



Evaluating GenAI's Ability to Map a Professional Engineering Skills Framework to Self-Agency and Feedback Literacy

Zachery Quince^a, Anna Lidfors Lindqvist^b, and Patcharin Chen^c

Centre for Teaching and Learning, Southern Cross University^a, School of Mechanical and Mechatronics Engineering, University of Technology Sydney^b, Department of Materials Science and Engineering, Monash University^c

Corresponding Author's Email: Zach.quince@scu.edu.au

ABSTRACT

CONTEXT

The use of thematic mapping between different frameworks or themes can present significant challenges, often introducing human error and bias. This issue is widely recognised as a key limitation in a mapping via thematic analysis. While digital tools and artificial intelligence (AI) have been integrated into mapping processes to enhance consistency, the potential of generative AI (GenAI) in mitigating these limitations remains underexplored.

GOAL

This research aims to evaluate the efficacy of GenAI tools in conducting thematic mapping processes within the domain of professional engineering skills frameworks. Specifically, it seeks to critically assess the ability of ChatGPT to perform thematic mapping and provide justifications for its outputs.

APPROACH

ChatGPT models will be employed to thematically map three contemporary themes of feedback literacy, self-agency, and digital tool proficiency against a professional engineering skills framework developed by the authors. The study will not only examine the accuracy and consistency of the mapping performed by GenAI but also assess the underlying justifications provided by GenAI. The justifications will be compared with expert's evaluations to determine an extent to which ChatGPT can align its outputs with professional standards and ethical considerations.

OUTCOMES

It is expected that ChatGPT will generate plausible justifications for some of the themes, demonstrating potential utility in thematic mapping. However, a deeper, more nuanced understanding of the themes may be absent, particularly in areas requiring contextual reasoning, ethical judgment, or domain-specific expertise. The study will explore whether iterative refinement or human intervention is necessary to bridge these gaps.

CONCLUSIONS

This research contributes to the growing discourse on the role of GenAI in engineering education by critically evaluating the strengths and limitations of ChatGPT in thematic mapping. The findings will provide insights into the feasibility of integrating AI-driven mapping processes in educational and professional contexts while highlighting areas where human expertise remains indispensable.

KEYWORDS

Thematic mapping, generative AI, engineering education, self-agency, feedback literacy.

Introduction

Thematic mapping is the process of systematically aligning content or activities to a set of themes or categories (Braun & Clarke, 2006). Beyond its recognition as a research methodology, it is widely used in curriculum accreditation and constructive alignment practice in education. This often takes the form of curriculum mapping, where program learning outcomes, course content and assessment tasks are mapped to overarching frameworks or standards (Carew et al., 2013). Such mapping provides a “snapshot” of what skills and knowledge are taught or helping educators identify gaps and overlaps in the curriculum (Quince et al., 2023). Whilst this process is good practice, it also serves as a function in enabling external stakeholders, such as accreditation bodies, to understand the structure behind a study program. In engineering education, programs must satisfy accreditation criteria, such as Accreditation Board for Engineering and Technology (ABET) outcomes or Engineers Australia’s (EA) Stage 1 competencies. The programs are mapped to the required graduate attributes to demonstrate alignment (Engineers Australia, 2019; Carew et al., 2013). Mapping can be conducted in reverse by mapping key skills or content from frameworks such as EA stage 1 competencies or the newly developed Competency in Professional and Authentic Skills for Success (COMPASS) framework (Quince et al, 2025) to investigate where and in which contexts these skills should be taught. The COMPASS framework is a derivation of the Washington Accord with enhanced language for stakeholders’ accessibility.

By auditing the curriculum against external or internal standards, educators can ensure that necessary skills are introduced, reinforced and assessed in alignment with institutional or professional expectations. The process inherently relies on expert judgment where individuals must interpret meaning and alignment. Different educators or professionals may reasonably disagree on these interpretations, leading to subjectivity and variability in mapping results (Uchiyama & Radin, 2009). Prior studies highlight the “pervasive vagueness and inconsistency” in how skills and outcomes are defined and understood by different stakeholders (Sumsion & Goodfellow, 2004). The perspective of each stakeholder varies as well, which can lead to certain themes being prioritised (Quince et al, 2024). Without careful moderation and consensus, one person’s thematic map might reflect personal biases or assumptions rather than a shared understanding of the data. To address this, researchers often use intercoder agreement checks, peer moderation or structured frameworks to improve consistency (O’Connor and Joffe, 2020).

Recent advances in digital tools including natural language processing (NLP) and generative artificial intelligent (GenAI) offer new approaches to curriculum and thematic mapping by enhancing scalability, consistency and processing time. NLP techniques have been applied to automatically map course outcomes to program outcomes or professional competencies with good performance (Zaki et al., 2023). These tools can analyse large volumes of text and identify thematic alignment using keyword extraction, similarity metrics or trained classification models. GenAI models like ChatGPT take this a step further by engaging in interpretive reasoning and simulating human-like responses to prompts (Hagendorff et al., 2023; Oyeniyi, 2024). GenAI can generate thematic groupings, propose mappings and justify decisions in language which makes them desirable for curriculum reviews. These capabilities offer promising ways to extend thematic mapping practices, particularly in large-scale evaluations or curriculum reform initiatives. Despite its strengths, GenAI still faces substantial limitations in making context-sensitive or ethically informed judgments. This limits their ability to navigate ambiguous or ethically complex interpretations (Quince & Nikolic, 2025).

Furthermore, biases embedded in training data can subtly influence how GenAI aligns skills or interprets themes which potentially reinforcing existing stereotypes or privileging dominant discourses (Rehman et al., 2025). This “black box” nature makes it difficult to assess when their outputs are genuinely insightful or merely plausible sounding. As engineering education increasingly explores AI integration, these limitations must be critically evaluated especially in domains involving complex judgment.

To explore the potential of GenAI tools in the mapping process, this study employs three GenAI models of ChatGPT 4o, ChatGPT o1 and Microsoft Copilot to perform thematic mapping of two contemporary skills, feedback literacy and self-agency, against the professional engineering COMPASS framework. Both skills are essential to engineers, but they are not considered within the competency standards as they belong to a broader skill set which allows engineers to successfully engage with various tasks rather than being an explicit competency.

Context

As the engineering profession become more complex, interdisciplinary and focus on social responsibility, it increasingly requires not just technical skills but also interpersonal and professional abilities (Male et al., 2011). These skills include initiative, adaptability and continuous learning. Central to this is self-agency, which is the ability of individuals to take ownership of their learning, make autonomous decisions and direct their own development (Bandura, 2006). Despite its clear relevance to engineering practice, self-agency is often assumed rather than explicitly cultivated in traditional curricula. Equally important is feedback literacy, which refers to the ability to understand, engage with and act on feedback to enhance performance (Carless & Boud, 2018). Engineers often work in iterative environments where the ability to reflect on performance, adapt to feedback and revise their approaches are essential skills. A lack of feedback literacy can hinder continuous improvement and collaborative problem-solving (Molloy et al., 2020). Both self-agency and feedback literacy capabilities reflect a shift in what it means to be “professionally ready”. The Engineering Futures 2035 (Crosthwaite, 2021; Crosthwaite et al., 2018) report emphasises the need for educational experiences that mirror real-world contexts, where engineers are not simply told what to do, but must determine how to proceed in uncertain and evolving situations. Embedding self-agency and feedback literacy into engineering education involves creating opportunities for students to set goals, make decisions, reflect on progress and seek or respond to feedback in meaningful ways (Ajjawi et al., 2022). These are not peripheral traits but central professional skills. An example of this practice in assessment design is the use of formative sprints and the prioritisation of feedback over marks (Lidfors Lindqvist et al. 2023; Lidfors Lindqvist et al. 2024).

The rise of GenAI in higher education highlights the increasing need to develop self-agency and feedback literacy. Tools like ChatGPT give students quick access to information and feedback. However, these tools risk encouraging passive rather than active engagement if the students do not have the ability to apply judgment, independence and accountability (Quince & Nikolic, 2025). Therefore, supporting students to use GenAI tools critically, ethically and strategically requires self-agency that must be fostered and not assumed. By making self-agency and feedback-literacy a visible and an intentional goal of engineering education, institutions can better prepare students to take ownership of their learning and contribute meaningfully to the profession. This shift benefits graduates and supports a more adaptive, reflective and future-ready engineering workforce.

Methods

GenAI Tools

This study evaluated three GenAI models from two providers for their ability to perform thematic mapping: ChatGPT 4o, ChatGPT o1 and Microsoft Copilot. ChatGPT 4o was chosen as a default model, while o1 (the internal reasoning model) was used to assess the system's interpretive capabilities. This approach allowed for a comparison of the generated explanations and also assessing thematic alignment of each tool when tasked with making internally consistent judgments. Microsoft Copilot was included as a comparator due to its increasing integration into productivity tools and adoption in both professional and educational settings.

Prompt Design

Designing the prompt for GenAI was a critical step that directly impacts the accuracy, consistency and relevance of the output. Recent literatures provide guidance on how to structure high quality prompts to elicit good performance from GenAI tools. When writing prompts for academic purposes, it is importance to use clear task directives, role-setting and instructional framing (Giray, 2023). Additionally, a typology of prompt formats such as instruction-based prompts is also important (Liu et al., 2024). Both Giray (2023) and Liu et al (2023)'s systematic survey also supports the use of prompts that explicitly define the task, provide context and anticipate model reasoning patterns.

Following this guidance, the prompt in this study was structured to include: (1) a clear instruction to align a target skill with a professional competency framework, (2) a description of the alignment rubric to guide response calibration (Table 1), and (3) an explicit request for explanation to improve interpretability. To ensure that the GenAI platforms were using the same definitions to map the skills, each GenAI tool was also prompted to define self-agency and feedback literacy skills in the broad sense without the engineering context. This structure was intended to promote both response relevance and transparency across the different models. The prompt included clear instructions, contextual information and a defined rubric for alignment levels. The prompt is as follows.

"I'd like you to analyse a specific skill or concept and determine where it aligns thematically within a given professional skill framework, even if it's not explicitly mentioned. Here's the skill or idea: 'INSERT. The framework is attached. Please infer alignment by identifying where this skill might be applied, expressed, or implied within the framework. Also, please provide an explanation as to why you believe there is alignment. Please use this guide to provide your conclusions"

Table 1: Description of level of alignment used for the mapping process

Alignment Level	Symbol	Description
Strong	✓✓	The skill is fundamental in supporting the competency in the framework.
Partial	✓	The skill supports the competency in the framework but is not essential.
Minimal/no alignment	-	There is no direct link between the skill and the competency in the framework.

Evaluation Method

For self-agency and feedback literacy skills, all three GenAI models were asked to perform a full thematic mapping against the 35 sub-descriptors of the COMPASS framework (Table 2).

Table 2: COMPASS framework

Code	Skill Sub-Descriptor	Code	Skill Sub-Descriptor	Code	Skill Sub-Descriptor
1.1	Identify, analyse, and solve complex engineering problems supported by evidence.	5.4	Consider environmental impacts in engineering decisions.	8.2	Communicate effectively with society.
1.2	Solve engineering problems with sustainability and holistic perspectives in mind.	5.5	Understand relevant legal frameworks in engineering.	8.3	Develop reports and presentations that respect cultural and linguistic diversity.
2.1	Design innovative engineering solutions to meet project needs.	6.1	Adhere to relevant national and international laws and regulations.	9.1	Apply financial decision-making to individual engineering projects.
2.2	Develop solutions considering public health and safety.	6.2	Demonstrate an understanding of diversity and inclusion.	9.2	Apply management principles in individual engineering tasks.

2.3	Develop solutions considering project lifecycle, costs, and societal impacts.	6.3	Uphold ethical standards and professional norms in engineering.	9.3	Contribute to team projects with financial decision-making skills.
3.1	Design experiments, analyse and interpret data, and synthesise information for valid conclusions.	7.1	Receive feedback, evaluate its relevance, and take necessary action.	9.4	Apply management principles to teamwork in multidisciplinary settings.
3.2	Use research methods to investigate complex engineering problems.	7.2	Provide constructive feedback.	9.5	Lead projects with sound economic decision-making.
4.1	Identify suitable modern tools and techniques, understanding their strengths and limitations	7.3	Contribute effectively as a team member.	9.6	Manage multidisciplinary projects effectively as a leader.
4.2	Select suitable modern tools and techniques, understanding their strengths and limitations	7.4	Lead teams effectively.	10.1	Undertake independent and lifelong learning to grow professionally.
5.1	Evaluate the sustainability of engineering solutions.	7.5	Perform self-directed tasks.	10.2	Adapt to emerging technologies in the engineering field.
5.2	Assess the socioeconomic impacts of engineering practices, decisions, and solutions.	7.6	Collaborate in diverse and multidisciplinary environments.	10.3	Develop critical thinking to navigate technological change.
5.3	Analyse health and safety concerns in engineering practices.	8.1	Communicate effectively within the engineering community.		

The responses from each GenAI tool were recorded and coded according to the provided alignment categorisation table (Table 1). In cases where justification was requested, models were instructed to explain why the alignment existed. Copilot interpreted this instruction as only providing justification when an alignment was identified (i.e., in cases of “–” alignment). This behavioural difference was noted during data collection and considered in the analysis.

After initial mapping, each GenAI tool’s output was compared to a consensus mapping developed by the authors, established through expert review and deliberation. The authors’ mapping served as a reference point for identifying divergence between the GenAI outputs and expert expectations. Further details of this mapping can be found in (Lidfors Lindqvist et al., 2025). Comparisons were made on two levels: (1) Quantitative agreement with the authors’ consensus, and (2) Interpretive divergence where critical differences were observed. No further analysis was conducted when there was alignment between the models’ and the authors’ interpretations. However, a detailed qualitative comparison of justifications was undertaken for sub-descriptors where critical divergence or misalignment occurred, that is when at least two models generated outputs that differed substantially from the authors’ consensus. This analysis focused on identifying whether the GenAI tools’ reasoning was plausible, internally consistent and sensitive to context.

Results and Discussion

Skill definition

Both ChatGPT and Copilot generated similar definitions for feedback literacy. Table 3 outlines the definitions of feedback literacy provided by three different GenAI models used.

Table 3: GenAI models interpretation of feedback-literacy

Tool	Skill definition
ChatGPT 4o	Feedback literacy as a skill is the capacity to understand, interpret, and effectively use feedback to improve performance or learning. It includes both giving and receiving feedback constructively, recognizing its value, and integrating it into future actions or work. Key components: Understanding the purpose and process of feedback. Interpreting feedback accurately (even when it's vague or critical). Seeking out feedback proactively. Applying feedback to improved. Responding reflectively and non-defensively.
ChatGPT o1	Feedback literacy, as a skill, is the capacity to effectively seek, interpret, and use feedback to learn, grow, and refine one's performance or understanding. It includes knowing how to ask for constructive input, critically evaluate the information received, respond without defensiveness, and apply insights to improve. A person with strong feedback literacy demonstrates open-mindedness, reflection, and adaptability, consistently using feedback loops to enhance their development.
Copilot	Feedback literacy is the ability to effectively understand, interpret, and utilize feedback. It involves receiving feedback with an open mind, analyzing its relevance and constructive points, and implementing necessary changes for improvement. Additionally, it includes the skill to provide clear, constructive, and actionable feedback to others, fostering a culture of continuous learning and development. This competency is crucial for personal and professional growth, as it enables individuals to leverage insights from others to enhance their performance and skills.

All definitions state the skill is to understand, interpret and use feedback that is provided. Both ChatGPT 4o and o1 models identified that “seeking feedback” and “responding non-defensively to feedback” were key attributes of feedback literacy. In contrast, Copilot described “providing feedback to others” as a key attribute. Whilst all three definitions are examples of how feedback literacy could be defined; the authors argue that a more accurate definition combines the ability to seek feedback, respond to feedback constructively and provide feedback. Accordingly, the definition of feedback literacy proposed by Carless and Boud (2018), which entails “*the ability to appreciate feedback, making judgments, manage affect and take action*” was adopted. Whilst there were some differences between the various descriptions, the themes were consistent and should not impact the mapping process.

All platforms agree that the definition of self-agency involves self-motivation, self-direction and ability to take responsibility for one's actions, as per Table 4. They also associate self-agency with personal and professional growth, goal setting and autonomy. However, there were differences in the emphasis on understanding values and motivations (ChatGPT o1) and acting in alignment with one's beliefs despite external pressures (ChatGPT 4o).

Table 4: GenAI models interpretation of self-agency

Tool	Skill definition
ChatGPT 4o	Self-agency as a skill is the ability to intentionally direct and take responsibility for one's own thoughts, actions, and decisions in pursuit of goals. It involves initiative, self-motivation, autonomy, and the capacity to act in alignment with one's values and beliefs, even in the face of external pressures or uncertainty. Key components: Goal-setting and follow-through, Ownership of choices and outcomes, Initiative and proactive behavior, Adaptability and problem-solving, Confidence in personal efficacy
ChatGPT o1	Self-agency, when considered as a skill, is the ability to recognize and exercise one's own power to make choices, take action, and shape personal or professional outcomes. It involves understanding one's goals, values, and motivations, and intentionally directing effort and resources toward them. A person skilled in self-agency typically demonstrates self-awareness, self-initiation, and a willingness to accept responsibility for the consequences of their actions.
Copilot	Self-agency refers to the capacity of individuals to act independently and make their own free choices. It involves self-motivation, self-direction, and the ability to take initiative and responsibility for one's actions. This skill is crucial for personal and professional growth, as it empowers individuals to navigate challenges, pursue goals, and adapt to changing circumstances with confidence and autonomy.

Mapping of Feedback Literacy

Using the prompt outlined in the Method section, all three GenAI models produced an assessment of the alignment of feedback-literacy against each of the 35 COMPASS sub-descriptors. Table 5 contains the alignment of feedback literacy as assessed by the GenAI models and the author's previous work. In Table 5, CM is the consensus mapping of the authors, 4o is ChatGPT model 4o, o1 is the ChatGPT reasoning model and Co is Copilot. The shaded cells represent misalignment between GenAI and CM.

Table 5: Feedback literacy mapping to the COMPASS framework

Skill	CM	4o	o1	Co	Skill	CM	4o	o1	Co	Skill	CM	4o	o1	Co
1.1	✓	✓	✓	–	5.4	–	✓	✓	–	8.2	✓	✓	✓✓	✓
1.2	✓	–	–	–	5.5	–	–	✓	–	8.3	✓	✓	✓	–
2.1	✓	✓	✓	–	6.1	–	–	–	–	9.1	–	✓	✓	–
2.2	✓	✓	✓	–	6.2	✓	✓	–	–	9.2	–	✓	✓	–
2.3	✓	✓	✓	–	6.3	–	✓	✓	–	9.3	–	✓	✓	–
3.1	✓	✓	✓	–	7.1	✓✓	✓✓	✓	✓✓	9.4	✓	✓	✓	–
3.2	–	✓	✓✓	–	7.2	✓✓	✓✓	✓✓	✓✓	9.5	–	–	✓	–
4.1	–	✓	✓✓	–	7.3	✓✓	✓	✓✓	✓	9.6	✓✓	✓	–	–
4.2	–	✓	✓✓	–	7.4	✓✓	✓✓	✓	✓	10.1	–	✓✓	✓	✓
5.1	✓	✓	✓	–	7.5	✓✓	✓	✓✓	–	10.2	–	✓	✓✓	–
5.2	✓	✓	✓	–	7.6	✓	✓✓	✓	–	10.3	–	✓	✓	✓
5.3	✓	✓	–	–	8.1	✓	✓✓	✓✓	✓					

Out of the 35 sub-descriptors, no clear pattern of alignment was observed. Eight showed critical misalignment. Among the remaining sub-descriptors, the authors rated 6 as having strong alignment, 15 with partial alignment and 14 with no alignment. Only one sub-descriptor (7.2) demonstrated full alignment across all four mappers.

ChatGPT 4o demonstrated either partial or strong alignment with 31 out of 35 sub-descriptors, including 6 with strong alignment of 10.1, 8.1, 7.6, 7.4, 7.2 and 7.1 sub-descriptors. The four sub-descriptors showing no alignment were 1.2, 5.5, 6.1 and 9.5. ChatGPT 4o matched the authors' consensus mapping for 18/35 sub-descriptors, in which four were assessed at a lower alignment level and 13 at a higher level. ChatGPT o1 showed partial or strong alignment with 30 out of 35 sub-descriptors, including 9 with strong alignment (3.2, 4.1, 4.2, 7.2, 7.3, 7.5, 8.1, 8.2 and 10.2). The five sub-descriptors with no alignment were 1.2, 5.3, 6.1, 6.2 and 9.6. ChatGPT o1 aligned with the authors' consensus on 14 out of 35 sub-descriptors where 6 were assessed at a lower alignment level and 15 were assessed at a higher level than the authors' ratings. Copilot documented either partial or strong alignment with 8 out of 35 sub-descriptors including 2 showing strong alignment (7.1 and 7.2). The 6 that were partially aligned were 7.3, 7.4, 8.1, 8.2, 10.1 and 10.3. Copilot aligned with the authors' consensus on 16 out of 35 sub-descriptors where 9 were assessed at a lower alignment level and 2 were assessed at a higher level. Table 6 shows a summary of the alignment.

Table 6: Summary of alignment with authors mapping for feedback-literacy

Tool	Strong alignment	Partial alignment	No alignment	Matches with authors	Lower than authors	Higher than authors
ChatGPT 4o	6	25	4	18 / 35	4	13
ChatGPT o1	9	21	5	14 / 35	5	16
Copilot	2	6	27	16 / 35	17	2

Overall, Copilot would be the preferred tool for this mapping task. As human oversight would always be required for this task, it would be easier to assess if one of the sub-descriptors is missing rather than added. When examining the justification for alignment, there were differences as shown in Table 7.

Table 7: Justification of alignment mapping for feedback-literacy for the three models

Code	Explanation 4o		Explanation o1
3.2	Research methods can be refined based on feedback and peer review.		Effective communication with engineers and non-engineers requires both giving and receiving feedback, making feedback literacy a fundamental part of communication clarity.
4.1	Feedback helps in recognizing tool strengths and limitations during implementation.		Teamwork and leadership rely heavily on continuous feedback to coordinate efforts, motivate, resolve conflict, and adapt leadership styles—making feedback literacy central here.
4.2	Feedback supports awareness of modern tool limitations and their effective selection.		Reflection on learning and performance is at the heart of feedback literacy. Applying feedback to personal development is a direct expression of this skill.
7.5	Self-directed work benefits from feedback to improve independent outcomes and gauge performance.		Leading teams demands coaching, mentoring, and performance reviews; all are grounded in proficient feedback exchange.
7.6	Collaborating across disciplines demands feedback exchange to align understanding and expectations.		Self-directed work is enhanced by feedback from peers or supervisors for quality checks, but one can be somewhat independent.
9.6	Leadership of projects involves responding to team and stakeholder feedback to adjust plans effectively.		Economic decisions typically rest on market data, cost analysis, and ROI—feedback literacy is relatively peripheral here.
10.1	ChatGPT 4o	ChatGPT o1	Copilot
	Lifelong learning is driven by engaging with feedback to grow and adapt.	Leading projects profitably often involves responding to feedback (team morale, deliverables, performance metrics).	Feedback literacy supports lifelong learning by enabling individuals to use feedback for continuous improvement.
10.2	Adapting to new technologies can be supported by feedback but not inherently dependant on it.		Lifelong learning is practically driven by seeking out and applying feedback for continuous improvement—this is a textbook application of feedback literacy.

Across the eight sub-descriptors, the explanations provided by ChatGPT4o, ChatGPT o1, and Copilot reveal distinct interpretive tendencies rather than clear-cut agreements or disagreements. There is a large variation in how each model interpreted the mapping and explained the justification. ChatGPT 4o tends to frame feedback literacy in more technical or task specific terms, often linking it to research refinement, tool selection or independent work. In contrast, the reasoning model generally emphasises the interpersonal and developmental aspects of feedback, such as its role in communication, leadership, reflection and personal growth. These differences suggest that 4o often interprets descriptors with a practical, implementation-focused lens, while the reasoning model adopts a broader, more human-centered perspective. In contrast, Copilot only provide an explanation for one descriptor in this set. Its explanation aligned with the reasoning model and 4o in viewing feedback as central to lifelong learning. Although the authors agree with the reasoning given by the AI models, we disagree with the mapping as the AI models fail to acknowledge whether the skill is essential to the competency itself, or it is a skill that supports the efficacy of said competency.

Mapping of Self-Agency

All GenAI models generated assessments of the relevance of self-agency to each of the 35 COMPASS sub-descriptors. Table 8 presents a comparison of the assessments generated by GenAI alongside the authors' prior evaluations, highlighting how each model interprets the role of feedback literacy across the framework. In Table 8, CM is the consensus mapping of the authors, 4o is ChatGPT model 4o, o1 is the ChatGPT reasoning model and Co is Copilot. The shaded cells represent misalignment between GenAI and CM.

Table 8: Self-agency mapping to the COMPASS framework

Skill	CM	4o	o1	Co	Skill	CM	4o	o1	Co	Skill	CM	4o	o1	Co
1.1	✓✓	✓	✓	✓	5.4	✓	✓	✓	✓	8.2	✓	✓	✓✓	✓
1.2	✓	✓	–	✓	5.5	–	–	✓	✓	8.3	✓	✓	✓	✓
2.1	✓	✓	✓	✓	6.1	✓✓	–	–	✓	9.1	✓	✓	✓	✓
2.2	✓	✓	✓	✓	6.2	✓✓	✓	–	✓	9.2	✓	✓	✓	✓
2.3	✓✓	✓	✓	✓	6.3	✓✓	✓	✓	✓✓	9.3	✓	✓	✓	✓
3.1	✓	✓	✓	✓	7.1	✓✓	✓✓	✓	✓	9.4	✓	✓✓	✓	✓
3.2	✓	✓	✓✓	✓	7.2	✓	✓	✓✓	✓	9.5	✓	–	✓	✓✓
4.1	✓	✓	✓✓	✓	7.3	✓	✓	✓✓	✓	9.6	✓✓	✓✓	–	✓✓
4.2	✓	✓	✓✓	✓	7.4	✓✓	✓✓	✓	✓✓	10.1	✓✓	✓✓	✓	✓✓
5.1	✓	✓	✓	✓	7.5	✓✓	✓✓	✓✓	✓✓	10.2	✓✓	✓	✓✓	✓
5.2	✓	✓	✓	✓	7.6	✓	✓	✓	✓	10.3	✓✓	✓	✓	✓
5.3	✓	✓	–	✓	8.1	✓	✓	✓✓	✓					

From the 35 sub-descriptors, no consistent pattern of interpretation was observed. Two sub-descriptors (6.1 & 6.2) were identified as showing critical divergence in how self-agency was understood and will be discussed in a later section. Of the remaining sub-descriptor, the authors rated 12 as showing strong alignment with self-agency, 22 as partial and one as showing no alignment. Twelve sub-descriptors demonstrated full alignment across all three GenAI evaluations and the authors' mapping.

ChatGPT4o documented either partial or strong alignment with 33 of the 35 sub-descriptors, including 6 assessed as strong and 2 as showing no alignment. The remaining 27 were assessed as partially aligned. ChatGPT 4o matched the authors' consensus on 26 of the 35 sub-descriptors, rated 8 at a lower alignment level, and 1 at a higher level. ChatGPT's o1 identified strong or partial alignment for all but 5 descriptors, rating 8 as strong, 22 as partial and 5 as having no alignment. This matched the authors' evaluations on 15 sub-descriptors, rated 12 lower and 8 higher than the authors. Copilot indicated partial or strong alignment for all 35 sub-descriptors, including 6 rated as strong and 29 as partial. It aligned the authors' consensus mapping on 26 out of 35 sub-descriptors, rated 7 at a lower and 2 at a higher level. Table 9 shows a summary between the authors' and GenAI's mapping of self-agency.

Table 9: Summary of alignment with authors mapping for self-agency

Tool	Strong Alignment	Partial Alignment	No Alignment	Matches with Authors	Lower than Authors	Higher than Authors
ChatGPT 4o	6	27	2	26 / 35	8	1
ChatGPTo1	8	22	5	15 / 35	12	8
Copilot	6	29	0	26 / 35	7	2

Overall, Copilot and ChatGPT 4o demonstrated minimal differences in their outputs compared to the authors' and each other. In contrast, the reasoning model from ChatGPT o1 was significantly different. Either Copilot or ChatGPT 4o would be the preference in models. Given that human oversight would always be required for this task, it would be easier to assess if one of the descriptors was missing, rather than removing.

Sub-descriptor 9.5 demonstrated critical divergence between the three GenAI models. While all models acknowledged some role for self-agency, they differed in the emphasis placed on its relevance. ChatGPT 4o and the reasoning model focused on initiative and accountability in managing projects, while Copilot framed self-agency as essential for leading multidisciplinary efforts. These cases highlight how the models diverge in interpreting complex and leadership-oriented descriptors, particularly when balancing procedural thinking with personal autonomy. The justifications are provided in Table 10.

Table 10: Justification of alignment mapping for self-agency for the three models

Code	ChatGPT 4o	ChatGPT o1	Copilot
9.5	Economic reasoning is data-driven and procedural, not strongly tied to agency.	Economic analysis relies more on systematic methods and data than on individual initiative.	Self-agency is fundamental to leading projects effectively.

Model Performance

Across both the feedback literacy and self-agency mappings, Copilot showed the highest level of alignment with the author's consensus mapping. It outperformed both ChatGPT models in mapping feedback-literacy and surpassed ChatGPT o1 in self-agency evaluation. However, Copilot showed a lack of specificity in its feedback and justification, which has been a common limitation of Copilot outputs when using institutional models. In contrast, ChatGPT o1 demonstrated more frequent divergence in judgment compared to the other models, particularly in interpreting complex and leadership-oriented descriptors. These findings highlight the importance of refining GenAI models to improve their accuracy and reliability in nuanced mapping tasks. Despite these differences, all models demonstrated a capacity to assist in thematic mapping. While AI tools such as ChatGPT and Copilot offer substantial support in initial analysis and processing a large volume of material, their varied interpretations of context-sensitive descriptors require human intervention. Given their ability provide compelling reasoning behind their rating, it is important to critically analyse their outputs as they can sometimes be misleading.

Potential Applications

AI-assisted mapping offers institutions the potential to make curriculum review and accreditation more efficient by rapidly processing large volumes of data. This could reduce the administrative burden on faculty and accreditation teams, allowing them to focus on higher-level decision-making rather than labour-intensive data compilation. Additionally, AI-driven mapping can also support educational practice by helping instructors to align course content and assessments with intended learning outcomes and standards. These potential applications inform the practical relevance of the findings. However, careful implementation and expert validation are necessary. AI systems are only as reliable as the data and parameters they are given, and precautions must be taken when handling sensitive data. To ensure that AI enhances efficiency without compromising data privacy and accuracy.

Conclusions

This study demonstrates that GenAI can play a transformative role in mapping skills such as feedback literacy and self-agency to professional frameworks like COMPASS. While ChatGPT 4o and Copilot produced the most consistent outputs, the divergence observed in the reasoning model reveals expert judgement remains essential. GenAI is most valuable for supporting large-scale or preliminary analysis when outputs are critically reviewed, validated by experts and transparently documented. When applied responsibly, GenAI can complement professional expertise and streamline the mapping processes.

References

- Ajjawi, R., Tai, J., Dollinger, M., Dawson, P., Boud, D., & Bearman, M. (2024). From authentic assessment to authenticity in assessment: Broadening perspectives. *Assessment & Evaluation in Higher Education*, 49(4), 499–510. <https://doi.org/10.1080/02602938.2023.2271193>
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307–337). Information Age Publishing. <https://search.gesis.org/publication/zis-Bandura2006Guide>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Carew, A., Doe, P., Hadgraft, R., Symes, M., & Henderson, A. (2013). *Engineering curriculum structure and mapping: Accreditation and beyond*. CQUniversity. <https://hdl.handle.net/10018/1017363>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Crosthwaite, C. (2021). *Engineering Futures 2035: Engineering education programs, priorities & pedagogies*. Australian Council of Engineering Deans. <https://www.aced.edu.au/downloads/Engineering%20Futures%202035%20R2%20report%20to%20ACED.pdf>
- Crosthwaite, C., Hargreaves, D., Wilson, J., Lee, P., Foley, B., Burnett, I., Goldfinch, T., & Symes, M. (2018). *Engineering futures 2035*. In *Proceedings of the 29th Australasian Association for Engineering Education Conference (AAEE 2018)* (pp. 668–669). Engineers Australia. <https://search.informit.org/doi/10.3316/informit.199769789675749>
- Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*, 51(12), 2629–2633. <https://doi.org/10.1007/s10439-023-03272-4>
- Hagendorff, T., Fabi, S., & Kosinski, M. (2023). Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nature Computational Science*, 3(10), 833–838. <https://doi.org/10.1038/s43588-023-00527-x>
- Lidfors Lindqvist, A., Quince, Z., & Chen, P. (2025). *Aligning feedback literacy and self-agency to professional engineering competency standards* (Accepted manuscript).
- Lindqvist, A. L., Willey, K., Lidfors, L., & Francis, B. (2023). *Formative sprints to improve feedback, learning, and fidelity in practice-based activities*. In *Proceedings of the 34th Australasian Association for Engineering Education Conference (AAEE 2023): Adapting to the Changing Expectations of Students and Industry*. Engineers Australia. <https://search.informit.org/doi/10.3316/informit.T2024110400009191489893623>
- Lindqvist, A. L., Willey, K., Lidfors, L., & Francis, B. (2024). *Advancing fidelity and feedback in practical activities with formative sprints*. In *Proceedings of the 35th Annual Conference of the Australasian Association for Engineering Education (AAEE 2024)*. Engineers Australia. <https://search.informit.org/doi/10.3316/informit.T2025032400009491006416357>
- Liu, J., Wang, C., Liu, Z., Gao, M., Xu, Y., Chen, J., & Cheng, Y. (2024). A bibliometric analysis of generative AI in education: Current status and development. *Asia Pacific Journal of Education*, 44(1), 156–175. <https://doi.org/10.1080/02188791.2024.2305170>
- Male, S. A., Bush, M. B., & Chapman, E. S. (2011). An Australian study of generic competencies required by engineers. *European Journal of Engineering Education*, 36(2), 151–163. <https://doi.org/10.1080/03043797.2011.569703>
- Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*, 45(4), 527–540. <https://doi.org/10.1080/02602938.2019.1667955>
- O'Connor, C., & Joffe, H. (2020). Inter-coder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1609406919899220. <https://doi.org/10.1177/1609406919899220>
- Oyeniya, J. (2024). Unveiling the cognitive capacity of ChatGPT: Assessing its human-like reasoning abilities. *International Research Journal of Modernization in Engineering Technology and Science*, 6(4), 2582–5208. <https://doi.org/10.56726/IRJMETSS52428>
- Quince, Z., Hills, C., & Maxwell, A. (2023). *Are engineering programs meeting the 2035 professional and personal requirements?* In *Proceedings of the 34th Australasian Association for Engineering Education Conference (AAEE 2023): Adapting to the Changing Expectations of Students and Industry*. Engineers Australia. <https://search.informit.org/doi/10.3316/informit.T2024110400009692147447182>
- Quince, Z., Hills, C., & Maxwell, A. (2023). *Are engineering programs meeting the 2035 professional and personal requirements?* In *Proceedings of the 34th Australasian Association for Engineering*

- Education Conference (AAEE 2023): Adapting to the Changing Expectations of Students and Industry.* Engineers Australia.
<https://search.informit.org/doi/10.3316/informit.T2024110400009692147447182>
- Quince, Z., & Nikolic, S. (2025). Student identification of the social, economic and environmental implications of using generative artificial intelligence (GenAI): Identifying student ethical awareness of ChatGPT from a scaffolded multi-stage assessment. *European Journal of Engineering Education*, 1–20. <https://doi.org/10.1080/03043797.2025.2482830>
- Quince, Z., Lidfors Lindqvist, A., & Chen, P. (2025). Preparing for the future: Aligning engineering competency standards. In *Proceedings of the 21st International CDIO Conference, Monash University, Melbourne, Australia, June 3–6, 2025*. CDIO Knowledge Library.
- Rehman, U., Iqbal, F., & Shah, M. U. (2025). Exploring differences in ethical decision-making processes between humans and ChatGPT-3 model: A study of trade-offs. *AI and Ethics*, 5(1), 279–289. <https://doi.org/10.1007/s43681-023-00335-z>
- Sumsion, J., & Goodfellow, J. (2004). Identifying generic skills through curriculum mapping: A critical evaluation. *Higher Education Research & Development*, 23(3), 329–346. <https://doi.org/10.1080/0729436042000235436>
- Zaki, N., Turaev, S., Shuaib, K., Krishnan, A., & Mohamed, E. (2023). Automating the mapping of course learning outcomes to program learning outcomes using natural language processing for accurate educational program evaluation. *Education and Information Technologies*, 28(12), 16723–16742. <https://doi.org/10.1007/s10639-023-11877-4>

Copyright Statement

Copyright © 2025 Authors' names Zachery Quince, Anna Lidfors Lindqvist, Patcharin Chen: The authors assign to the Research in Engineering Education Network (REEN) and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction, provided that the article is used in full, and this copyright statement is reproduced. The authors also grant a non-exclusive license to REEN to publish this document in full on the World Wide Web (prime sites and mirrors), on portable media and in printed form within the REES and RCEE & RHEd 2025 proceedings. Any other usage is prohibited without the express permission of the authors.