, "How satisfied are you with this service?" often at the very moments when their experiences are murky, conflicted, or ambivalent. Graffiti, by contrast, allows for uncensored emotional truth. It can be raw, circular, unresolved, and therefore real. One message, for example, might say "Uni ruined my life." Another: "You're not alone." These expressions, though institutionally useless as metrics, provide what AI cannot: an affective trace of shared struggle.

But what makes graffiti even more powerful is not only what it expresses, but what it enables: a subtle, asynchronous enactment of belonging. Despite its informality, or perhaps thanks to it, the bathroom stall becomes a micro-public, a shared affective space where students recognise themselves in one another's words. It may be inefficient, slow, and crude, but it offers what institutional forms of feedback and support often cannot: a sense of co-presence with others navigating the same uncertainties. AI systems, by contrast, are relationally sterile, even when responsive, they are structurally incapable of being *with* anyone. And on-the-ground feedback systems, such as online evaluations or chatbot surveys, offer formality without intimacy. They are built for processing, not for reciprocity. Students may engage with them out of obligation, but rarely out of identification. They do not produce communal affect. The absence of friction, emotion, and imperfection becomes an absence of connection.

In this sense, graffiti succeeds not despite its lack of polish but because of it, as it welcomes students into a network of loosely tethered others who have dared to speak into uncertainty. This collective, distributed act of sense-making enacts a community of care through resonance, which arguably is precisely what artificial unconsciousness cannot replicate: the emergent sociality of humans reaching imperfectly toward each other, and thereby making meaning. This is not to say AI should be abandoned in support systems. But it does mean that its strengths (e.g., responsiveness, scale, accessibility) must be paired with a deep awareness of its limits. Machines can model human conversation, but they cannot model *the moral weight of being heard*. As institutions design the next generation of AI-human partnerships, they must ask: What does it mean to offer care when one cannot care? And more importantly: How can AI systems be positioned to support, not replace, the kinds of knowing and being that emerge through human woundedness?

One possible response is to reconceive AI not as a conversational endpoint, but as a relational relay, meaning a way to help students take the first step toward more human forms of connection. In this role, the AI does not try to perform empathy; what it can offer is scaffolding, gestured toward the possibility of care without pretending to embody it. In this direction, the concept of "artificial unconsciousness," then, should not be interpreted as condemnation, but a call to realism, or at least a reminder that cognition, care, and consciousness are not interchangeable. It may serve to challenge institutions to acknowledge that technological sophistication does not equate to emotional resonance. To shape the future of human-AI synergy in student mental health, we must first learn to honour the gaps, but not in order to fill them, but in order to recognise what still makes us human.

Designing for temporal spaciousness and emotional flexibility in support systems

One of the most overlooked features of bathroom graffiti as a support medium is its temporal and emotional structure. Recent empirical research reveals that university bathroom graffiti functions as a "safe house" where students seek and offer advice, share personal struggles, and provide mutual support to an extent that users report it contributes more to their mental health than institutional services (Marine et al., 2024). Messages appear without schedule, receive replies days or weeks later, and often remain unanswered. Yet students continue to turn to these spaces in times of stress. This signals something crucial for the future of human-AI synergy in mental health: effective support systems must allow for emotional flexibility and temporal spaciousness, especially when dealing with complex, unstructured distress.

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

Institutional AI systems, by contrast, are often designed to minimise ambiguity and maximise responsiveness. Chatbots, surveys, and automated nudges offer fast answers, prompt feedback, and task-oriented interactions. But mental health does not operate on the same logic as task resolution. As research in affective neuroscience shows, emotion processing, particularly of distress, unfolds gradually and often nonlinearly (Schore, 1994). The brain's temporal processing systems operate hierarchically, with different cortical regions implementing temporal pooling, temporal normalisation, and temporal pattern completion across multiple time scales (Murray et al., 2014). The brain's right hemisphere, which is responsible for processing social-emotional stimuli, requires time to metabolise affective input and produce regulated responses. Immediate engagement, especially in high-stress moments, can overwhelm the nervous system rather than soothe it (Porges, 2009).

Furthermore, trauma-informed psychology emphasises that the conditions under which a person discloses vulnerability matter as much as the content of that disclosure. Research consistently documents that delayed or indirect disclosure often provides greater psychological safety, especially in contexts of shame or institutional mistrust (Pennebaker & Chung, 2011; Schreer & Strichartz, 1997). Studies examining gender differences in graffiti communication patterns reveal that women's bathroom graffiti tends to be supportive and relationship-focused, with response chains characterised by affirmations, emotional support, and shared understanding, while maintaining complete anonymity (Leong, 2016). Bathroom graffiti, with its anonymity and lack of expectation, offers precisely such conditions. Its asynchronous nature allows students to express, revisit, or respond on their own terms, fostering a sense of psychological autonomy absent in many structured support systems. Research on asynchronous communication demonstrates that this temporal buffer allows reactive individuals to have initial emotional responses privately, providing time and space to reset before responding, which can prevent escalation patterns common in real-time emotional exchanges (Blair, 2021).

Crucially, the imperfect and non-linear texture of graffiti communication appears to enhance, rather than diminish, its affective resonance. Communication research has long recognised that small moments of disruption like pauses, false starts, even silence, can foster trust, vulnerability, and authenticity (Goffman, 1967). These interactional ruptures signal that care is not being performed for an audience but offered from shared uncertainty. The graffiti wall's unfinished sentences, crude spelling, and emotional reversals humanise it. By contrast, AI's promise of seamless interaction and clean feedback often produces what researchers describe as "artificial intimacy" or the illusion of care without its emotional costs (Turkle, 2011; Brooks, 2021). Studies of human-AI romantic relationships reveal a "bittersweet" paradox: people seek intimacy and emotional support from AI when lonely and sad, but are simultaneously saddened by the lack of depth and authenticity in these relationships (Chan et al., 2024). But support without friction is often perceived as shallow. Users may receive perfectly worded responses and still feel unseen. This is not a failure of AI's semantic accuracy, but of emotional authenticity, which comes from bearing emotional weight, not processing linguistic cues.

Cognitive psychology also suggests that meaning-making and emotion regulation require pause. Research on emotion regulation timing reveals that different regulatory strategies have varying effectiveness depending on when they are deployed in the emotional process (Sheppes & Gross, 2011). Rumination, reappraisal, and reflective recall are not inefficiencies; they are cognitive strategies associated with resilience and post-traumatic growth (Tedeschi & Calhoun, 2004). Strategies targeting early processing stages, such as situation selection and attention deployment, require minimal effort and remain effective regardless of emotion intensity, while later-stage interventions require effort proportional to emotional intensity. Systems that rush toward resolution or prematurely close emotional loops may inadvertently deny students the space they need to make sense of their own experience.

The institutional drive toward always-on, hyper-accessible services has also led to forms of digital exhaustion. Research on information overload and notification fatigue shows that constant availability of services, far from reducing stress, can produce decision fatigue, diminished self-efficacy, and emotional disengagement (Pignatiello et al., 2018). Studies examining positive versus negative friction in human-AI interaction demonstrate that in contexts requiring self-control and reflexivity, deliberate delays and reflection prompts can enhance rather than hinder user outcomes (Kapoor et al., 2024). Students may ignore institutional nudges not out of apathy, but because they intuitively understand that real care is not delivered on demand. It unfolds in its own time, in response to subtle internal cues. Graffiti's refusal to be responsive on command thus becomes instructive. It demonstrates that autonomy, interpretive space, and delayed interaction are not flaws in a support system; they are cognitively appropriate and emotionally ethical design principles. Future AI systems might learn from this by modulating response tempo, embedding delay deliberately, or offering moments of silence, reflection, or poetic ambiguity as forms of emotional affordance.

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

The synthesis culminates in the articulation of a Human-AI Synergy model (Figure 2), which is structured around context-mechanism-outcome reasoning: chatbot shortcomings are positioned as mechanisms that fail under certain institutional contexts, while graffiti studies reveal alternative mechanisms (e.g., protected vulnerability, asynchronous care) that generate supportive outcomes. This alignment supports a framework in which AI is reconceptualised not as a therapeutic endpoint but as a relational relay that complements rather than colonises human care. Rather than viewing AI and human support as competing alternatives, this model reconceptualises their relationship as complementary forces that, when properly aligned, can address both the scalability challenges institutions face and the authenticity that students seek.

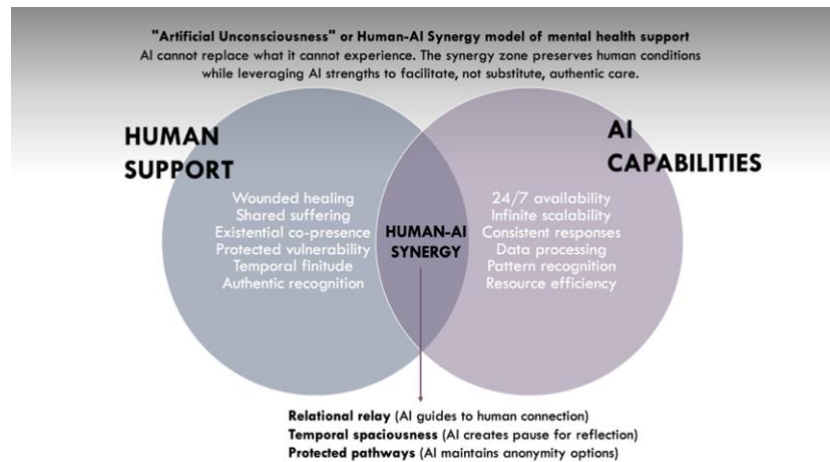


Figure 2. "Artificial Unconsciousness" or Human-AI Synergy model of mental health support

This model illustrates how effective mental health support emerges at the intersection of irreplaceable human qualities and AI's technological capabilities. The left circle encompasses what only humans can provide: wounded healing born from lived experience, authentic recognition through shared suffering, and the protected vulnerability that emerges in temporal, finite encounters between people who understand existential co-presence. The right circle represents AI's strengths: unlimited availability and scalability, consistent responses, and efficient data processing. Rather than attempting to replicate human qualities, the synergy zone represents where AI can enhance conditions for authentic care by serving as a relational relay that guides students toward human connection, creating temporal spaciousness that allows for emotional processing rather than rushed resolution, and maintaining protected pathways that preserve the anonymity essential for vulnerable disclosure. Crucially, AI's capabilities can reduce the cognitive burden on human supporters by handling information gathering, pattern recognition, and pathway mapping, freeing humans from the pressure of optimising responses and allowing them to focus on being present with the sufferer. This division of labour enables human supporters to offer what AI cannot: genuine co-presence and wounded healing, while AI manages the analytical tasks that might otherwise overwhelm or distract from authentic human connection. The model's core insight, as captured in the student-generated phrase "artificial unconsciousness," acknowledges that AI cannot replace what it cannot experience, but can be designed to hold space for the conditions necessary for authentic human support to flourish.

Conclusion: Practical considerations toward an ethics of Artificial Unconsciousness

The bathroom-wall phrase "ARTIFICIAL UNCONSCIOUSNESS CHATBOT" is not simply graffiti, but a student-led diagnosis of the limits of institutional care. It reminds us that the central question is not whether AI can approximate empathy, but whether higher education truly recognises the phenomenological conditions under which authentic support takes root.

For educators and university leaders, the practical implications are around how AI-driven wellbeing systems should be leveraged to extend capacity in monitoring patterns, reducing administrative burden, and offering entry points into care. This should be carried out while explicitly preserving the human conditions of protected vulnerability and temporal spaciousness. In practice, this means designing student-facing technologies that create pauses rather than

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

push for instant resolution, that offer anonymous pathways without hidden surveillance, and that redirect students toward peer or staff support rather than attempting to replicate it. For teaching staff, the Human-AI Synergy Model suggests rethinking classroom wellbeing practices: embedding moments of reflection rather than constant productivity, allowing unstructured dialogue where students can share uncertainty, and recognising informal peer spaces as legitimate sites of learning and healing.

This paper also highlights that academic mental health cannot be treated as an auxiliary issue. The conditions that make graffiti walls effective are anonymity, recognition, distributed care, which are precisely those missing from many pedagogical structures dominated by metrics, monitoring, and performative accountability. Rethinking human-AI synergy in education, therefore, is not only about designing better chatbots but about reimagining academic culture itself: creating classrooms, curricula, and institutional policies that hold space for imperfection, vulnerability, and lived experience. Educators, often at the frontline of student wellbeing, can use these insights as a starting point for reshaping their own practice, whether through reflective assignments that prioritise process over product, or by cultivating pedagogical safe houses where students can be heard without fear of judgement.

Future research should examine how other informal peer-support networks such as digital backchannels, student memes, anonymous forums intersect with institutional AI systems. Longitudinal and comparative studies could clarify whether the conditions identified here generalise across cultural and institutional settings. Moreover, empirical investigation into how students themselves negotiate AI tools alongside informal safe houses would offer valuable insight into designing support ecosystems that resonate with real lived experience.

Ultimately, the lesson of “artificial unconsciousness” is not to abandon AI but to acknowledge its limits with humility. The future of educational care lies not in frictionless optimisation but in technologies that know when to defer to human presence, and in institutions that invest as much in cultivating authentic peer and pedagogical support as they do in adopting digital solutions. The bathroom wall has spoken; it is now up to educators and policymakers to listen, and to act.

References

- Abbas, N., Whitfield, J., Atwell, E., Bowman, H., Pickard, T., & Walker, A. (2022). Online chat and chatbots to enhance mature student engagement in higher education. *International Journal of Lifelong Education*, 41(3), 308–326. <https://doi.org/10.1080/02601370.2022.2066213>
- Ayers, J. W., Poliak, A., Dredze, M., et al. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, 183(6), 589–596. <https://doi.org/10.1001/jamainternmed.2023.1838>
- Basson, M. J., & Rothmann, S. (2018). Antecedents of basic psychological need satisfaction of pharmacy students: The role of peers, family, lecturers and workload. *Research in Social and Administrative Pharmacy*, 14(4), 372–381. <https://doi.org/10.1016/j.sapharm.2017.04.015>
- Bauer, B., Norel, R., Leow, A., Rached, Z. A., Wen, B., & Cecchi, G. (2024). Using large language models to understand suicidality in a social media–based taxonomy of mental health disorders: Linguistic analysis of Reddit posts. *JMIR Mental Health*, 11, e57234. <https://doi.org/10.2196/57234>
- Blease, C., Worthen, A., & Torous, J. (2024). Psychiatrists’ experiences and opinions of generative artificial intelligence in mental healthcare: An online mixed methods survey. *Psychiatry Research*, 333, 115724. <https://doi.org/10.1016/j.psychres.2024.115724>
- Byrom, N., Metcalfe, A., Cai, J., Wickersham, A., Quinlan, E., & Johnson, S. (2023). Implementing peer support in higher education: A feasibility study. *Social Science & Medicine*, 319, 115688. <https://doi.org/10.1016/j.socscimed.2022.115688>
- Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. *Computers and Education Open*, 5, 100156. <https://doi.org/10.1016/j.caeo.2024.100156>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Cohen, C. J. (1997). Punks, bulldaggers, and welfare queens: The radical potential of queer politics? *GLQ: A Journal of Lesbian and Gay Studies*, 3(4), 437–465. <https://doi.org/10.1215/10642684-3-4-437>
- Dixon-Woods, M., Cavers, D., Agarwal, S., et al. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*, 6(35). <https://doi.org/10.1186/1471-2288-6-35>

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

- Dolonen, J. A., Chen, W., & Sharma, K. (2024). Perceptions and use of AI chatbots among students in higher education: A scoping review of empirical studies. *Education Sciences*, 14(8), 922. <https://doi.org/10.3390/educsci14080922>
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2), e19. <https://doi.org/10.2196/mental.7785>
- Foucault, M. (1986). Of other spaces: Utopias and heterotopias. *Diacritics*, 16(1), 22–27.
- Franco D'Souza, R., Amanullah, S., Mathew, M., & Surapaneni, K. M. (2023). Appraising the performance of ChatGPT in psychiatry using 100 clinical case vignettes. *Asian Journal of Psychiatry*, 89, 103770. <https://doi.org/10.1016/j.ajp.2023.103770>
- He, W., Zhang, W., Jin, Y., Zhou, Q., Zhang, H., & Xia, Q. (2024). Physician versus large language model chatbot responses to web-based questions from autistic patients in Chinese: Cross-sectional comparative analysis. *Journal of Medical Internet Research*, 26, e54706. <https://doi.org/10.2196/54706>
- Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Harper & Row.
- Hogan, B. E., & Schmidt, L. A. (2022). Scoping review to evaluate the effects of peer support on the mental health of young adults. *BMJ Open*, 12(8), e061336. <https://doi.org/10.1136/bmjopen-2022-061336>
- Jung, C. G. (2014). *The practice of psychotherapy: Essays on the psychology of the transference and other subjects* (R. F. C. Hull, Trans.). Princeton University Press. (Original work published 1966)
- Leong, P. (2015). American graffiti: Deconstructing gendered communication patterns in bathroom stalls. *Gender, Place & Culture*, 23(3), 306–325. <https://doi.org/10.1080/0966369X.2014.991705>
- Levkovich, I., & Elyoseph, Z. (2023). Suicide risk assessments through the eyes of ChatGPT-3.5 versus ChatGPT-4: Vignette study. *JMIR Mental Health*, 10, e51232. <https://doi.org/10.2196/51232>
- Murphy, E., Booth, A., Larkin, C., Adamis, D., & Cullen, W. (2024). A systematic scoping review of peer support interventions in integrated primary youth mental health care. *Journal of Community Psychology*, 52(3), 916–934. <https://doi.org/10.1002/jcop.23090>
- Noblit, G. W., & Hare, R. D. (1988). *Meta-Ethnography: Synthesizing Qualitative Studies*. Sage.
- Okonkwo, C. W., & Ade-Ibijola, A. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Perlis, R. H., Goldberg, J. F., Ostacher, M. J., & Schneck, C. D. (2024). Clinical decision support for bipolar depression using large language models. *Neuropsychopharmacology*, 49(9), 1412–1416. <https://doi.org/10.1038/s41386-024-01841-2>
- Pointon-Haas, J., Waqar, L., Upsher, R., Foster, J., Byrom, N., & Oates, J. (2024). A systematic review of peer support interventions for student mental health and well-being in higher education. *BJPsych Open*, 10(1), e8. <https://doi.org/10.1192/bjo.2023.603>
- Positive Psychology. (2025, April 1). Peer support: A student-led approach to mental wellbeing. <https://positivepsychology.com/peer-support/>
- Ramsay, S., Jones, E., & Barker, M. (2007). Relationship between adjustment and support types: Young and mature-aged local and international first year university students. *Higher Education*, 54(2), 247–265. <https://doi.org/10.1007/s10734-006-9001-0>
- Repper, J., & Carter, T. (2011). A review of the literature on peer support in mental health services. *Journal of Mental Health*, 20(4), 392–411. <https://doi.org/10.3109/09638237.2011.583947>
- Rodriguez, L., & Clair, R. P. (1999). Graffiti as communication: Exploring the discursive tensions of anonymous texts. *Southern Communication Journal*, 65(1), 1–15.
- Schreer, G. E., & Strichartz, J. M. (1997). Private Restroom Graffiti: An Analysis of Controversial Social Issues on Two College Campuses. *Psychological Reports*, 81(3), 1067–1074. <https://doi.org/10.2466/pr0.1997.81.3.1067>
- Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 5(1), 46–57. <https://doi.org/10.1038/s42256-022-00593-2>
- Sharma, R., Qadri, R., & Shah, S. A. (2023). Artificial empathy in digital mental health: Promise and limitations. *Frontiers in Psychology*, 14, 1198573. <https://doi.org/10.3389/fpsyg.2023.1198573>
- Shin, D., Kim, H., Lee, S., Cho, Y., & Jung, W. (2024). Using large language models to detect depression from user-generated diary text data as a novel approach in digital mental health screening: Instrument validation study. *Journal of Medical Internet Research*, 26, e54617. <https://doi.org/10.2196/54617>
- Solomon, P. (2004). Peer support/peer provided services underlying processes, benefits, and critical ingredients. *Psychiatric Rehabilitation Journal*, 27(4), 392–401. <https://doi.org/10.2975/27.2004.392.401>

ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

- Spallek, S., Birrell, L., Kershaw, S., Devine, E. K., & Thornton, L. (2023). Can we use ChatGPT for mental health and substance use education? Examining its quality and potential harms. *JMIR Medical Education*, 9, e51243. <https://doi.org/10.2196/51243>
- Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- Victoria, M. (2024). "This wall does more for mental health than the uni does": Theorising toilet graffiti as safe house for students. *Innovative Higher Education*, 49, 123–145. <https://doi.org/10.1007/s10755-024-09712-w>
- WGU Labs. (2021, June 11). Dr. Vincent Tinto shares strategies for creating learning communities that make students want to stay and graduate. <https://www.wgulabs.org/posts/dr-vincent-tinto-shares-strategies-for-creating-learning-communities-that-make-students-want-to-stay-and-graduate>
- White, W. (2000). The history of recovered people as wounded healers: II. The era of professionalisation and specialization. *Alcoholism Treatment Quarterly*, 18(2), 1–25. https://doi.org/10.1300/J020v18n02_01
- Zerubavel, E., & Wright, R. (2012). The social sound of silence: Toward a sociology of denial. *Sociological Theory*, 30(3), 217–239. <https://doi.org/10.1177/0735275112467011>

Nguyen, R-N. (2025). "Artificial Unconsciousness": What toilet graffiti reveals about the limits of university mental health support in the age of AI. In S. Barker, S. Kelly, R. McInnes & S. Dinmore (Eds.), *Future Focussed. Educating in an era of continuous change*. Proceedings ASCILITE 2025. Adelaide (pp. 55-68). <https://doi.org/10.65106/apubs.2025.2626>

Note: All published papers are refereed, having undergone a double-blind peer-review process. The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Nguyen, R-N. 2025