

From AI Consumption to Co-Thinking: How Structured Evaluation Enhances Year 9

Girls' Critical Thinking in Science Inquiry

Jo Oreo

Ruyton Girls' School, Victoria, Australia

Abstract

This action research examines how structured evaluation of AI-generated outputs shapes girls' critical thinking. This study was motivated by concerns that generative artificial intelligence (GAI) may encourage surface-level learning and cognitive offloading when girls use it primarily for quick responses, particularly in the absence of explicit scaffolding for critical and ethical engagement. A purpose-designed co-thinking framework was implemented within a Year 9 Science inquiry unit in an independent girls' school to support girls to "co-think" with AI while maintaining ownership of their ideas. Data from student portfolios, surveys, classroom observations, and interviews indicated a shift from using AI for quick answers toward more deliberate co-thinking, with greater attention to accuracy, bias, and limitations, alongside increased cross-checking, prompt revision, and clearer differentiation between students' ideas and AI contributions. Overall, the findings indicate that the impact of GAI depends less on the technology itself and more on pedagogical design. When evaluation is explicitly scaffolded, students engage more critically with AI, using it to test and refine ideas rather than accept outputs at face value. Student feedback from this research informed the refinement of this framework, now supporting broader whole-school approaches to critical thinking and ethical engagement with GAI.

Glossary

Agency (with AI): The capacity for learners to make intentional, critically informed decisions about their thinking and AI use, including when, how, and whether to engage with AI, and how to act on or challenge its outputs.

Cognitive offloading: The reduction of mental effort by outsourcing thinking processes to an external tool, such as AI.

Co-thinking / thinking with AI: A structured approach where human judgement remains central and AI functions as a cognitive partner. When using GAI, learners analyse, question, and refine ideas using evidence and reasoning, while retaining responsibility for evaluating the accuracy, relevance, and bias.

Evaluative Judgement: The ability to make informed decisions about the quality, credibility, and usefulness of ideas, including AI-generated content.

Generative Artificial Intelligence (GAI): Digital tools that generate text, images, or other content in response to user prompts, based on patterns learned from large datasets.

Prompt Revision / Re-Prompting: An iterative process in which students modify their original AI prompt in response to their evaluation of an initial output.

SPARK framework: A five-stage structured co-thinking framework designed to scaffold students' critical and evaluative judgement when working with AI-generated outputs. **S**tart with what you know, **P**rompt with purpose, **A**ssess for accuracy and bias, **R**efine and reference, **K**now your voice (see Appendix A).

From AI Consumption to Co-Thinking: How Structured Evaluation Enhances Year 9

Girls' Critical Thinking in Science Inquiry

The rapid emergence of Generative Artificial Intelligence (GAI) in education has created both opportunity and concern, particularly in relation to student thinking, authorship, and agency. Early responses to GAI in schools have often centred on tools—which platforms are permitted, which chatbots are blocked, and how outputs might be detected—rather than how students can be supported to think ethically, critically, and with integrity when working alongside AI. These tensions are especially visible in inquiry-based subjects such as science, where evaluating evidence, synthesising ideas, and developing concepts are central to learning.

Rather than focussing on what AI platforms are accepted or restricted, my interest stemmed from a broader pedagogical concern: how to support girls to think with AI. I became interested in how girls might be explicitly supported to evaluate AI-generated outputs and retain ownership of their ideas, reflecting a reframing of AI in learning away from tools and platforms and toward the thinking skills girls need regardless of what technologies are permitted. As girls continue to access AI beyond the classroom, it is critical that they question information, recognise bias, and apply judgement ethically.

Early observations and conversations with Year 9 girls and colleagues highlighted uncertainty about how GAI should be used in learning tasks. Many girls were engaged, yet hesitant to challenge AI-generated responses that appeared fluent or authoritative. Without explicit scaffolding, girls often defaulted to copying or lightly paraphrasing AI outputs, limiting opportunities for critical thinking and diminishing ownership of ideas. These observations prompted the central research question: *How does student evaluation of AI-generated outputs influence critical thinking during Year 9 girls' science inquiry activities?*

This study used action research because it enabled iterative cycles of planning, action, observation, and reflection within a real classroom context, allowing me to refine practice as the inquiry unfolded. Early attempts to incorporate GAI relied on loosely

structured prompts and stand-alone activities, which did not consistently deepen thinking. This led to a deliberate shift toward thinking with, and critically evaluating, AI. Over time, the project evolved toward a more intentional design approach. Drawing on research related to critical thinking, metacognition, and ethical technology use, I developed the SPARK framework (see Appendix A) as a shared language for co-thinking with AI. SPARK evolved from discrete “thinking tools” embedded within the inquiry portfolio into an overarching framework that shaped lesson design, classroom routines, and student reflection. The framework was implemented through seven inquiry phases (see Appendix B) and a one-page Critical Thinking Toolbox (see Appendix C) and has since informed broader inquiry planning across the school. Throughout the inquiry, girls used routines to make their thinking and evaluation visible (see Appendix D), providing a basis for analysing changes in critical thinking.

Literature Review

Critical thinking has consistently been recognised as a core competency for lifelong learning and workforce readiness (OECD, 2025; OECD; Sun et al., 2025). In the context of GAI, this competency is becoming increasingly important, as these tools create new opportunities for feedback, inquiry, and critical engagement, while also raising concerns about over-reliance, accuracy, and academic integrity (Abdelghani et al., 2023; Mulder et al., 2025; Oliveira et al., 2025).

However, while GAI tools are increasingly available, the extent to which they strengthen or undermine students’ reasoning is still under debate. As Gerlich (2025) cautions, “the pervasive availability of AI tools, which offer quick solutions and readymade information, can discourage users from engaging in the cognitive processes essential for critical thinking” (p.2). This tension is reflected in patterns observed in my Year 9 Science classroom, where students are drawn to AI because it is fast and polished, yet can discourage the slower, more thoughtful analysis that science inquiry requires.

Critical thinking is commonly defined as the ability to apply, analyse, evaluate, synthesise, and judge information for its reliability, relevance and logic (ACARA, 2024). Gerlich (2025) emphasises that critical thinking is a multifaceted cognitive process that supports problem solving, informed decision-making, knowledge building, and ultimately academic performance. In science, critical thinking enables students not only to recall scientific knowledge, but to weigh up the evidence that supports it. However, when collaborating with GAI tools, students must recognise AI limitations, such as built-in assumptions, biases, or hallucinations (Bianchi, 2024) that can compromise their evaluation of such evidence. Without scaffolds that prompt students to pause and evaluate, many may skim AI outputs and engage superficially (Oliveira et al., 2025; King, 2025). This risk intensifies when AI is treated as an answer-giver rather than a thinking partner (Abdelghani et al., 2023; Gerlich, 2025; King, 2025; Paulson, 2024).

The structure of traditional assessments can further amplify the risks of AI use (Mulder et al., 2025). When emphasis falls on final products, the implicit message students receive is that flawless results matter most. The Girls' Day School Trust (2023) notes that girls, in particular, are vulnerable to this pressure, with performance-based self-esteem driving perfectionism. These perfectionistic tendencies make the neat, polished outputs of AI especially appealing, encouraging over-reliance. Assessments that reward the final product rather than the learning process can reinforce these unproductive habits (Mulder et al., 2025; Pearce & Chiavaroli, 2023). This concern is also reflected in wider educational articles, as The Atlantic (2024) observes:

It's not about whether AI helped write an essay, but about how students directed the AI, how they explained their thought process, and how they refined their approach based on AI feedback. These metacognitive skills are becoming the new metrics of learning. (para. 14)

Assessments, therefore, need to value evaluation, reflection, and revision.

To counter perfectionistic pressures, students need routines and strategies that help them track their thinking and treat AI as a thinking partner (Abdelghani et al., 2023; Bianchi, 2024). Worked examples can model how to check AI responses for reliability, logic, and assumptions. Paulson (2024) uses three analogies—the gardener’s tree, the navigator’s map and the sculptor’s stone—to show how students generate, evaluate, and refine ideas. Together, these align with Bloom’s higher-order thinking skills and offer practical ways to scaffold evaluation in science inquiry (ACARA, 2024; Paulson, 2024).

Research on inquiry and constructivist learning highlights that authentic assessment must capture process as well as product. Inquiry-based approaches, such as hands-on modelling and action-centred learning, provide opportunities for students to test and question ideas (Bronstein et al., 2024; Chang et al., 2023). This emphasis on process reflects the iterative, non-linear nature of learning, where students revise their thinking in response to new evidence (The Atlantic, 2024). When combined with GAI, this iterative inquiry enables students to evaluate explanations against their own prior knowledge or evidence, strengthening critical thinking and maintaining ownership of their ideas.

While combining inquiry with GAI offers opportunities, risks remain. Unstructured AI use can lead to cognitive offloading, with students outsourcing reasoning, while polished outputs may mask authentic learning (Gerlich, 2025; Melisa et al., 2025). AI-generated content can also reflect embedded bias, reinforcing the need for students to critically evaluate outputs (OECD, May, 2025). These risks are particularly relevant for girls, where confidence and perfectionism may reduce critical engagement (Australian Government, 2024).

The research indicates that developing AI literacy is essential. The OECD AI Literacy Framework (May, 2025), for example, reports that almost half of secondary students struggle to critically evaluate AI outputs, with only 15% having received formal education. AI Swiss (2025) and King (2025) similarly note that intentional teaching and carefully designed

learning activities are needed to help students use AI as a cognitive partner rather than a replacement.

Although research in this area is expanding, few studies specifically focus on how girls evaluate AI in the middle years of science. This makes classroom-based action research well-suited to investigating such gaps (Mertler, 2020). This study examines how structured evaluation of AI-generated outputs can strengthen students' critical thinking and evaluative judgement, while supporting girls' engagement and confidence in science and emerging technologies.

Research Context

Ruyton Girls' School is an independent girls' school in Melbourne, Australia, with approximately 950 students across Years K-12, and a strong emphasis on inquiry-based learning and student voice.

The action research was conducted over a six-week period with 18 Year 9 students in my science class and was embedded within a unit focussed on climate change and human impacts. As part of the curriculum, students investigated contemporary environmental issues and proposed an evidence-informed solution to the school's sustainability committee. Although the entire Year 9 cohort completed the inquiry task as the key assessment for the unit, data analysed for this study were drawn from my class of 18 girls.

All students were informed about the purpose of the research, and participation in interviews was voluntary. Parental and student consent was obtained in accordance with school research protocols. Student data were anonymised, and all identifiable information was removed from reporting.

The school-approved AI tool used during the project was Microsoft Copilot. Students were explicitly instructed not to upload personal data into external AI systems. AI-generated responses were printed and annotated by students and were collected for analysis.

The Action

The literature review informed the redesign of the science unit to prioritise critical thinking and make the learning process more visible, rather than emphasising a polished end-product. This was translated into practice through the SPARK co-thinking framework (Start, Prompt, Assess, Refine, Know your voice—see Appendix A), designed to support students to evaluate and refine GAI outputs. The framework drew on Paulson’s (2024) three critical thinking analogies—the Gardener’s Tree, Navigator’s Map, and Sculptor’s Stone—which emphasise idea generation, verification, and iterative refinement. These ideas were integrated with Mertler’s (2020) action-research cycle to reflect the iterative process of planning, feedback, and refinement central to co-thinking with AI.

Students engaged with an excerpt from Amelia King’s (2025) *Thinking with AI* and applied the “lift versus stairs” analogy to evaluate the role of AI in learning. This positioned AI as a tool that could either support or bypass thinking, explicitly addressing the risks of cognitive offloading. Through a whole-cohort discussion, girls explored authenticity, voice, and ethical tool use, recognising that while opportunities to reproduce ideas have always existed, generative AI intensifies the need for explicit evaluation and ownership beyond traditional assessments. These principles guided the inquiry design, structured across seven phases from individual idea generation to collaborative synthesis (see Appendix B).

In the early phases, students engaged in tech-free brainstorming followed by structured co-thinking with AI. Later phases focussed on comparing ideas, engaging in ethical discussion, and developing a group two-minute shark tank-style pitch to the school’s sustainability committee. Delivered without notes, the pitch required students to articulate their reasoning clearly, justify claims with evidence, and respond authentically. This task emphasised idea development and evaluative judgement over a polished presentation.

A structured inquiry portfolio was developed to support this process and included a one-page Critical Thinking Toolbox (see Appendix C). Girls used routines such as tech-free brainstorming, annotated GAI conversations, prompt refinement, mind-mapping, and colour coding to distinguish student thinking from AI content (See Appendix D). These routines

made thinking visible by externalising reasoning, supporting revision of AI interactions, and clarifying the boundary between AI-generated ideas and girls' own scientific thinking.

Data Collection

Qualitative and quantitative data were collected using a mixed-methods approach to examine how structured co-thinking with AI influenced Year 9 girls' critical thinking. These data included student learning portfolios, which were polyangulated with pre- and post-intervention surveys, classroom observations, and focus-group interviews to ensure reliability and depth (Mertler, 2020).

Baseline conversations with colleagues and students indicated that AI was being used but rarely discussed in terms of reasoning or evaluative judgement. These insights shaped the design of an intervention that modelled authentic, collaborative learning, aligned with 21st century skills to enhance critical thinking and ethical AI use (AI Swiss, 2025; Australian Government, 2024; OECD, May, 2025).

Each datum source served a distinct purpose: surveys captured shifts in student confidence and attitudes; student portfolios (structured through SPARK-aligned routines) visualised conceptual change; and interviews revealed deeper metacognitive insight. A whole class pre-intervention survey and small group student interviews were conducted to establish a baseline of how students used AI and their confidence in evaluating AI-generated outputs. Several survey items were adapted from Stupple's (2017) critical thinking toolkit (CriTT), and Gerlich's (2025) *AI Tools in Society*, focussing on analysis, inference, and critical thinking when using AI. Quantitative Likert scale responses provided a reference point for measuring growth, while open-ended questions offered qualitative insights into student perceptions.

Throughout the intervention, I collected annotated AI printouts, colour-coded mind maps, and written reflections showing how students evaluated, refined, and justified their ideas when co-thinking with AI. Classroom observations and field notes captured student dialogue and patterns of questioning, while recorded focus-group interviews (conducted mid-

and post-intervention) provided deeper reflection on shifts in thinking. A post-action survey, mirroring the pre-action survey, measured changes in students' self-reported confidence, reasoning, and ethical awareness when working with AI.

Data Analysis

Data analysis followed Mertler's (2020) three-phase process of organise, describe and interpret. Quantitative survey data were analysed to compare pre- and post-intervention responses, while qualitative data were coded and themes identified in terms of metacognition, reasoning and ethical AI use.

Qualitative themes were analysed alongside quantitative survey trends to identify patterns in students' perceptions and behaviours in their AI use over the inquiry task. Polyangulation across surveys, portfolios, and interviews ensured consistency of interpretation, while maintaining a reflection journal and having a colleague conduct classroom observations helped reduce bias and informed iterative adjustments to lesson and task design.

Together, these methods provided a coherent data set for examining how structured co-thinking with AI influenced students' critical thinking. Findings will inform future iterations of AI-supported inquiry tasks.

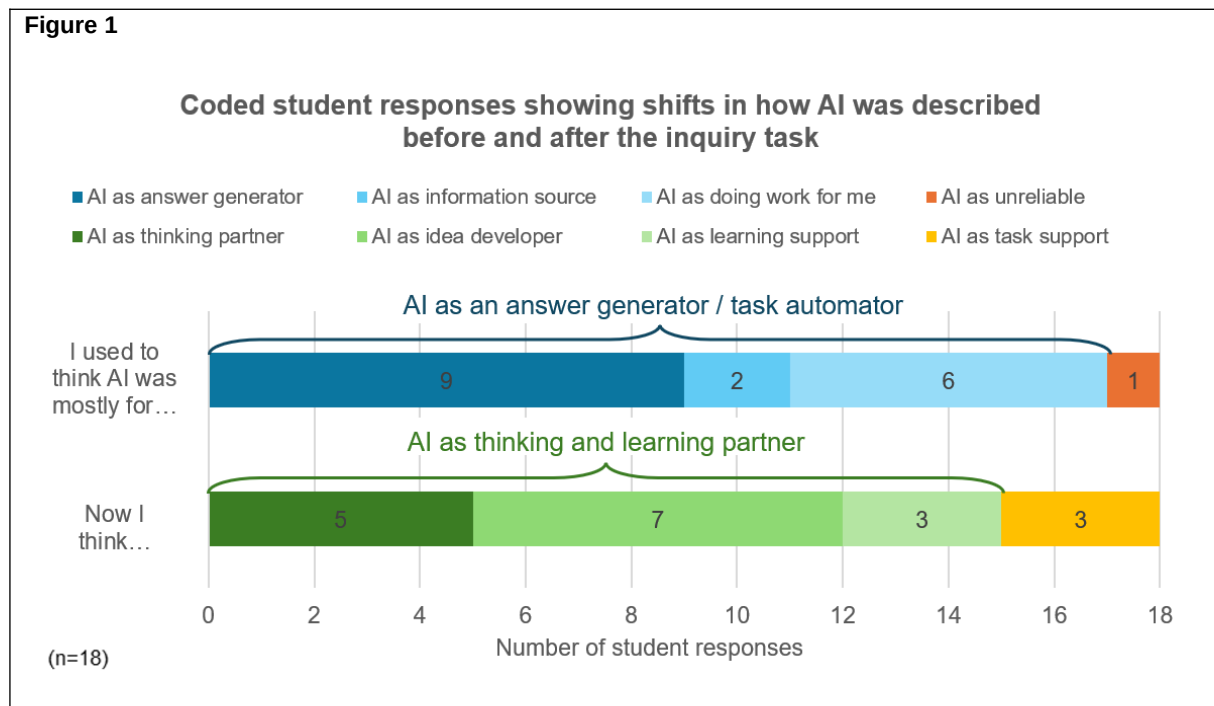
Discussion of Results

The following identified themes focus on changes in cognitive engagement, evaluative practices, and agency during the Year 9 inquiry.

Co-Thinking With AI Extends and Connects Girls' Ideas

Within the SPARK framework, AI was positioned as a co-thinking tool that supported girls to extend and connect their existing ideas, rather than generating answers. Before the inquiry task, girls primarily approached AI as a tool for consuming answers and improving efficiency rather than thinking. Pre-survey data indicate that many girls reported using AI "to explain things," "to help me understand content," or for "generating answers quickly." This aligns closely with "I used to think; now I think" responses (see *Figure 1*), indicating a clear

shift in girls' perceptions. Rather than initially viewing AI as an information source or task automator, the "Now I think" responses indicate that girls increasingly used AI to support and extend their own thinking. One student reflected, "I used to think AI was mostly there to be used as a source of information, and now I think it can help me transform my own ideas into more information whilst using my thinking."



The change in thinking was reflected in how girls worked with GAI during the inquiry task. In their portfolios, most students (15 of 18) described AI as supporting them to connect and extend existing concepts, rather than introducing entirely new content, particularly through clarification, organisation, and the connection of prior knowledge (see Appendix D). Portfolio evidence indicates that girls used AI conversations to generate new inquiry questions, extending their thinking beyond consolidation. This finding is reinforced by post-survey responses, where girls increasingly described AI as supporting them to “extend my thinking” or “generate new questions.” Many girls attributed this shift to the initial tech-free lesson, which required them to document and mind map what they already knew before engaging with AI (see Appendix D). One student noted that the tech-free lesson “allowed me to jot down what I KNEW first and not just rely on AI’s answers... I realised that I knew way

more about the topic where I was forced to use my own thinking first,” while another reflected that “the tech-free questions connected all my learning... it made me realise how this unit was all connected.” This sequence reflects the importance of activating prior knowledge before AI use, supporting girls to engage more critically with generated content rather than defaulting to AI as a primary source of ideas (OECD, May, 2025).

Post-survey responses reinforce the pattern of “pre-AI thinking,” with 12 out of 18 girls reporting that they preferred using AI as a thinking partner rather than an answer generator. Furthermore, this finding was echoed in portfolio reflections, where girls used colour coding and annotation to separate and track their ideas from AI (see Appendix D). One student reflected, “tracking my thinking using different colours helped me see what was my idea and what came from AI, which helped me stay in control of my work,” while another noted, “that I was reading the information it gave me rather than just looking at the answer, because I had to read the information to continue the conversation.” Together, these patterns suggest that evaluation was not simply about checking accuracy, but about maintaining ownership of reasoning within the inquiry process, aligning with Paulson’s (2024) emphasis on evaluative judgement as central to independent learning and decision making.

Increased confidence and persistence in thinking was evident during the inquiry. Classroom observations indicated early discomfort during the initial tech-free lesson, with some girls expressing concern about making incorrect connections. As the inquiry progressed, girls demonstrated increased engagement and confidence in articulating emerging ideas. One girl stated, “I didn’t realise I knew so much. I made lots of connections.” Rather than outsourcing difficulty to AI, girls remained cognitively engaged, using AI to extend and consolidate rather than replace their thinking.

Across portfolio and survey data, girls increasingly positioned AI as a support for extending thinking rather than a shortcut for answers. This shift reflects broader literature that suggests AI can function as a cognitive partner that deepens understanding, rather than providing answers (AI Swiss, 2025; King, 2025). SPARK’s co-thinking routines strengthened

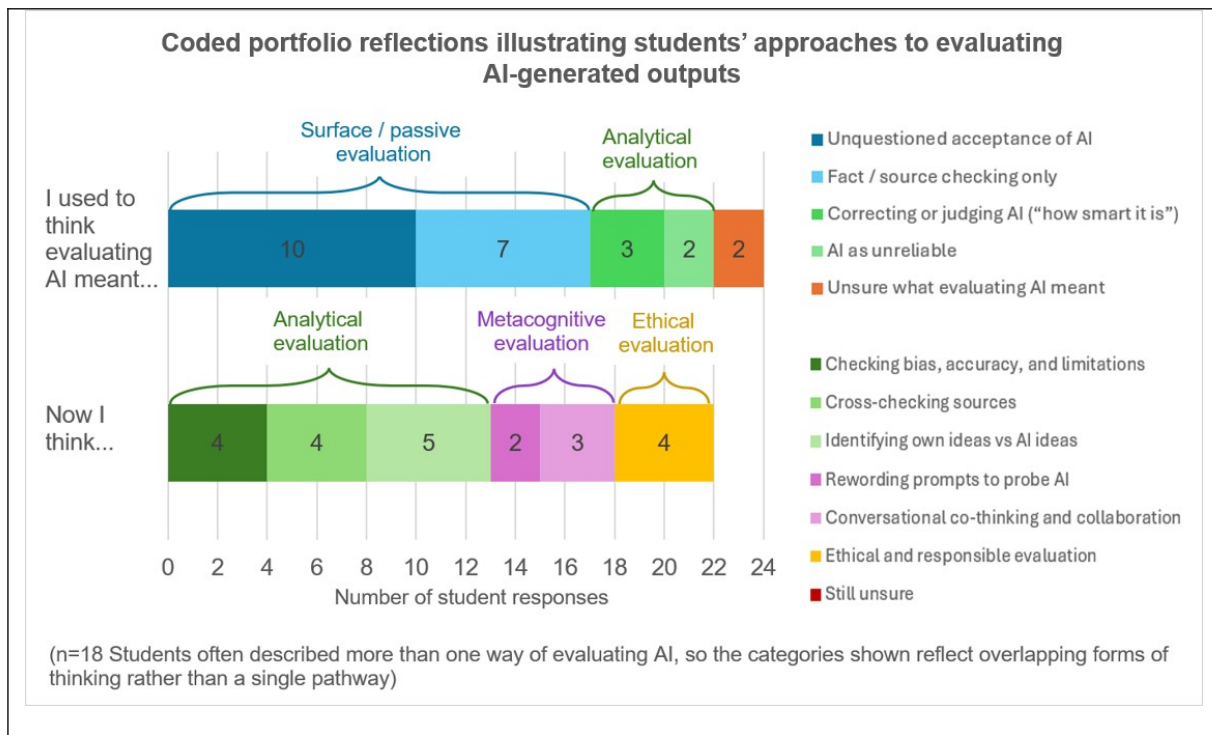
girls' cognitive engagement and agency, encouraging productive struggle as part of learning rather than relying on AI for quick answers.

Co-Thinking Routines Develop Girls' Critical Thinking

Girls' evaluation of AI-generated outputs developed gradually within the SPARK framework, becoming more deliberate and sophisticated as they engaged with structured co-thinking routines. Only six students initially described cross-referencing AI outputs or recognising limitations in AI-generated explanations. This suggests that critical evaluation was not initially evident in girls' AI use but developed through deliberate routines that slowed thinking and required students to cross-check AI outputs against other sources, encouraging reflection on accuracy, bias, relevance, and uncertainty.

Analysis of reflections and portfolios indicates that girls' understanding of evaluation broadened over time. *Figure 2* shows a clear increase in analytical, metacognitive, and ethical evaluation, indicating that students moved beyond basic fact-checking toward deeper interrogation of AI-generated knowledge. For example, while annotating her own AI conversation, one student reflected: "I tried to affirm this, but I am not able to at first! I just started to question this response. Yet, I realised that AI is able to form varied opinions by synthesising information." This reflection illustrates a shift from checking accuracy to questioning how AI generates and synthesises explanations.

Figure 2

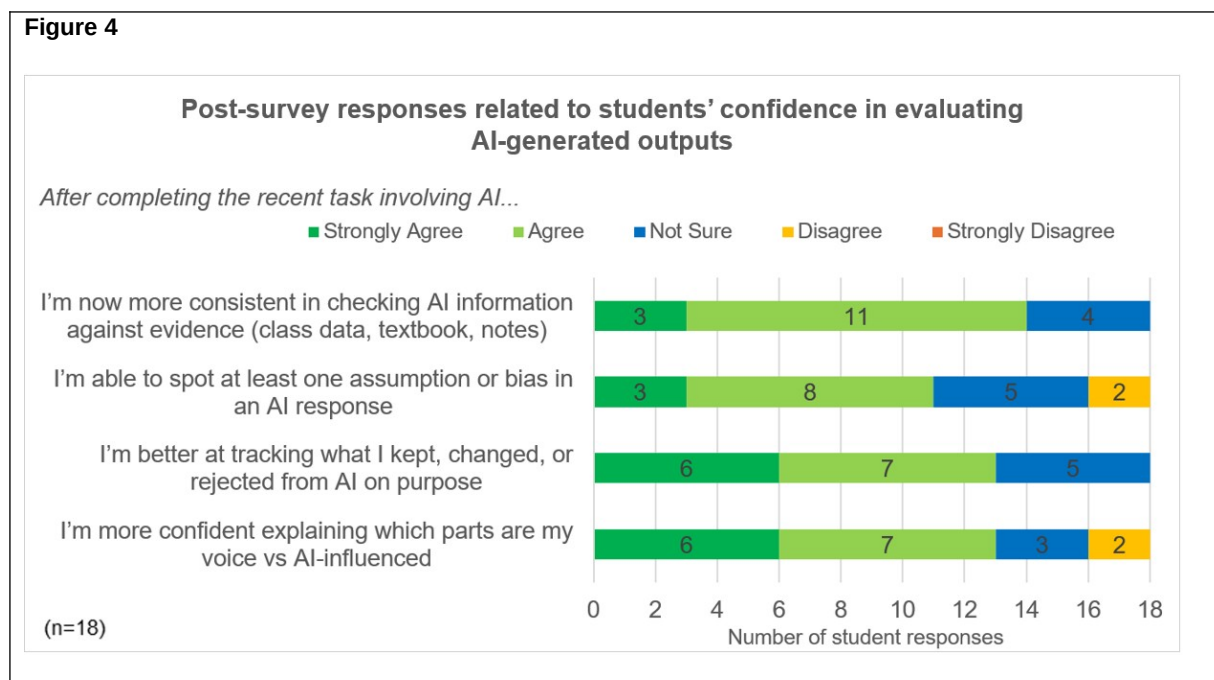
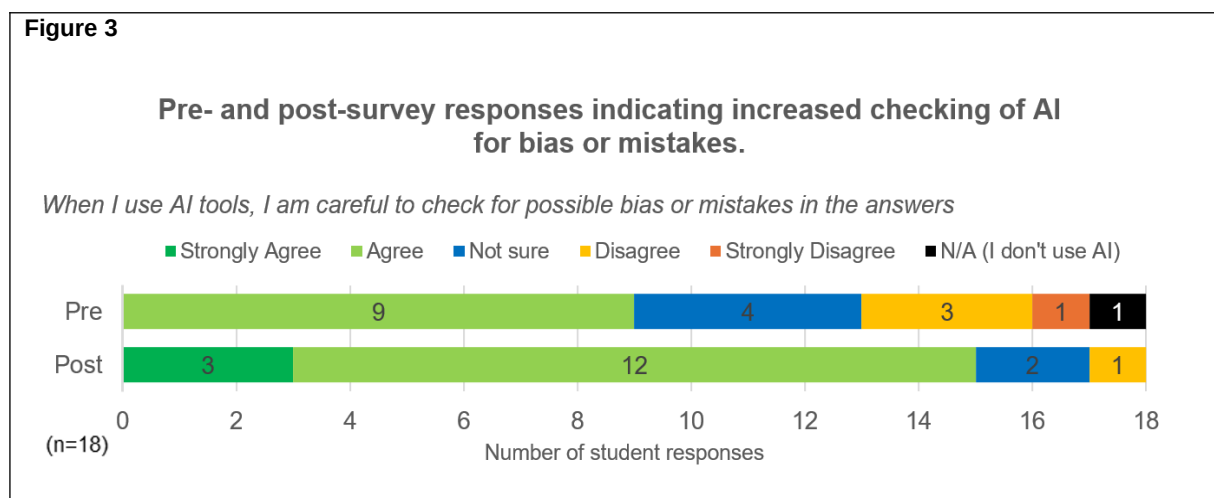


Post-survey responses indicate that girls increasingly described evaluation as an intentional process rather than automatic. Rather than accepting AI outputs at face value, students reported pausing to question relevance, accuracy, and perspective before deciding whether to incorporate or reject ideas. This shift was visible in both survey responses and annotated portfolios, particularly during the structured tech-free reflections following AI conversations, where evaluation became embedded within the inquiry process.

Through engagement with the Assess and Refine/Reference stages of the SPARK framework, including the cross-referencing routines and research validation table, girls increasingly interrogated AI outputs, which enabled them to identify missing perspectives and recognise limitations in AI-generated explanations. Several students noted dominant global or US-centric perspectives in AI outputs or overly negative representations of human impact, consistent with research highlighting embedded bias in AI-generated content (Encinas-Martin, 2025; OECD, May, 2025). One student commented, “A lot of sources were from the US, so the information may reflect their national ideas rather than a global view.” Another student reflected, “The AI feedback expanded my thinking but I rejected some statistics

because through further research I realised they were inaccurate.” These judgements shaped the direction of their inquiry, not just the accuracy of their notes.

Girls’ evaluative judgement is evident in post-survey data. *Figure 3* shows an increase in students actively checking AI outputs for bias or mistakes (from 9 to 15), while *Figure 4* indicates increased confidence in evaluating AI responses, including identifying bias, tracking decision-making, and distinguishing between their own thinking and AI-influenced ideas. This increased confidence in evaluative judgement is associated with greater student voice and engagement (Bray et al., 2020).



Classroom observations aligned with this interpretation of more deliberate and critical evaluation of AI outputs, with evaluative talk often occurring when girls were guided through co-thinking routines that focussed on reliability, relevance, and uncertainty. Interview reflections further reinforced the shift, with students describing evaluation as judging the usefulness of AI outputs rather than simply checking for correctness. One student explained that without providing context, “you’re just going to get irrelevant information or information that’s too convoluted for you to understand,” illustrating that evaluative judgement developed through structured co-thinking routines rather than existing as a default AI literacy skill.

Co-Thinking Routines Develop Girls’ Agency

Beyond evaluating AI outputs, co-thinking routines shaped girls’ sense of agency by supporting intentional decisions about when, how, and if to engage with AI in their learning. As students became more deliberate in their evaluation, their critical thinking deepened, leading them to exercise agency over their thinking and tool use. AI was positioned over time not as a source of answers, but as a tool to support, extend or challenge their ideas.

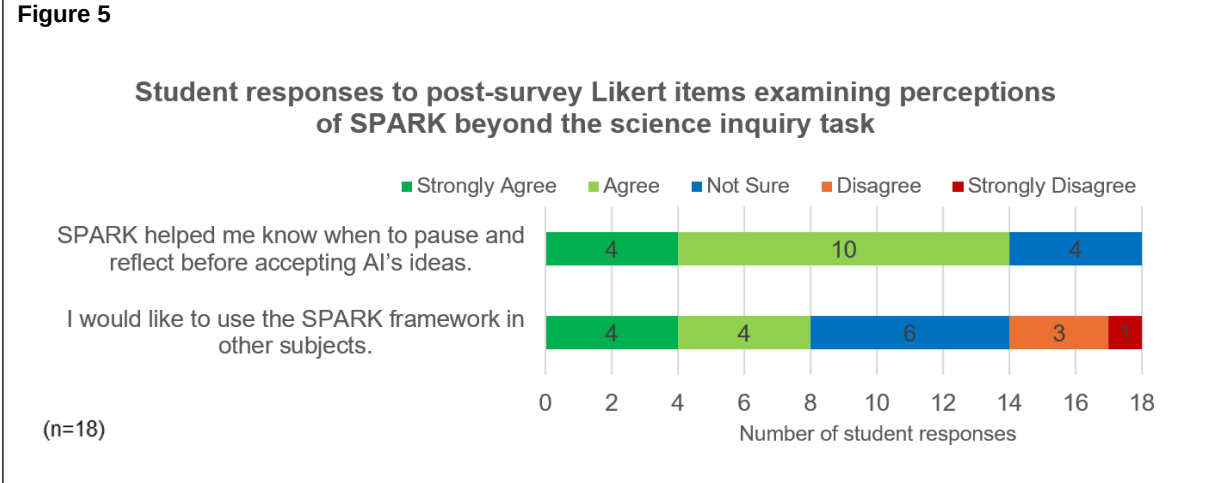
The girls’ uptake of AI use, however, was not automatic. A small number of girls initially chose not to use AI, citing limited experience, environmental concerns or desires to preserve their own thinking, including concerns about “stealing other people’s ideas” or AI generating inaccurate information. Over time, however, all girls chose to trial AI to varying degrees, suggesting that initial reluctance reflected caution rather than disengagement. One student articulated this stance clearly, noting in the mid-intervention interview, “I’m not sure if I plan to use AI in the future ... but if I was to use it, I’d make sure [I was] being specific about the context and how much you already know, because it dumps a lot of information at you at first.” This reflection illustrates evaluative judgement about AI use, demonstrating that critical thinking extended beyond judging outputs to judging the conditions under which AI should be used.

Student agency was evident in the way that girls re-prompted AI as part of the refinement process, and in some cases, chose to redirect or disregard initial AI outputs to

better align with their own reasoning (see Appendix D). While all students recognised limitations in initial prompts, most (14 out of 18) developed more purposeful prompts that deepened explanations or generated more targeted inquiry. For example, one student shifted from a broad request for ‘researched data’ to a focussed query linking human impacts on wildlife and interconnected environmental processes. However, fewer students challenged AI assumptions (such as default perspectives, generalisations, or the polish and completeness of AI-generated explanations), suggesting this aspect of evaluative judgement required more explicit scaffolding.

These patterns indicate that structured evaluation shaped how girls used AI, leading to more purposeful refinement, although the depth of critique varied. Collectively, this reflects emerging control over AI as a thinking partner rather than a content generator. One student captured this shift, stating “It [SPARK] helped me choose the harder path, not the easy AI answer.”

During the inquiry, most students suggested that co-thinking routines could be applied in other subjects. However, post-survey responses (see Figure 5) indicate mixed confidence in transferring co-thinking routines, suggesting that agency developed within the inquiry did not consistently transfer across subjects.



Interview reflections explained the reasons behind this uncertainty. The way AI was introduced appears to have shaped girls’ sense of agency. For example, one student

described initially assuming the task was a test, before realising that AI was permitted across different inquiry lessons, explaining that “a lot of us have never been able to actually use AI as part of a task [and] not being sneaky with it ... it was kind of a new experience, having an open conversation about it [using AI].” This suggests that explicit permission and open discussion supported girls to evaluate AI more confidently and deliberately, rather than using it covertly or uncritically.

While students valued this openness and described SPARK as “very simple and easy to understand, but they’re just not everywhere [as classroom posters],” several also expressed concern that AI use in other subjects might be perceived as “not doing anything,” or “sneaky.” Some students also noted that routines such as annotating AI conversations (see Appendix D), while helpful for making evaluation visible, may not be routinely expected in other subjects. SPARK was perceived as most relevant for science or research-based subjects, with English viewed as a higher-risk for outsourcing; as one student noted, “a lot of the time people are just like, ‘Oh write me an essay’.”

The findings indicate that while students recognised the value of evaluative routines, their application beyond the inquiry depended on subject-specific norms and expectations. Although co-thinking routines strengthened critical thinking, agency, and evaluative control during the inquiry, transfer beyond science was shaped by inconsistent expectations and permissions surrounding AI use across subjects. This suggests that evaluative judgment, while strengthened through structured inquiry, is context-sensitive and requires consistent modelling across subjects rather than being assumed to transfer automatically.

Conclusions

Across the three themes, analysis indicates that the SPARK framework functioned as a shared language for co-thinking, rather than a set of instructions. The routines embedded evaluation within the inquiry, shifting AI from an answer generator to a tool students could question, test, and refine.

Developing critical and evaluative judgement depended less on the tool itself and more on how learning was structured around it. Intentional routines that required girls to pause, annotate, revise, question, and justify their thinking made judgment visible rather than assumed. For educators, this reinforces a simple but important idea: design the thinking first, then decide where AI fits. When evaluation was visible, girls were more likely to retain ownership of their reasoning.

The findings of this study are specific to one Year 9 Science class of 18 girls within a strong culture of academic accountability and learning for impact. They may not translate directly to other school contexts, particularly those with different expectations around technology use. Survey findings relied on self-reported Likert data and may be subject to response bias. In addition, my role as both teacher and researcher inevitably shaped the design, implementation, and interpretation of the study, despite efforts to triangulate data and include colleague observations.

Future research could explore three areas in regard to co-thinking routines : transfer across disciplines; contribution to academic buoyancy and confidence in STEM pathways for girls; and the sustainability of evaluative judgement with AI over time.

Taken together, the shift was not driven by AI itself, but by pedagogical design. When girls are explicitly taught how to evaluate AI, rather than simply allowed to use it, AI becomes less about efficiency and more about building the habits of mind that science inquiry depends on.

Reflection Statement

This action research project has been one of the most energising, affirming and transformative experiences of my professional life. It gave me the space to explore something I care deeply about—how we ensure girls develop confidence, sound judgement, and agency when working with emerging technologies. What began as a question about AI in the classroom became a much deeper exploration of thinking, ownership, and growth.

Throughout this process, I have grown not only as a classroom teacher, but as a leader within our school. I've come to better understand the value of what I bring—a willingness to question, to design with intention, and to place student voice at the centre of change. Being acknowledged through my new leadership role as Head of STEM and AI Education feels like a meaningful continuation of this work and an opportunity to further develop thoughtful, ethical AI approaches across the school.

I am grateful to our Principal, Kim Bence, for supporting this journey and recognising the contribution this project makes to our broader vision for learning.

My thanks go to my GARC research advisor, AnnMarie Zigrossi, for her ongoing encouragement and belief in the initial concept. Her thoughtful discussions and feedback consistently pushed my thinking further and strengthened the depth of this work.

I am grateful to my Ruyton mentor, Cathryn Furey, for her time, generosity, and steady support throughout the process. She played a central role in shaping my growth in co-thinking with AI, encouraging me to pursue professional learning opportunities and to think more deeply and boldly about this work. Her guidance, curiosity and support have been instrumental in the development of this project.

My thanks also extend to my 2025 Year 9 teaching colleagues, Katrina Mount, Ruth Rae, and Maria Divitto, for their feedback, guidance, and willingness to engage in something new. This inquiry task was different from anything we had done before, and their openness to iteration and refinement supported portfolio developments that evolved over time for students.

I also thank my Ruyton colleague, Margarita Condello, for providing classroom observations and thoughtful feedback. Her insights and professional honesty strengthened the rigour of this work.

To my Year 9 students of 2025, thank you for your honesty, curiosity and willingness to trial something new. Your feedback, questions and ideas have shaped SPARK V4.2 in ways that will influence learning well beyond your cohort.

And finally, to my family—thank you for living alongside the many portfolios and papers that occasionally took over the living room, and for your quiet support during busy submission weeks. Your patience and encouragement made this possible.

This project has reinforced my belief that schools must not simply react to AI, but design for it—with intention, clarity, and care. This work has already begun to evolve within our school, with SPARK continuing from structured co-thinking routines into more flexible “thinking moves” that can be applied across contexts. I am proud of what has been created and excited about where it will lead next.

References

- Abdelghani, R., Sauz on, H., & Oudeyer, P. (2023). Generative AI in the Classroom: Can Students Remain Active Learners? <https://doi.org/10.48550/arXiv.2310.03192>
- ACARA. (2024). *Critical and Creative Thinking | V9 Australian Curriculum*. (Australian Curriculum, Assessment and Reporting Authority) [australiancurriculum.edu.au. https://v9.australiancurriculum.edu.au/curriculum-information/understand-this-general-capability/critical-and-creative-thinking](https://v9.australiancurriculum.edu.au/curriculum-information/understand-this-general-capability/critical-and-creative-thinking)
- AI Swiss. (2025). *White Paper: Human-AI Co-Thinking: Transforming Swiss Education*. <https://a-i.swiss/resources>
- Australian Government. (2024). STEM Equity Monitor Data report 2024. *Australian Government Department of Industry, Science and Resources*. <https://www.industry.gov.au/sites/default/files/2024-07/STEM-Equity-Monitor-data-summary-report-2024.pdf>
- Bianchi, I. (2024). Enhancing Critical Thinking in Education: Leveraging AI Tools and Technologies in Pedagogical Practice. *October*. <https://www.researchgate.net/publication/384632236>
- Bray , A., Byrne, P., & O'Kelly, M. (2020). A Short Instrument for Measuring Students' Confidence with "Key Skills" (SICKS): Development, Validation and Initial Results. *Thinking Skills and Creativity*, 37(100700). <https://doi.org/10.1016/j.tsc.2020.100700>
- Bronstein, B., Griffiths, M., Ludwig, C., Morrison, D., Smith, H., & Steffens, B. (2024). Action and Engagement in Climate Science. *Connected Science Learning*, 6(4), 179-181, <https://doi.org/10.1080/24758779.2024.2376483>
- Chang, Y.-S., Wang, Y.-Y., & Ku, Y.-T. (2023). Influence of online STEAM hands-on learning on AI learning, creativity, and creative emotions. *Interactive Learning Environments*, 32(8), 4719–4738. <https://doi.org/10.1080/10494820.2023.2205898>

- Encinas-Martin, M. (2025, April 15). *Ensuring equitable AI: the role of education in shaping an inclusive digital future*. OECD. <https://www.oecd.org/en/blogs/2025/04/ensuring-equitable-ai-the-role-of-education.html>
- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Girls' Day School Trust (GDST). (2023). *The Girls' Futures Report: Shaping skills, driving confidence and creating leaders*.
<https://www.gdst.net/wp-content/uploads/2023/09/Girls-Futures-Research-updated-01.08.23.pdf>
- King, A. (2025). *Thinking with AI*. United Kingdom: Thinking EdTech.
- Melisa, R., Ashadi, A., Triastuti, A., Hidayati, S., Salido, A., Ero, E., Marlini, C., Zefrin, Z., & Fuad, Z. (2025). Critical Thinking in the Age of AI: A Systematic Review of AI's Effects on Higher Education. *Educational Process International Journal*, 14(1).
<https://doi.org/10.22521/edupij.2025.14.31>
- Mertler, C. A. (2020). *Action Research: Improving Schools and Empowering Educators (Sixth Edition)*. SAGE Publications.
- Mulder, R., Baik, C., & Ryan, T. (2025). *Rethinking Assessment in Response to AI*. The University of Melbourne. Melbourne Centre for the Study of Higher Education.
https://melbourne-cshe.unimelb.edu.au/__data/assets/pdf_file/0004/4712062/Assessment-Guide_Web_Final.pdf
- OECD. (2025). *Introducing the OECD AI Capability Indicators*. OECD Publishing.
<https://doi.org/10.1787/be745f04-en>
- OECD. (May, 2025). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education (Review draft)*. OECD.
<https://ailiteracyframework.org>

- Oliveira, L., Tavares, C., Strzelecki, A., & Silva, M. (2025). Prompting Minds: Evaluating how Students Perceive Generative AI's Critical Thinking Dispositions. *Electronic Journal of ELearning (EJEL)*, 23(2), 1–18. . <https://doi.org/10.34190/ejel.23.2.3986>
- Paulson, E. (2024, Dec 3). *Critical Thinking with AI: 3 Approaches*. Conestoga.on.ca. <https://tlconestoga.ca/critical-thinking-with-ai-3-approaches/>
- Pearce, J., & Chiavaroli, N. (2023). Rethinking assessment in response to generative artificial intelligence. *Australian Council for Educational Research (ACER)*. https://research.acer.edu.au/higher_education/77
- Stupple, E., Maratos, F., Elander, J., Hunt, T., Cheung, K., & Aubeeluck, A. (2017). Development of the Critical Thinking Toolkit (CriTT): A measure of student attitudes and beliefs about critical thinking. *Thinking Skills and Creativity*, 23, 91–100. <https://doi.org/10.1016/j.tsc.2016.11.007>
- Sun, D., Cheng, G., Leung Ho Yu, P., Jia, J., Zheng, Z., & Chen, A. (2025, Feb 11). Personalized stem education empowered by artificial intelligence: a comprehensive review and content analysis. *Interactive Learning Environments*, 1–23. <https://doi.org/10.1080/10494820.2025.2462156>
- The Atlantic. (2024, 07 16). The Metacognition Revolution: AI is Playing a Central Role in Reshaping How we Learn. *The Atlantic*. <https://www.theatlantic.com/sponsored/google/metacognition-revolution/3924/>

AI Use Statement

All concepts, themes and ideas presented in this report are the author's own. The research questions, intellectual synthesis and interpretation of findings, were independently written by the author.






AI tools were used in a limited and supportive capacity to assist with proofreading, refining clarity of expression, and acting as a cognitive partner to extend and test the author's thinking. All AI-assisted outputs were guided by deliberate prompts, critically reviewed, and substantially edited to ensure accuracy, coherence and alignment with the research intent.

Consistent with the approach outlined in this project, AI was used to support thinking—not replace it—with human agency, expertise and professional judgement remaining central throughout.

Appendix A

SPARK Framework

The intervention centred on the development of the SPARK framework, designed to scaffold explicit evaluation and critical engagement with AI-generated outputs during inquiry. It drew on Paulson's (2024) three critical thinking analogies - the Gardener's Tree, Navigator's Map, and Sculptor's Stone—which emphasise idea generation, verification, and iterative refinement. These ideas were integrated with Mertler's (2020) action-research cycle to reflect the iterative process of planning, feedback, and refinement central to co-thinking with AI.

S  Start with what you Know	Begin tech-free - brainstorm, sketch or write what you know.
P  Prompt with Purpose	Write clear, specific prompts. Keep a log of your prompts and conversations.
A  Assess for Accuracy & Bias	Evaluate AI's responses for reliability, logic, missing perspectives, data or bias.
R  Refine, Reflect, Reference	Use your knowledge to improve an AI output. Track what you kept, changed, or rejected. Validate concepts and facts through referencing.
K  Know your Voice	Label or identify what's yours Vs AI. Explain how AI influenced your prompts and thinking.

Appendix B

Critical Thinking Phases Within the Inquiry Portfolio

Track Your Thinking With AI

PART A

Phase 1

Task Introduction

- Explore a human impact on climate change through both your own thinking and by co-thinking with AI.
- Human – ‘brain dump’ (your mind map)

Remains at school

Phase 2

10-15min Conversation with AI.

- Use sample prompts provided to help you get started and track your own prompts and ideas

Print and Annotate your Conversation.

- Validate AI outputs - Find at least five points in your conversation to cross reference.
- Use colour codes to track your thinking.

May be taken home

Phase 3

Update Mind Map

- Use a different colour pen to update and annotate your mind map.
- Show how AI has helped you form new links and ideas.
- Identify what you want to explore next.

Reflect On:

- accuracy and any bias of AI outputs
- how might you adjust your prompts to change the AI output?
- new learning and how you have used AI in this task.
- Etc.

At school

Collaborate

PART B

Phase 4

Form groups

- Compare mind maps and find common ideas

Share ideas

- Select one that could be actionable for the school community

At school

Phase 5

Using Tech Ethically

- Review and discuss a series of infographics
- Write an AI use statement to show how your group wants to work with technology

At school

Phase 6

Research and Draft Pitch

At school

Phase 7

Shark Tank! Pitch and Prizes

Group presentations (2 min)

At school

Appendix C:

Critical Thinking Toolbox (One-Page Student Reference Guide)

🔥 IGNITE YOUR THINKING - Critical Thinking Toolbox

S	Start with what you Know	Begin tech-free - brainstorm, sketch or write what you know.
P	Prompt with Purpose	Write clear, specific prompts. Keep a log of your prompts and conversations.
A	Assess for Accuracy & Bias	Evaluate AI's responses for reliability, logic, missing perspectives, data or bias.
R	Refine, Reflect, Reference	Use your knowledge to improve an AI output. Track what you kept, changed, or rejected. Validate concepts and facts through referencing.
K	Know your Voice	Label or identify what's yours Vs AI. Explain how AI influenced your prompts and thinking.

Be intentional - How are you going to use Tech/AI?


S/P **Gardener's Tree:** Begin with your own ideas, and use Tech/AI to connect and grow ideas.
Example: brainstorm essay topics, then expand through peer discussions, references and technology

A **Navigator's Map:** Tech/AI output is a 'map', compare it with real evidence.
Example: check an AI summary against textbook notes or an experiment.

R/K **Sculptor's Stone:** Tech/AI gives a rough 'block'; you refine it with your own thinking & clear prompts.
Example: AI suggests 5 dot-points for a science conclusion – you combine, edit, and rewrite them into a clear paragraph in your own words.

S-P-A **Track your thinking**

1. Use different pen colours to track your thinking, co-thinking with tech/AI, and enhancing concepts.
2. Use a table to document your thinking. You may adapt this table and include:
 - What makes you more curious?
 - What information consolidated or extended your thinking?
 - What AI outputs do you need to validate and check?



My prompt or conversation with AI...			
My Thinking	AI output	I'm now wondering about... My next prompt will include	I need to check...
What I already know about the topic	What AI said in response		
My first attempt at an answer or explanation	AI-generated version		
What I agree/disagree with and why	Edits I made after reflection		

A-R-K **Ethical AI Use**

- Label what is YOUR voice vs AI's voice.
- You may choose not to use AI, but explain why.
- Always cite AI by including screenshots or prompt logs.
- Remember: One AI query uses resources (like water and energy). Use it wisely.

A-R-K **Bias & Reliability Check**

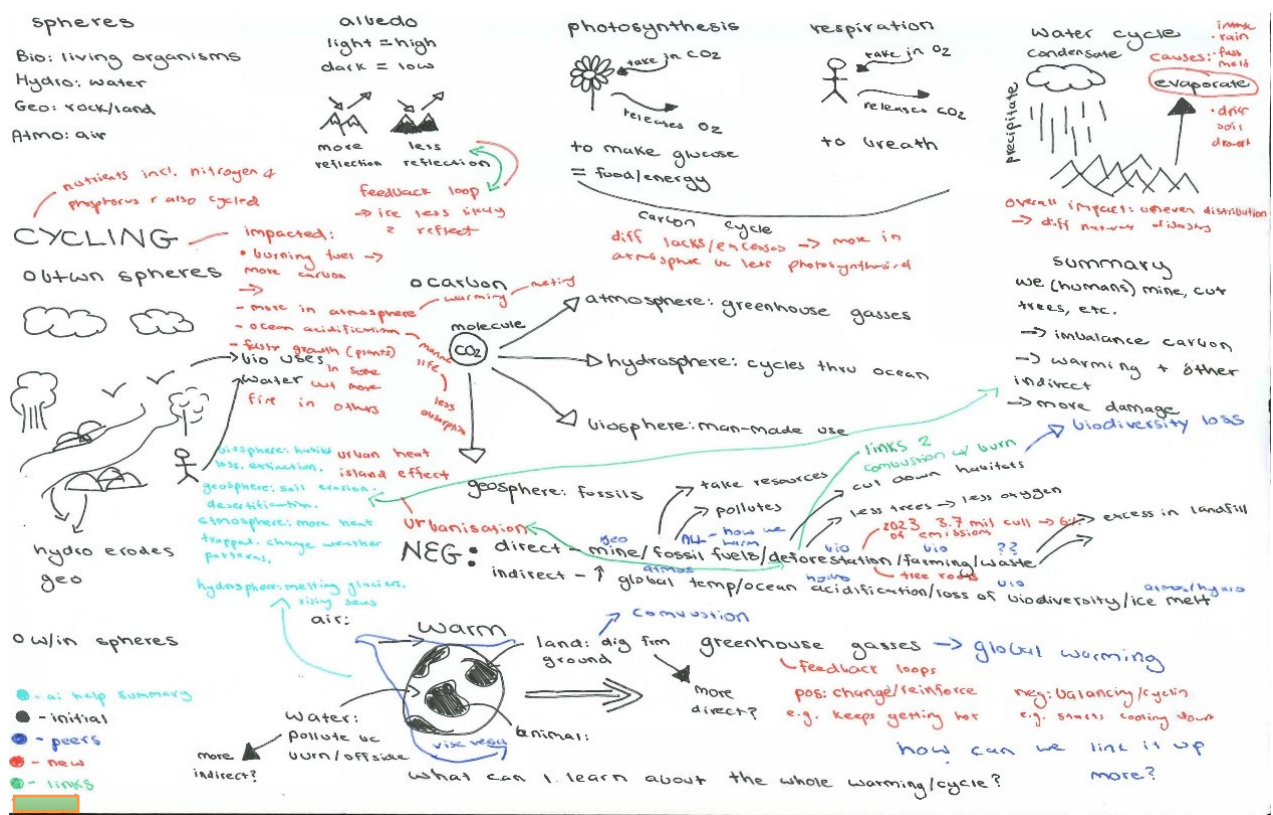
- Whose perspective might be missing?
- Does this sound too polished or general?
- How do I know this information is reliable? Can I confirm with textbook / experiment / teacher?

Appendix D:

Student Artefacts Demonstrating Co-Thinking and Visible Reasoning

Student artefacts include paired examples of annotated AI conversations and corresponding mind maps, illustrating how ideas developed through co-thinking.

In their mind maps, students used colour coding to represent different stages of thinking: beginning with their own ideas (black), collaborating with peers (blue), extending ideas through co-thinking with AI (red), and linking new ideas with existing knowledge (green/purple).



= conformation
= new/expanded
= incorrect

request Hi, I'm a Year 9 Australian student. Below is my summary about human impacts on Earth's spheres. Please give me 2 more points about human's impacts on Earth's spheres and highlight any misconceptions to clarify. My summary is that humans impact climate change across all 4 spheres both directly and indirectly. Direct impacts such as mining of fossil fuels impact many spheres (including the geosphere and atmosphere). This impact on the atmosphere results in further impact on other spheres because of a rise in greenhouse gasses and associated warming of the planet causing further climate impacts such as ice melts effecting the hydrosphere.

point

link & climate change

Thanks for sharing your summary — you're clearly thinking deeply about how human activities affect Earth's systems! Here's how you can build on it:

Two More Human Impacts on Earth's Spheres

- Deforestation and its ripple effects**
 - Biosphere:** Removing forests directly affects plant and animal life, reducing biodiversity.
 - Atmosphere:** Fewer trees means less carbon dioxide is absorbed, increasing greenhouse gases. → expand/research further - in 2023 3.7 mil cut → 6% emissions
 - Geosphere:** Tree roots help hold soil together, so deforestation can lead to erosion and land degradation.
 - Hydrosphere:** Erosion can cause sediment to enter rivers and streams, affecting water quality and aquatic life.
- Urbanisation and land use change**
 - Geosphere:** Building cities alters land surfaces, replacing natural soil with concrete and asphalt.
 - Atmosphere:** Cities often trap heat (urban heat island effect), contributing to local warming.
 - Hydrosphere:** Impermeable surfaces reduce water absorption, increasing runoff and flood risk.
 - Biosphere:** Urban expansion can fragment habitats, making it harder for wildlife to survive.

Misconceptions to Clarify

- Spelling note:** "Effecting" should be "affecting" in this context. "Affecting" means causing change, while "effecting" means bringing something about intentionally.
- Climate change vs. general human impact:** Climate change is a major result of human activity, but not all human impacts are climate-related. For example, plastic pollution in oceans affects the hydrosphere and biosphere without directly contributing to climate change.
- Greenhouse gases and warming:** You're right that greenhouse gases warm the planet, but it's helpful to mention that this warming leads to **feedback loops** — like melting ice reducing Earth's ability to reflect sunlight, which causes even more warming.

Would you like help turning this into a polished paragraph or visual diagram for your assignment?

pos: changing/reinforcing so it just continues heat up
 neg: balancing cycles - cool so it just would (slowly) cool

Can you explain greenhouse gasses and the impact of warming to me further by using the examples of deforestation and urbanisation

Absolutely! Let's break it down using your examples of **deforestation** and **urbanisation**, and show how they connect to **greenhouse gases** and **global warming**.

Deforestation and Greenhouse Gases

- Trees absorb carbon dioxide (CO₂),** a major greenhouse gas, through photosynthesis. Forests act like giant sponges, soaking up CO₂ from the atmosphere.
- When forests are **cut down or burned**, two things happen:
 - Less CO₂ is absorbed** because there are fewer trees.
 - More CO₂ is released** into the atmosphere from the burning or decay of trees.
- This increase in CO₂ traps more heat in the atmosphere, contributing to **global warming**.
- Warming leads to **climate feedbacks** like:
 - Melting ice (affecting the **hydrosphere**)
 - Changing rainfall patterns (affecting the **biosphere** and **geosphere**)

(electrification, renewables, efficiency, protecting forests) reduce each source.

first time hearing about electrification - according to energy.gov.au, it is the process of replacing non-electric fuels with electric ones to reduce carbon footprint

known by prior knowledge learned in past years during class

One misconception to clarify

How greenhouse gases trap heat

Greenhouse gases do not create new infrared waves by themselves; they absorb outgoing infrared radiation from Earth and then re-emit that energy in all directions.

Some of that re-emitted energy goes back toward the surface, slowing the loss of heat to space and raising the planet's temperature. This is a radiative transfer process, not the gas "making" extra heat out of nothing! Greenhouse gases also differ in lifetime and strength, so CO₂'s long atmospheric lifetime makes cumulative emissions especially important.

proven by site - technically NOT correct, re-emitted energy in the form of infrared waves hits other molecules of greenhouse gases, NOT directly hits the surface like it suggests, however, it is correct that it slows the loss of heat to space as heat stays near Earth's surface for longer.

proven by site - I understand / know this already, but I must have not inputted a clear enough explanation of my understanding of this concept to AI.

proven by site

proven by ScienceDirect.com - RT is the transport of electromagnetic radiation

QUESTION: comparison of different greenhouse gases' lifetimes? through media is shorter or longer lifetime better for our planet?

One focused research question you could investigate

How much Arctic summer sea ice is lost per tonne of CO₂ emitted?

Research approach: look for peer-reviewed studies that quantify sea-ice area loss per tonne of CO₂, compare results across different time windows, and check whether the numbers differ for summer vs annual ice. Use datasets from climate research groups and link the per-tonne loss to personal or national emissions to make the result relatable.

proven by National Institutes of Health - linear relationship of a loss of 3 ± 0.3 sq metres of sept. sea ice per metric ton of CO₂ emissions

If you want, I can:

- Draft a short paragraph you could add to each expanded idea for your report, or
- Find simple numbers and classroom-friendly graphs (e.g., CO₂ concentration history, per-tonne sea-ice loss) and explain them step-by-step. Which would help most?

Could you find data to support your question on 'How much Arctic summer sea ice is lost per tonne of CO₂ emitted, and a comparison on the CO₂ emissions for household uses vs factories?

- The geosphere stores fossil fuels — ancient carbon from dead organisms.
- When we extract and burn these fuels, we release that stored carbon into the atmosphere.
- The land surface (part of the geosphere) also affects the energy budget:
 - Dark surfaces like asphalt absorb more solar energy.
 - Deforestation and mining change land reflectivity (albedo), increasing heat absorption.

Already said this, but liked how it says again - more helpful and helps me understand in each situation.

like that link.

→ combustion

Summary for Your Science Project

"The geosphere stores fossil fuels, which humans burn to produce energy. This releases greenhouse gases into the atmosphere, trapping heat and disrupting Earth's energy budget. As more energy is absorbed than released, global temperatures rise — linking the geosphere, atmosphere, and climate system."

Would you like help turning this into a visual diagram or a short explainer paragraph for Year 9 students? I can also help you link this to Australian examples like coal mining or bushfire impacts.

Edit in a page

can you give me more ideas based off of this to expand my knowledge and what I could research further

Absolutely — here are some layered ideas you can explore to deepen your understanding and expand your research around fossil fuels, the geosphere, greenhouse gases, and Earth's energy budget:

Topics to Expand Your Knowledge

1. Carbon Cycle and Human Disruption