CriSTaL
CRITICAL STUDIES IN TEACHING & LEARNING

2025, Volume 13, Special Issue 2
DOI: 10.14426/cristal.v13iSI2.2506

Towards pedagogies of distrust: Higher education learning in the age of generative artificial intelligence

Dalene Joubert# and Albert Strever
Stellenbosch University, South Africa
#Corresponding Author: dvermeulen@sun.ac.za

Dalene Joubert | Albert Strever

(Submitted: 17 October 2024; Accepted: 12 August 2025)

Abstract

Since the release of ChatGPT, generative artificial intelligence (AI) tools have become ubiquitous in higher education teaching-learning-assessment (TLA). This paper explores how generative AI impacts trust relationships within the TLA context between students and lecturers in relation to AI technologies. Framed by Rolfe, Freshwater and Jasper's (2001) reflective model of *what, so what, now what*, we draw on practical experiences to demonstrate how an integrated model of AI literacies can enhance student engagement and foster critical interaction with generative AI, ultimately cultivating a criticality toward traditional ways of knowledge construction in the classroom. Instead of fostering unquestioning trust, we propose a *pedagogy of distrust* – an environment of healthy scepticism where students (and lecturers) critically interrogate both generative AI and human contributions to knowledge creation. This approach encourages reflective learning, deeper engagement, and the development of lifelong learning skills. It urges lecturers to embrace evolving roles, shifting from sole knowledge sources to facilitators who enable students to navigate complex learning environments independently.

Keywords: AI literacies; generative AI; higher education pedagogy; trust in education; critical thinking

Introduction

In a reflective piece, Eaton asserts that 'trust is not a goal, but something earned through continuous practices of living, leading, and learning' (2024a). As humans, we seem to inherently trust others, a tendency that remarkably seems to extend to our interactions with technological artefacts (Hancock, et al., 2023). Eaton's viewpoint, however, compels us to re-examine the concept of trust within higher education, particularly in the current age of widely accessible generative artificial intelligence (AI) systems.

Since the release of ChatGPT in November 2022, generative AI tools have become embedded in higher education teaching-learning-assessment[1] (TLA). ChatGPT, along with tools like Claude, Gemini, Copilot, and Pi, are large language models designed to analyse massive datasets and generate sophisticated, seemingly intelligent text (Van Dis, et al., 2023). ChatGPT employs neural networks and the combination of big data, computing power, and algorithms to produce these human-like responses (Yu, 2023). It can generate text in multiple languages, aside from English, including South African languages like Afrikaans (OpenAI, 2024a).

The integration of generative AI into higher education introduces significant disruptions to traditional TLA practices, asking us to reconsider the relationships within these learning environments. This necessitates a re-evaluation of existing dynamics between role players, urging an expansion beyond the traditional student-lecturer-content triad (Fraser, 2006) to acknowledge and potentially include generative AI technologies (Joubert & Strever, 2023). The inclusion of generative AI also necessitates a reconsideration of the structure and flow of our learning opportunities, echoing the earlier identified imperative to re-examine trust within these evolving pedagogical relationships.

Our approach, much like Jacobs' (2005) academic literacies project, is grounded in a collaborative journey between an academic developer and lecturer navigating the uncertainty of generative AI's role in higher education. We combined TLA knowledge, disciplinary expertise, and practical experience to explore generative AI's impact on student learning and trust relationships. Our so-called 'insider-outsider' dynamic reflects our shared position as both practitioners and researchers, bringing together diverse perspectives to understand how generative AI reshapes educational environments. Following Rolfe, et al.'s (2001) reflective model, we structured our work around the guiding questions *what, so what,* and *now what*. This framework allows us to make meaning of our practical experiences by first describing the situation, then making sense of it through reflection and theory, and finally considering ways forward for higher education learning in the age of generative AI.

Through this paper, we aim to provide insights into how AI literacies can be integrated into higher education in ways that foster critical engagement, building towards a proposed pedagogy of distrust – an environment of healthy scepticism where students (and lecturers) critically interrogate both generative AI- and human-generated knowledge creation.

Trust and higher education in the age of generative AI

Conceptualising trust

Although trust is considered to be one of the binding forces of society (Hancock, et al., 2023; Hoy & Tschannen-Moran, 1999), it is a challenging term to research because of its lexical and semantic relatedness to terms such as 'rely upon', 'co-operate', 'trustworthiness', 'distrust', and

[1] We agree with Ashwin (2012) and Dann (2014) who argue that TLA are interconnected processes, requiring input and interaction from both teachers and students, debunking the notion that teaching is solely for teachers and learning only for students.

'confidence' (Gerlich, 2024; Robbins, 2016). Tavani and Zimmer (2025) indicate that, philosophically, trust between humans has been viewed as an 'attitude' (following the work of Baier [1986]) in 'relational terms' (Robbins, 2016) or as a general 'expectation' (Walker, 2006). What exactly trust is, and how it is perceived in the higher education space and the academic sense that we are referring to in this article, varies across disciplines and on a societal level between contexts and cultures (Kumar, et al., 2023).

In an educational context, trust mainly seems to be viewed in relational terms (Okoye, 2023; Zhou, 2023). Hoy and Tschannen-Moran's (1999: 189) widely cited definition of trust within education offers a useful starting point – namely, an individual's 'willingness to be vulnerable to another party based on confidence in that party's benevolence, reliability, competence, honesty, and openness'. This definition not only highlights relationality, but it also supposes a specific relationship of trust and vulnerability between various parties. A trust relationship such as this involves two key roles: the *trustor*, the party who places their trust in another, and the *trustee*, the party whom the trustor trusts (Mayer, et al., 1995). Hoy and Tschannen-Moran's (1999) work underscores the intricate and multidimensional nature of trust within (primary or elementary) education, including trust between faculty members, leaders, and external stakeholders like parents. Their findings suggest that trust between various stakeholders is interrelated, with higher trust between, for example, faculty members and their leaders correlating with higher trust in other relationships. This has led to their conceptualisation of trust as 'contagious', wherein trust in one domain spreads to others – and distrust can have a broadly negative effect. Although their research was conducted within the primary education setting, their insights into the interrelatedness of trust relationships are a valuable addition when considering similar dynamics in higher education. Trust in higher education settings often involves multiple stakeholders, such as students, lecturers, administrators, and external/industry partners, mirroring some of the relationships researched by Hoy and Tschannen-Moran (1999).

To deepen our understanding of trust in the context of higher education, it is to some extent useful to draw on research from organisational studies, where trust has been thoroughly conceptualised.[2] Mayer et al. (1995) outlined the foundational characteristics of trust relationships within organisations; more recently, Hancock, et al. (2023) expanded on this work by examining the characteristics of the trustor and trustee, and the influence of contextual factors. This broader framework provides valuable insights into how trust operates in institutions, such as higher education spaces, where complex dynamics are often at play. Hancock, et al. (2023) categorised the factors influencing trust into three main areas:

- **Trustor factors:** The characteristics of the individual who trusts, including the propensity to trust and the perception of the risks involved in trusting.
- **Trustee factors:** The characteristics of the individual in whom trust is placed, including

[2] We do take note that the uncritical application of definitions of trust from various fields has also been criticised (see Zhou, 2023).

perceived trustworthiness, integrity, and ability.

- **Contextual factors:** Situational elements that influence the interaction between the trustor and trustee.

Recognising that trust operates on and across multiple levels within an organisation, Hancock, et al. (2023) also categorised the directionality of trust, labelling 'downwards trust' as the trust a supervisor or manager places in subordinates; while 'upwards trust' refers to the trust that a subordinate bestows on their supervisors or managers. We consider the identified trustor-trustee characteristics in both downward and upward trust relationships relevant to interactions found in the TLA environment within higher education.

Zhou (2023), however, notes that, although relational, trust in higher education settings could make more sense when viewed in a so-called 'network model', where relationships are not defined so strictly. Drawing on the work of Walker (2006), Zhou (2023: 7) argues that trust in higher education does not necessarily rely on a 'relation between two individuals (who already know each other) but exists as a potentiality between individuals insofar as they share and interact in a common social environment, institution or space'. With the network model Zhou (2023: 1;7) shows that trust can be found in 'social environments, institutions or spaces' where individuals can rely on others 'to behave in acceptable ways', leaving room for trust to form between groups of people (not just individuals), or even between humans and non-human entities.

Trust within the TLA context

When applied to higher education, the network model of trust suggests that students' trust is not based solely on their interactions with lecturers, but rather on a shared connection to broader networks (such as their educational institution or the broader education system at large). Students may, for example, already trust their university due to its established reputation, which in turn leads them to trust the lecturers associated with it.[3] Within this model, where trust is diffused across these interconnected relationships, Zhou (2023) perhaps unintentionally relieves lecturers of the pressures of maintaining the high relational expectations students seem to foster and expect, according to Okoye (2023), such as knowing students' names or connecting with them personally.

If we zoom in on the higher education classroom, and the TLA environment specifically, various types and levels of relationships coexist, each influencing students' learning differently. The traditional learning environment can be understood through the triad of lecturer, content, and student (Fraser, 2006), forming an intricate network and symbiosis within this diffused model of trust. These three basic interactions, however, do not account for other interactions and relationships that shape students and their learning within higher education.

[3] There is a potential linkage here with Lukyanenko, et al.'s (2022) foundational trust framework. Due to the limited scope of this article, we have decided not to expand on this system's approach to trust, but we do believe that it can be researched within a larger project on higher education, trust, technology, and generative AI.

The social constructivist approach to student learning (Bates, 2022; Liu & Matthews, 2005; Stellenbosch University, 2022; Vygotsky, 1978) is widely acknowledged and utilised in South African higher education institutions as a meaningful way in which to understand and facilitate student learning. Social constructivists believe that, in addition to the above-mentioned lecturer-student-content triad, knowledge is also co-constructed through social interactions with peers, peer-learning facilitators (such as tutors, TLA assistants, and demonstrators), and the context in which learning takes place (Akpan, et al., 2020). At our institution, we further subscribe to the learning-centred approach (Stellenbosch University, 2025; 2022; Whetten, 2007), which encourages the use of active learning strategies, such as inquiry-based learning, collaborative learning, and problem-based learning, to enhance student engagement and, ultimately, their learning and success. Within this overarching social constructivist approach to student learning, various role players in the TLA environment are meant to contribute to students' learning. The lecturer designs and facilitates learning, while peers and other facilitators can act as more knowledgeable others, guiding the students when they reach an impasse. The contributors in this carefully crafted and curated active learning environment are trustworthy, or at the very least are supposed to be. One could even argue that this kind of TLA environment has been built on a pedagogy of trust (Curzon-Hobson, 2002); it is exactly this foundation of trust that faces disruption with the introduction of easily accessible generative AI tools.

Researchers note that higher education students utilise, dare we say *trust*, generative AI tools such as ChatGPT as an 'autonomous social actor' in the learning environment (Dai, et al., 2023:87), despite these tools not subscribing to the aspects of trust contained, for instance, in Hoy and Tschannen-Moran's (1999) stricter definition of trust, nor even Zhou's more accommodating expectancy that it should 'behave in acceptable ways' (2023: 1). Although the utilisation of generative AI in the TLA environment justifies its inclusion in it, we argue that this new role player brings with it other ways of defining and understanding trust, especially considering the inherent problems around conceptualising its trustworthiness (Ryan, 2020). On a philosophical level, Ryan (2020: 2749) argues that AI cannot be perceived as 'trustworthy' because it does not meet the 'requirements of the affective and normative accounts of trust' – it does not 'possess emotive states' nor can it be responsible for its actions . Therefore, humans can only *rely* on AI – which should not be confused with *trusting* it. This correlates with biological research conducted by Montag, et al. (2023), who discovered that human brains react differently when displaying trust towards humans and AI. In practice, this differentiation does not seem to prevent humans from perceiving AI as trustworthy or utilising it as such. We acknowledge Ryan's (2020) primary concerns, but also try to make sense of the dynamics that come into play when these technologies are used in higher education TLA environments.

Due to its utilisation and consequent inclusion in TLA environments, students may be conditioned to perceive generative AI as yet another 'trusted' role player within this space, despite ethical concerns and its (current) limitations. Keep in mind that general generative AI tools were not necessarily created for educational purposes (Heaven, 2023); they do therefore have different functionality compared to purpose-built learning technologies or educational AI tools, such as

those described in Choi, et al. (2023). Generative AI tools are not reliable knowledge or learning partners, nor are they necessarily more knowledgeable others in the same sense that social constructivists intend (Akpan, et al., 2020; Ding, et al., 2023). Due to how generative AI tools were developed and trained, they can and do hallucinate, they perpetuate biases, and their outputs are not necessarily factually correct – all interactions and outputs with and by these tools should therefore be evaluated critically (Lindgren, 2023; OpenAI, 2024b).

Despite the problems outlined above, generative AI models are in our learning spaces and students are making use of them, which means that we as higher education practitioners should critically consider how to approach their presence in our educational spaces. Perhaps a starting place is considering how to negotiate trust between the various role players in the TLA environment when one of them is completely outside the realm of trustworthiness.

Generative AI and shifting trust dynamics

Researchers have already started investigating the trust relationships in learning spaces that include generative AI. When it comes to the interpersonal trust relationships in the learning environment, Luo (2024) argues that the presence of generative AI here is leading to increasing distrust, especially between lecturers and students. With AI-detection software being used as surveillance tools and students being required to submit AI chat histories along with their submissions of written work, the already strained lecturer-student relationship seems to be fraying.[4] Researching from the student's perspective, Luo (2024) highlights several key factors that shape this relationship. A first finding is that students are fearful of being wrongfully accused of misusing generative AI, leading them to underreport its use (even when it is permitted in assignments), suggesting a level of distrust that students feel toward their lecturers. Students further expect lecturers to be AI-literate, teach these literacies to students, and be transparent about their own use of the technology. Additionally, students feel that lecturers have little trust in them – largely because they don't know the students on a personal level, which is an issue that Okoye (2023) also identifies as important to students' needs.

These changing expectations and issues of trust are not limited to interpersonal relationships between students and lecturers. Ding, et al. (2023) shift the focus to how students report trust in generative AI itself, particularly when used as a virtual tutor. In their study, they explored how first-year physics students interacted with the free version of ChatGPT despite its relatively well-known limitations. They found that students' level of trust in the tool as a tutor fell into three distinct categories: full trust; partial trust; and complete distrust. Interestingly, those who fully trusted ChatGPT were more likely to view it as a machine or robot, while those who trusted it less tended to perceive it as more human-like. This finding contrasts with much of the literature on AI, which suggests that anthropomorphism typically fosters greater trust (Carter, et

[4] Although Luo (2024) attributes the fraying interpersonal relationships to the presence of generative AI, we believe that these relationships were already suffering (Okoye, 2023) due to large classrooms, the massification of higher education, and other factors. Generative AI is simply bringing these existing issues to the forefront (Adendorff, Herman & Joubert, 2025).

al., 2023; Ding, et al., 2023; Ryan, 2020). This may hold further implications for higher education learning environments, portrayed in Figure 1, as it might signal students' increasing trust in technology and distrust of humans.

Students' perceived trust in generative AI models can be linked with Hancock, et al.'s (2023) observation that humans tend to project the same kind of trust onto AI models that we place in other humans, as these technologies often fulfil roles similar to those of people. However, Hancock, et al. (2023) emphasise that the relationship between a user and (broader) AI should resemble a hierarchical one, comparable to the downward trust between a manager and a subordinate. In this dynamic, the characteristics transferred from the human manager-subordinate relationship that signal trust are performance, reliability, and trustworthiness.[5] While they do not explicitly address whether generative AI specifically meets these criteria, we argue that, given the known conceptual and practical limitations discussed earlier, generative AI (currently) falls short of earning trust.

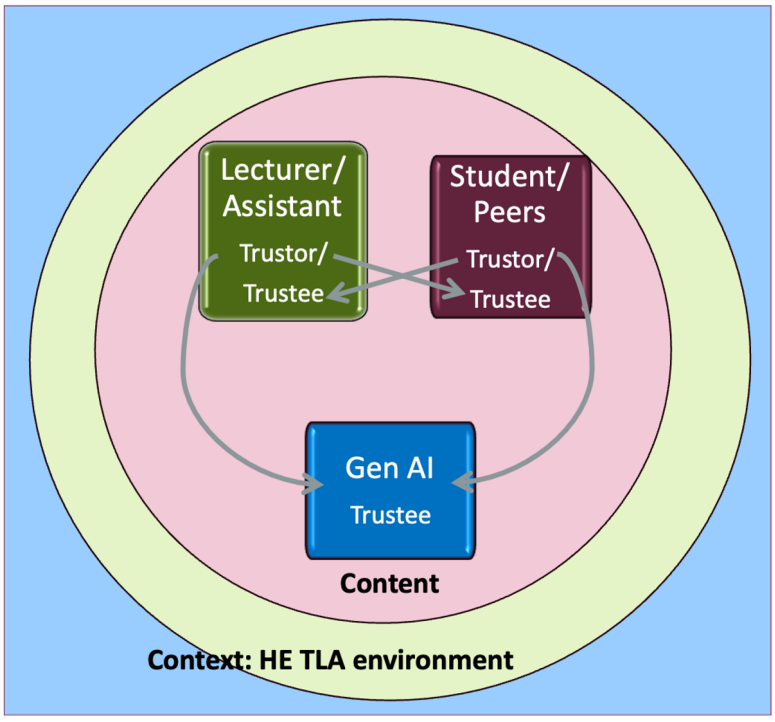


Figure 1: Model of the higher education teaching-learning-assessment environment visualised as role players in a networked model of symbiotic trust, with generative AI being the latest inadvertent addition to the space

Hancock, et al.'s (2023) portrayal of trust as a multifaceted construct shaped by individual traits and the role one assumes (trustor or trustee), alongside situational factors, can be adapted to account for the complexities of multidirectional trust relationships in the TLA environments of

[5] Hancock, et al. (2023) do seemingly not consider Ryan's (2020) argument that AI cannot be conceptualised as a trustworthy entity.

higher education institutions. As shown in Figure 1, we focus on the two most prominent stakeholders in a learning situation: the lecturer and the student. Both are embedded in a network of diffused trust, situated within broader interpersonal and societal systems. Depending on the specific context, they can alternate between being the trustors and/or trustees or even occupy both roles simultaneously within a given interaction. For instance, students (trustors) entrust their educational journey to the institution and its representatives, namely their lecturers, who act as trustees on behalf of the institution (Adendorff, et al., 2025). In an ideal scenario, this is a balanced relationship of mutual trust; however, if we draw on metaphors for higher education where students are positioned as clients and institutions are service providers, this could be perceived as a downward trust relationship, with lecturers and universities catering to the students' needs (see Adendorff, et al., 2025). At the same time, lecturers as trustors place their trust in students, expecting them to take responsibility for their learning, thereby positioning students as trustees. This dynamic creates a situation where trust is multidirectional and context-dependent, flowing both ways in a complex network of expectations and responsibilities.

We argue that when generative AI models enter the TLA environment, it further complicates this multidirectional trust, expanding the roles and responsibilities of both students and lecturers. Ideally, as Hancock, et al. (2023) suggest, both parties should engage with generative AI as supervisors or managers, maintaining a downward trust relationship where the technology serves a supportive, rather than an authoritative, role.

Higher education institutions now face the challenge of navigating the complexities of AI tools in TLA environments, mitigating the potentially negative aspects while leveraging the positive. AI literacies, particularly those specific to TLA in higher education, play a critical role in this process (Adendorff, 2024; Kumar, et al., 2023). Rather than leaving students to navigate AI tools through trial and error, with limited success (Fourie, 2024), we advocate for an integrated approach to teaching with and about these technologies. Aligned with Jacobs' (2005) findings on the integration of academic literacies, we argue that AI literacies are most effectively developed when linked to specific disciplines and incorporated into both the knowledge and skills domains of the academic curricula. In our experience, integrating AI into teaching practices can also present an opportunity to venture away from traditional teaching approaches toward more interactive, engaging, and inquiry-based learning. In the following section, we showcase how integrating AI literacies can, aligning with Hutchinson (2024), encourage students to ask questions, engage in critical thinking, and explore topics more deeply, fostering self-directed learning and problem-solving skills.

An integrated approach to generative AI in TLA

In this section, we consider the *what* (Rolfe, et al. 2001), by describing and reflecting on co-author Dr Albert Strever's innovative integration of generative AI into his TLA, demonstrating the practical impact of AI literacies on academic content and pedagogy. His approach, shared during a webinar (Joubert & Strever, 2023) and at Stellenbosch University's Scholarship of Teaching and Learning Conference in 2023, exemplifies how AI can reshape learning environments and better

equip students to navigate the complexities of generative AI.

Strever's one-semester module, situated within the context of the Department of Viticulture and Oenology in the Faculty of AgriSciences, is designed to guide undergraduate students through the academic research process specific to their discipline. Students typically work in small groups for a semester to research and address real-world industry issues by using scientific sources to produce a research report.

The advent of ChatGPT created a need to support students in responsibly navigating generative AI tools, prompting a shift in the teaching and learning strategy to integrate AI literacies into the existing academic literacies curriculum. This was achieved by turning the existing multi-stage formative assessment into an integrated, experiential learning opportunity, embedding AI literacies within the disciplinary context. In doing so, we draw on three AI literacies principles: selecting appropriate tools for specific tasks with awareness of their affordances, limitations, and ethical dimensions; using them effectively without over-reliance; and critically evaluating outputs for credibility, bias, and value (Stellenbosch University, 2023a). Some of the other objectives of the revised module are to engage students with real-world research problems and encourage them to co-create knowledge in a group setting where peer support fosters collaborative learning and supports engagement with new technologies. Throughout this process, students are encouraged to embrace mistakes as part of their learning, while the lecturer and teaching assistants serve as guides on the side (King, 1993), assisting students through each stage's complexities. Weekly check-ins allowed for progress sharing, critical discussion of AI's limitations and benefits, and feedback throughout the stages of the research project.

The first step is that students self-select groups of about four members who will work together throughout on a project. They are presented with a list of current research topics drawn from industry and choose one that interests them. In the first few contact sessions, students are encouraged to analyse their topics, identify the underlying research problems, and begin formulating research questions. During this phase, students are encouraged to use generative AI tools, specifically large language models of their choice, to help break down complex topics into smaller, manageable components. These tools are also employed as brainstorming partners to assist with developing arguments, in line with institutional AI guidelines[6] (Stellenbosch University, 2023b).

In the information gathering and research phase, students are asked to bring a list of references that they plan to use for the contact session. When students bring hallucinated reference lists generated by large language models, the lecturer and the faculty librarian guide them through proper academic research methods, emphasising the importance of validating sources and searching academic databases. Students are also introduced to AI research tools and made aware of their limitations. By the time students reach the final phase where they are tasked with writing their research reports, they have developed a strong understanding of the

[6] Even though the institutional guidelines had not yet existed during the first iteration of this formative assessment, Strever's classroom experience and contributions were key to its development.

affordances and limitations of various generative AI tools across different phases of the research process. During the writing phase, students are permitted to use large language models to generate feedback on their writing. They may also use language editing tools to improve clarity and grammar, particularly valuable in our multilingual context, where many students write in a language other than their first[7]. The shared understanding was that the final work submitted should represent the student's original thinking and voice, with AI functioning as a supportive tool rather than the primary creator of content.

When it came to assessing and grading the students' work, the assessment criteria were adapted to account for integrating AI tools into the process. We realised that the inclusion of generative AI, to an extent, raises the bar for what lecturers can expect from students, enabling them to deliver higher-quality work, such as submissions that are free of unnecessary language errors. This shift in expectations potentially reflects the broader impact of AI tools on enhancing the quality of student outputs.

Another critical point to highlight is that throughout the module (and in line with institutional guidelines) students are informed about how and when the lecturer uses generative AI to support their academic process. For example, generative AI was employed to refine topics before presenting them to students and to brainstorm assignments and learning flows before trialling them in the classroom. By sharing this practice, we, as higher education practitioners, model transparency and responsible use of AI (as described by Dignum, 2019), offering students a real-world example to navigate these tools ethically and effectively.

Key understandings

Through critical reflection and collegial discussion, we engaged with Rolfe, et al.'s (2001) *so what* question in relation to the integration of AI literacies into this learning and assessment opportunity. This led to three key understandings that highlight how students grapple with problem formulation, build engagement through interaction, and navigate the complexities of generative AI in ways that reconfigure relationships of trust and scepticism in higher education.

Problem identification and formulation

In line with Acar's (2023) view, we assert that effective problem formulation, rather than prompt engineering, is a critical skill for students – both in their academic studies and in their future worlds of work. Recognising that students often struggle with identifying and formulating research problems, we encouraged the critical use of generative AI as a brainstorming partner to help them diagnose the core issue, break it into smaller, manageable components, and analyse it. Through this interaction, students learn how to approach complex problems for their own sense-making, practise the skill of problem formulation itself, and communicate effectively with AI in ways that support, rather than substitute, their thinking. An added benefit is that students

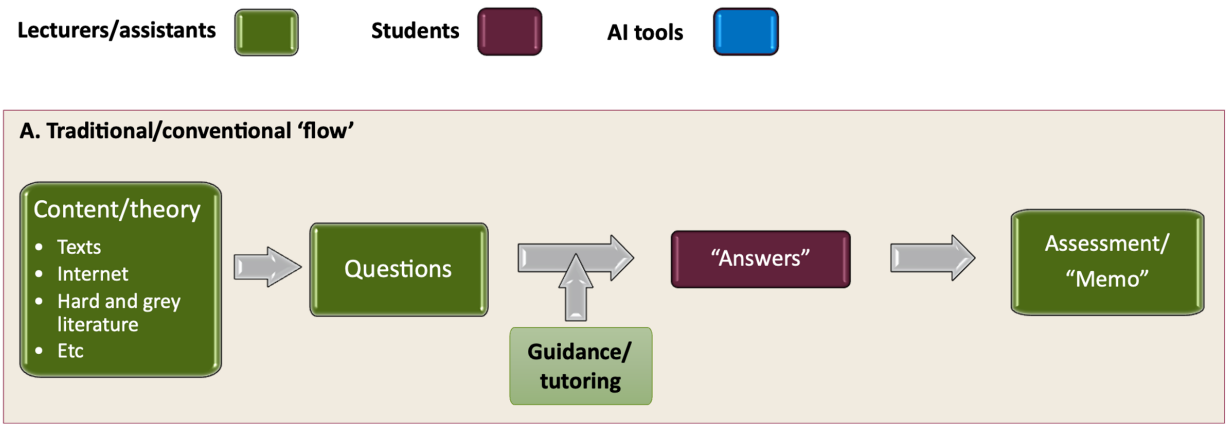
[7] In South African HEI's, English is often the academic language and lingua franca, despite it only being the fifth largest home language in our multilingual country, with isiZulu, isiXhosa, Afrikaans, and Sepedi ranking above it (StatsSA, Census 2022).

are less likely to become stuck when analysing their topics and beginning to formulate research questions, which creates more space for innovative thinking and creative solutions to their research problems. This type of interaction offers students the opportunity to form a downward trust relationship with AI, treating it as a tool that can enhance their work and support new insights, but one that still needs to be engaged with critically.

Enhanced student interaction and engagement

The integrated approach to generative AI in this module led to noticeably higher levels of student participation and engagement throughout the learning and formative assessment process, likely due to the involvement of novel technologies, which are also fun and interesting to engage with. Research supports the idea that learning-centred approaches and active student engagement both contribute to overall student success (Wheeler & Bach, 2021). In our case, students appeared more involved than in previous iterations of the module, probably because of their constant (critical) engagement with AI tools. Based on these observations, we propose that this integrated approach to generative AI could foster similar levels of heightened student involvement in other academic contexts.

Figure 2 presents a proposed adaptation of 'traditional learning flows' through the integration of generative AI tools. The first 'A' block shows the traditional flow of information (or 'learning') in the lecturer-student-content triad (Fraser, 2006). Lecturers or teaching assistants present the content (potentially consisting of academic texts, internet/multimedia sources, and other resources) and formulate questions based on these inputs. They provide support and guidance to students so that they can answer the posed questions, and afterwards, an assessment and/or memorandum provides students with the 'correct' answers to the questions. Within this configuration, students have a very small role to play in their learning environment. Considering the directionality and number of arrows indicating the flow of information, students do not seem to be very active in the learning process. They are fed information and expected to answer questions about the content, not necessarily learning or grappling with the work.



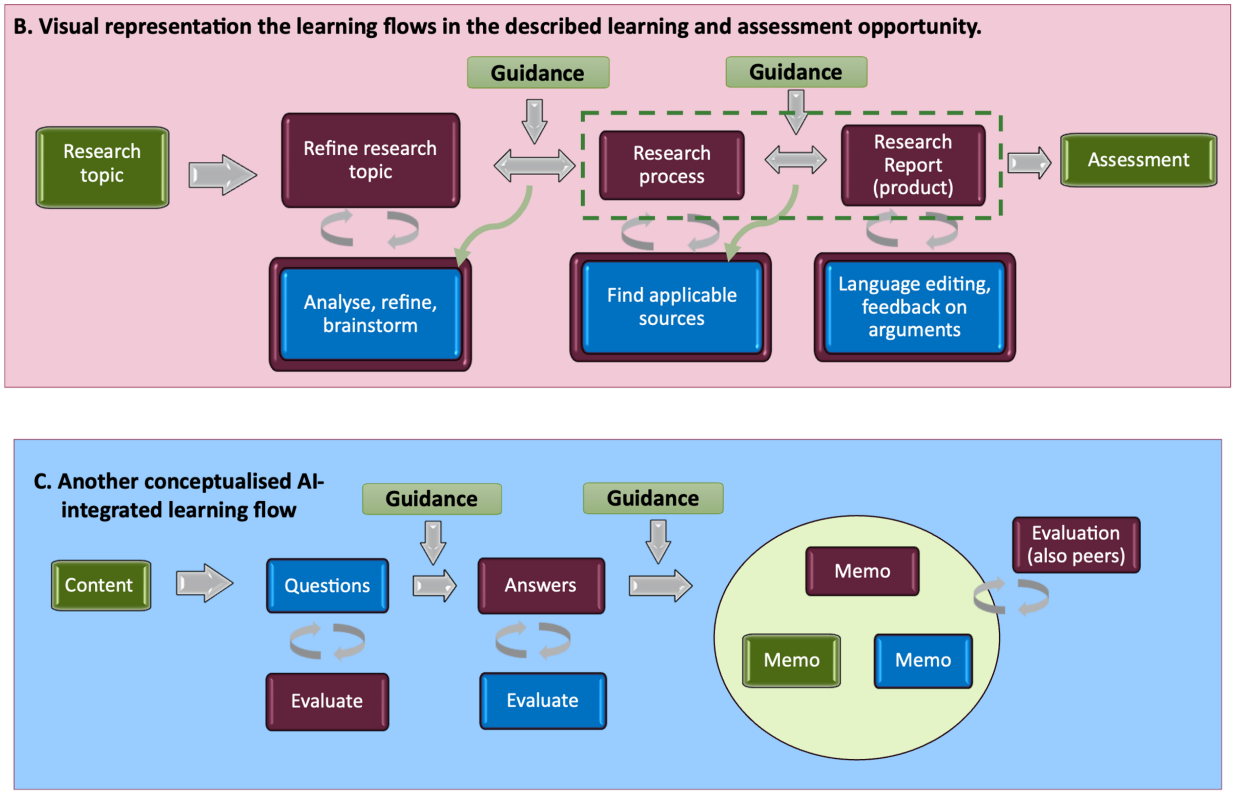


Figure 2: Proposed adaptation of 'traditional learning flows' through an integrated AI approach

The second, B block of Figure 2 is a visual representation of the learning flow in our described learning and assessment opportunity. The lecturer initiates the process by providing broad research topics, but from that point forward, students take the lead in shaping the project. They actively refine their chosen topics with the help of generative AI for problem formulation, conduct research using appropriate AI tools, and later utilise AI for language editing and feedback during the writing phase. In the diagram, the blue 'AI' blocks are literally enclosed within maroon 'student' blocks, symbolising that students actively control and direct the tasks undertaken with generative AI tools. Throughout this process, students are active agents and the primary doers in the learning environment. The iterative nature of the process is represented by multidirectional and circular arrows, indicating ongoing revisions based on interactions with generative AI tools, group members, and the lecturer or teaching assistants. Lecturers remain available to guide students through each phase and support them in navigating their interactions with generative AI when necessary. While assessment appears at the end of the diagram, the dotted line illustrates that assessment encompasses both the research process and the final product, thereby valuing and evaluating the entire learning process. Compared to block A, this flow illustrates significantly greater student agency – quite literally shown by the increase in maroon 'student' blocks – signalling a shift towards students as the primary actors in the learning environment.

To further argue the case for integrating AI literacies, we have conceptualised another way that it can play out in TLA practice, as visualised block C of Figure 2. Unlike the extended,

research-focused process in block B, this AI-integrated learning flow highlights shorter iterative cycles of questioning, answering, evaluating, and memo-building. Here, AI is embedded not only as a writing support but as part of the dialogic process through which students generate, test, and refine knowledge collaboratively. Here, the lecturer remains a guide on the side, providing input only when necessary to encourage student participation. The lecturer designs and delivers specific module content, after which students, possibly with the help of generative AI, formulate questions, create answers using their learning materials and AI tools, and refine these answers through further research and reflection. They are encouraged to critically evaluate all outputs, both their own and those generated by AI, by comparing them with course content and reflecting on their understanding. At this stage, lecturers can offer their own memorandum of answers alongside one generated by AI for students to scrutinise. Students then compare their work with both sets of answers, evaluating the differences and limitations in each. This simulates a peer-evaluation environment where, as research shows, students' evaluation of multiple exemplars enables deeper learning (Nicol & McCallum, 2022). This can also lead to class discussions around divergent interpretations, highlighting the differences in results produced by various AI tools, depending on the problem analysis and prompting techniques used and the inclusion (or exclusion) of validated sources. Compared to Block A, this flow reflects heightened student engagement, evident in the greater number of maroon 'student' blocks, while also positioning AI less as a peripheral tool and more as a learning partner in the dialogic process.

Introducing generative AI tools and the associated AI literacies in the ways set out above creates new opportunities for students to engage more deeply not only with the module content, but also with the lecturer, teaching assistants, and peers in the TLA environment. The AI-integrated learning flows promote an interactive process, encouraging students to iterate, re-evaluate their work, and refine their understanding through their engagement with generative AI. This is consistent with inquiry-based and problem-based learning approaches, particularly when the problem statement becomes central to the integration of generative AI. In these flows, generative AI serves as a brainstorming partner, providing instant feedback on students' ideas and arguments. Meanwhile, the lecturer and teaching assistants remain available to offer guidance, share subject expertise, and assist when students encounter challenges or over-rely on the technology.

Navigating the relationship with generative AI in TLA

As students interact with and assess AI outputs, they gain first-hand experience of both the strengths and weaknesses of generative AI technologies, deepening their engagement with the research process in ways that traditional teaching methods might not have achieved in the past.

One significant outcome of integrating AI literacies in the learning opportunity described, is that students learn to navigate their relationship with generative AI within their specific disciplinary contexts. Traditionally, within a social constructivist approach, the roles of lecturers, teaching assistants, and peers are designed to guide student learning as reliable and trustworthy figures (see Figure 1). However, through guided first-hand experience with generative AI, the

students in the module described don't run the risk of regarding generative AI as a more knowledgeable entity or an equal partner in the TLA environment. This is a crucial insight for navigating AI in educational spaces. Humans tend to project interpersonal trust onto AI, but as Hancock, et al. (2023) suggest, this relationship should be conceptualised as downward trust. In this view, AI is placed in a subordinate role, where it 'reports' to humans – in this case, the students. Through this positioning, students remain accountable for their learning, while continuously assessing generative AI outputs. Block B in Figure 2 illustrates some of this dynamic by framing generative AI within the context of students' interactions, emphasising their control of the tools and responsibility towards their learning.

A second, interesting outcome of integrated AI literacies was recognition of the importance of critically evaluating knowledge and information, irrespective of whether it is offered by generative AI or a human. As discussed earlier, generative AI tools often lack the depth, nuance, or contextual understanding that human instructors possess, necessitating a sceptical approach when engaging with it. Interestingly, though, these tools can at other times provide surprising insights or suggest solutions that may even surpass those offered by lecturers. This paradox, where AI can be simultaneously less reliable yet occasionally insightful, requires that both students and lecturers continuously and rigorously test AI-generated outputs through conventional methods of critical evaluation and fact-checking, thereby developing their evaluative judgement skills (Bearman, et al., 2024; Tai, et al., 2018). We argue that these instances would further compel students to take a greater critical stance towards all forms of knowledge and information in the TLA space, encouraging a criticality towards the inputs provided by the human role players. In other words, we believe that through integrating AI literacies, students learn to distrust not only the outputs of generative AI, but also to transfer healthy scepticism to other partners, such as lecturers and teaching assistants, and critically evaluate all forms of knowledge shared. As students engage with generative AI, they not only scrutinise the technology; they also re-examine their assumptions about knowledge and authority. Perhaps lecturers are not the infallible sources of knowledge as once perceived, and perhaps other learning partners, such as peers or TLA assistants are not to be trusted blindly either.

Conclusion: Towards pedagogies of distrust?

Through this reflection on a learning and assessment opportunity, we argue that integrated AI literacies enhance student engagement and learning, while refining skills such as problem formulation and fostering a critical stance. The criticality should extend not only toward generative AI tools, but also toward other role players in the TLA environment, such as lecturers and teaching assistants, who have traditionally been regarded as the more knowledgeable authorities. A healthy scepticism toward all participants in the higher education classroom may prompt a reconceptualisation of trust in TLA – framing trust as more relative and less automatic.

This sentiment aligns with that of Eaton (2024b) who states that, in our current global climate and within higher education specifically, trust should not be accepted as a given. She frames trust as 'not merely a goal' to be achieved, 'but something earned through continuous

practices of living, leading, and learning' (Eaton, 2024a). Rolfe, et al.'s (2001) *now what* question prompts us to take this argument even further, contemplating whether an environment of trust should be cultivated or earned at all. Instead of overprotecting students in an environment of trust, should we not be preparing them for the world beyond higher education, which is already (socially speaking) in a state of distrust (Norris, 2022)?

In this light, should we aim to introduce a 'pedagogy of distrust' in higher education? Not the kind of negative distrust that fosters surveillance and suspicion of students (Luo, 2024), nor the managerialism that replaces trust with oversight in the higher education workplace (Carless, 2009). Instead, we advocate for a kind of healthy scepticism, one that encourages critical thinking and interrogation of the basic assumptions of constructs such as knowledge and who shares it (Norris, 2022). In this environment of productive distrust, all role players in the TLA space remain open to critique and challenge, leading to the creation of new insights and the construction of shared knowledge. By critically engaging with the knowledge presented, analytically interacting with generative AI, and fostering slower, reflective learning (Zournazi, 2022), we can create spaces where students and lecturers alike can learn from one another on a more equal footing.

To this end, lecturers may also need to reconsider their roles in the higher education classroom. No longer solely the source of all theoretical knowledge and the ultimate specialist authority, they can instead embrace new roles of assisting students in self-navigating the unknown towards a goal of lifelong learning, augmented by technologies such as generative AI. This shift reflects the age-old adage of 'teaching someone to fish' – equipping students with the skills, mindset, and tools to learn independently, critically, and continuously in an ever-changing world.

Author Biographies

Dalene Joubert is a Senior Advisor (Higher Education) at the Centre for Teaching and Learning and the Faculty of AgriSciences at Stellenbosch University. Her research explores the intersection between generative AI and TLA in higher education, and the South African context specifically.

Albert Strever is a Senior Lecturer and Chair of the Department of Viticulture and Oenology, also coordinating innovation and entrepreneurship in the Faculty of AgriSciences, at Stellenbosch University. He is a TAU fellow and specialises in the development of entrepreneurship and its integration into non-business school curricula, as well as the use of AI in TLA and research at the university.

AI declaration

We used generative AI tools (ChatGPT, Copilot, Claude) in the same way we asked of students: as a brainstorming partner, critical friend and proofreader – thereby modelling responsible use in academic work. We remained mindful of the limitations and biases of generative AI tools and did not rely on their outputs as sources of information.

References

Acar, O.A. 2023. AI prompt engineering isn't the future. *Harvard Business Review*, 6 June. Online at: https://hbr.org/2023/06/ai-prompt-engineering-isnt-the-future (Accessed: 18 September 2024).

Adendorff, H.J. 2024. AI literacy a critical component in 21st-century learning. *University World News*, 5 September. Available at: https://www.universityworldnews.com/post.php?story=20240902234739542 (Accessed: 25 September 2024).

Adendorff, H.J., Herman, N. & Joubert, D. 2025. Avoiding human deepfakes: A learning and care-centred approach to higher education. In Pellissier, R. (Ed.). *Critical Conversations in Higher Education.* Stellenbosch: Cape Higher Education Consortium. Stellenbosch: SUN Press. 7-29.

Akpan, V.I., Igwe, U.A., Mpamah, I. & Okoro, C.O. 2020. Social constructivism: Implications on teaching and learning. *British Journal of Education*, 8: 49–56.

Ashwin, P. 2012. *Analysing Teaching-Learning Interactions in Higher Education: Accounting for Structure and Agency*. London: Bloomsbury.

Baier, A. 1986. Trust and antitrust. *Ethics*, 96: 231–260.

Bates, A.W. 2022. *Teaching in a Digital Age* (3rd ed.). Vancouver: Tony Bates Associates Ltd.

Bearman, M., Tai, J., Dawson, P., Boud, D. & Ajjawi, R. 2024. Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49(6): 893–905.

Carless, D. 2009. Trust, distrust and their impact on assessment reform. *Assessment & Evaluation in Higher Education*, 34: 79–89.

Carter, O.B.J., Loft, S. & Visser, T.A.W. 2023. Meaningful communication but not superficial anthropomorphism facilitates human-automation trust calibration: The Human-Automation Trust Expectation Model (HATEM). *Human Factors*, 66(11): 2485–2502.

Choi, S., Jang, Y. & Kim, H. 2023. Influence of pedagogical beliefs and perceived trust on teachers' acceptance of educational artificial intelligence tools. *International Journal of Human–Computer Interaction*, 39(4): 910–922.

Curzon-Hobson, A. 2002. A pedagogy of trust in higher learning. *Teaching in Higher Education*, 7(3): 265–276.

Dai, Y., Liu, A., Lim, C.P. 2023. Reconceptualizing ChatGPT as a student-driven innovation in higher education. Procedia CIRP, 119: 84-90.

Dann, R. 2014. Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2): 149–166.

Ding, L., Li, T., Jiang, S. & Gapud, A. 2023. Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education*, 20: 63.

Dignum, V. The ART of AI: Accountability, Responsibility, Transparency. *Medium*, 4 March.

Available at: https://medium.com/@virginiadignum/the-art-of-ai-accountability-responsibility-transparency-48666ec92ea5 (Accessed: 25 September 2024).

Eaton, S.E. 2024a. Trust as a Foundation for Ethics and Integrity in Educational Contexts. *HECU11*. Available at: https://sites.google.com/ru.ac.za/hecu11/sarahs-think-piece#h.ofj5r1phofu (Accessed: 24 September 2024).

Eaton, S.E. (Ed.). 2024b. *Second Handbook of Academic Integrity*. Springer International Handbooks of Education. Cham: Springer Nature.

Fourie, J. 2024. My students used AI to write essays. It went wrong. Badly. *News24.com*, 16 June. Available at: https://www.news24.com/fin24/opinion/johan-fourie-my-students-used-ai-to-write-essays-it-went-wrong-badly-20240615 (Accessed: 30 September 2024).

Fraser, J.D.C. 2006. 'Mediation of learning. In Nieman, M.M. & Montai, R.B. (Eds.). *The Educator as Mediator of Learning*. Pretoria: Van Schaik.

Gerlich, M. 2024. Exploring motivators for trust in the dichotomy of human: AI trust dynamics. *Social Sciences*, 13: 251.

Hancock, P.A., Kessler, T.T., Kaplan, A.D., Stowers, K., Brill, J.C., Billings, D.R., Schaefer, K.E. & Szalma, J.L. 2023. How and why humans trust: A meta-analysis and elaborated model. *Frontiers in Psychology*, 14: 1081086.

Heaven, D.H. 2023. The inside story of how ChatGPT was built from the people who made it. *MIT Technology Review*, 3 March. Available at: https://www.technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/ (Accessed: 25 September 2024).

Hoy, W. & Tschannen-Moran, M., 1999. Five faces of trust: An empirical confirmation in urban elementary schools. *Journal of School Leadership*, 9(3): 184–208.

Hutchinson, E. 2024. Navigating tomorrow's classroom: The future of information literacy and inquiry-based learning in the age of AI. *Journal of Information Literacy*, 18.

Jacobs, C. 2005. On being an insider on the outside: New spaces for integrating academic literacies. *Teaching in Higher Education*, 10(4): 475–487.

Joubert, D. & Strever, A.E. 2023. 'AI-enabled learning'. *AI2 Discussion Series*. Available at: https://www.youtube.com/watch?v=ZiWhieHB4Lg (Accessed: 25 September 2024).

Kumar, R., Eaton, S.E., Mindzak, M. & Morrison, R. 2023. Academic integrity and artificial intelligence: An overview. In Eaton, S.E. (Ed.). *Handbook of Academic Integrity*. Singapore: Springer Nature, 1–14.

King, A. 1993. Sage on the stage to guide on the side. *College Teaching*, 41(1): 30-35.

Liu, C.H. & Matthews, R. 2005. Vygotsky's philosophy: Constructivism and its criticisms examined. *International Education Journal,* 6(3): 386–399.

Lindgren, S. 2023. *Critical Theory of AI*. Cambridge: Polity Press.

Lukyanenko, R., Maass, W. & Storey, V. C. 2022. Trust in artificial intelligence: From a foundational trust framework to emerging research opportunities. *Electronic Markets*, 32: 1993–2020.

Luo, J. 2024. How does generative AI affect trust in teacher-student relationships? Insights from

students' assessment experiences. *Teaching in Higher Education*, *30*(4): 991–1006.

Mayer, R.C., Davis, J.H. & Schoorman, F.D. 1995. An integrative model of organizational trust. *The Academy of Management Review*, 20(3): 709–734.

Montag, C., Klugah-Brown, B., Zhou, X., Wernicke, J., Liu, C., Kou, J., Chen, Y., Haas, B.W. & Becker, B. 2023. Trust toward humans and trust toward artificial intelligence are not associated: Initial insights from self-report and neurostructural brain imaging. *Personality Neuroscience*, 6(3): 1–8.

Nicol, D. & McCallum, S. 2022. Making internal feedback explicit: Exploiting the multiple comparisons that occur during peer review. *Assessment & Evaluation in Higher Education*, 47(3): 424–443.

Norris, P. 2022. Evidence. In Norris, P. (Ed.). *In Praise of Skepticism: Trust but Verify*. Oxford: Oxford University Press: 52-94.

Okoye, F. 2023. Reinventing student–teacher relationship in higher education institutions of developing nations: Lessons from the University of the Free State. *HAYEF: Journal of Education*, 20(2): 119–127.

OpenAI. 2024a. GPT-4. Available at: https://openai.com/index/gpt-4-research/ (Accessed: 9 September 2024).

OpenAI. 2024b. Terms of use. Available at: https://openai.com/en-GB/policies/row-terms-of-use/ (Accessed: 9 September 2024).

Robbins, B.G. 2016. What is trust? A multidisciplinary review, critique, and synthesis. *Sociology Compass*, 10(10): 972–986.

Rolfe, G., Freshwater, D., Jasper, M. 2001. *Critical reflection in nursing and the helping professions: a user's guide.* Basingstoke: Palgrave Macmillan.

Ryan, M. 2020. 'In AI we trust: Ethics, artificial intelligence, and reliability'. *Science and Engineering Ethics*, 26: 2749–2767.

Statistics South Africa. CENSUS 2022 RESULTS. Available at: https://census.statssa.gov.za/assets/documents/2022/Census_2022_SG_Presentation_10102023.pdf (Accessed: 30 September 2024).

Stellenbosch University. 2022. Stellenbosch University Assessment Policy. Available at: https://www.sun.ac.za/english/learning-teaching/ctl/Documents/SU%20Assessment%20Policy.pdf (Accessed: 30 September 2024).

Stellenbosch University. 2023a. AI literacies Framework'. Online at: https://www.sun.ac.za/english/learning-teaching/ctl/Documents/AI%20literacies%20framework.png (Accessed: 15 September 2025).

Stellenbosch University. 2023b. Draft Interim SU Guidelines on Allowable AI Use and Academic Integrity in Assessment. Available at: https://www.sun.ac.za/english/learning-teaching/ctl/Documents/Interim%20SU%20guidelines%20on%20allowable%20AI%20use%20and%20academic%20integrity.pdf (Accessed: 30 September 2024).

Stellenbosch University. 2025. Teaching-Learning Policy. Available at:

https://www.sun.ac.za/english/learning-teaching/ctl/Documents/Final%20T-L%20Policy_Council%20approved%202%20Dec%202024_English.pdf (Accessed: 15 September 2025).

Tai, J., Ajjawi, R., Boud, D., Dawson, P. & Panadero, E. 2018. Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education*, 76: 467–481.

Tavani, H. & Zimmer, M. 2025. Search engines and ethics. *The Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/entries/ethics-search/ (Accessed: 15 September 2025).

Van Dis, E.A.M., Bollen, J., Zuidema, W., van Rooij, R. & Bockting, C.L. 2023. ChatGPT: Five priorities for research' *Nature*, 614: 224–226.

Vygotsky, L.S. 1978. *Mind in Society: Development of Higher Psychological Processes*. Cambridge: Harvard University Press.

Walker, M.U. 2006. *Moral Repair: Reconstructing Moral Relations after Wrongdoing*. Cambridge: Cambridge University Press.

Wheeler, L.B. & Bach, D. 2021. Understanding the impact of educational development interventions on classroom instruction and student success. *International Journal for Academic Development*, 26(1): 24–40.

Whetten, D.A. 2007. Principles of effective course design: What I wish I had known about learning-centered teaching 30 years ago. *Journal of Management Education*, 31(3): 339–357.

Yu, H. 2023. Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, 14.

Zhou, Z. 2023. Towards a new definition of trust for teaching in higher education. *International Journal Scholarship of Teaching & Learning*, 17(2): 1–13.

Zournazi, M. 2022. Building dwelling caring: Some reflections on the future of learning. *Critical Studies in Teaching & Learning*, 10, 151–163.