Beyond Artificial Intelligence Literacy: Conceptualising Digital Immunity against Hallucination Risks in Education

Nur Ashiela Abdul Manaf., Mohd Effendi @ Ewan Mohd Matore

Faculty of Education, National University of Malaysia, Selangor, Malaysia

DOI: https://doi.org/10.47772/IJRISS.2026.100500039

Received: 26 April 2026; Accepted: 02 May 2026; Published: 22 May 2026

ABSTRACT

The rapid integration of generative Artificial Intelligence (AI) tools into students' academic practices has brought not only benefits but also a hidden threat, namely AI hallucination. This phenomenon occurs when systems generate outputs that appear convincing but are in fact inaccurate, thus posing a critical challenge to academic integrity. This concept paper aims to elaborate and clarify the position of the concept of digital immunity against AI hallucination risk in educational contexts by developing a conceptual framework that integrates cognitive, behavioural and contextual dimensions of AI use. Its specific objectives are to explain the forms of AI hallucination risk in education, identify research gaps in AI literacy, propose a risk-assessment instrument framework and discuss the implications of its implementation for learners, teachers and educational institutions. Methodologically, the paper employs document analysis and synthesis of recent literature, drawing on AI literacy frameworks, dual-process theory, risk perception theory and information quality frameworks to shape the constructs of digital immunity and hallucination-risk domains. The conceptual findings indicate the need for a valid and reliable psychometric instrument to assess risk awareness, the ability to detect false information and dependence on AI as a basis for planning pedagogical interventions. The main limitation of this study is the absence of empirical data to validate the proposed digital-immunity framework and AI hallucination-risk constructs. Hence, future studies may focus on developing and empirically validating a psychometric instrument to measure levels of digital immunity and AI hallucination risk among educational users. Fostering digital immunity is a crucial step to protect the cognitive safety of educators and learners, uphold academic integrity and align AI use with the aspirations of the national Digital Education Policy.

Keywords: Generative Artificial Intelligence; AI hallucination risk; Digital Immunity; Education.

INTRODUCTION

The rapid integration of generative artificial intelligence (AI) tools, such as ChatGPT and Gemini, has become increasingly common in students' academic practices, fundamentally transforming the ways in which they access information and complete academic tasks (Dwivedi et al., 2023). Although these technologies offer substantial advantages in terms of efficiency and access to knowledge, they also introduce significant challenges. One of the most pressing challenges arising from this technological advancement is the phenomenon of AI hallucination. This phenomenon refers to situations in which AI systems generate outputs that appear convincing and authoritative but are in fact inaccurate, irrelevant, or entirely fabricated (Ji et al., 2023). AI hallucinations are not merely technical errors; rather, they constitute a latent threat that can undermine academic integrity and students' cognitive development.

A major risk associated with AI hallucinations in educational contexts concerns students' levels of awareness and their ability to critically evaluate whether AI-generated information contains hallucinations or false data. According to Ciubotaru (2025), this phenomenon can compromise the accuracy and reliability of digital learning, while Erümit & Sarıalioğlu (2025) emphasise that students who become overly reliant on AI are at risk of accepting incorrect information, which may negatively affect their conceptual understanding. More concerningly, prolonged exposure to persuasive yet erroneous AI outputs may erode students' critical thinking skills, diminish trust in technology more broadly, and contribute to the wider dissemination of misinformation within society.

In response, this concept paper focuses on the development and cultivation of digital immunity as a capability that enables secondary school students to counter the risks arising from the use of generative AI tools in learning. The primary population of interest comprises secondary school students within formal school-based educational contexts. The key variables addressed include AI hallucination risk and dimensions of digital immunity, operationalised through AI literacy, critical thinking, and information verification practices. This proactive approach is essential for understanding the scale of the problem and for designing effective pedagogical interventions. Accordingly, this concept paper outlines a strategic justification for addressing critical gaps in digital literacy, with the aim of ensuring that future generations are equipped to engage with AI in a wise, critical, and responsible manner.

Purpose And Objectives

Main Purpose

The primary purpose of this concept paper is to explicate the concept of digital immunity against AI hallucination risks in educational contexts through the development of a conceptual framework that integrates cognitive, behavioural, and contextual dimensions of AI use.

Specific Objectives

i. To explain the nature and forms of AI hallucination risks encountered by generative AI users in educational contexts.

ii. To identify research and practice gaps related to AI literacy and hallucination-risk assessment that hinder the development of digital immunity.

iii. To propose a conceptual framework for a risk-assessment instrument integrating cognitive and behavioural factors to identify AI-resilient user profiles.

iv. To discuss implementation strategies and implications for students, teachers, and educational institutions in fostering systematic and integrated digital immunity.

Contribution of the Study

Awareness of AI hallucination in generative AI outputs is essential for developing digital immunity, which refers to the cognitive and behavioural capacity to critically question, verify, and filter AI-generated information before accepting it as valid. Digital immunity reduces passive reliance on AI and encourages responsible, critical engagement with technology. Students who are able to recognise factual, contextual, multimodal, and logical hallucinations are better equipped to avoid inaccuracies, misinterpretations, and fabricated references in academic work, thereby preserving academic integrity (Ciubotaru, 2025; Elsayed, 2024; Ji et al., 2023)

An understanding of AI hallucination risks and the proposed digital immunity framework offers teachers and schools a practical basis for designing instruction that integrates critical thinking, AI literacy, and verification practices. Rather than restricting generative AI use, the framework supports its responsible pedagogical integration in AI-enabled learning environments (Dwivedi et al., 2023; Fulsher, Pagkratidou, and Kendeou, 2025). It may also inform the development of training modules, assessment tools, and school-level guidelines, supporting teacher professional development and the implementation of the 2027 School Curriculum, which identifies AI competence as a core skill.

Theoretically, this concept paper extends existing AI literacy discourse by moving beyond technical competencies to foreground AI hallucination risk and digital immunity as education-specific constructs. Practically, it offers a foundation for developing policies, intervention modules, and psychometric assessment instruments aimed at cultivating resilient and ethical digital citizens. Collectively, these contributions enhance the epistemic safety of AI use in education and align with global efforts to promote responsible and trustworthy educational practices.

Research Issues and Problems

The primary issue addressed in this study is AI hallucination, which remains insufficiently understood in relation to how secondary school students manage it in academic tasks. AI hallucinations are not merely common technical errors but often manifest as fluent, well-structured, and seemingly scholarly texts, making them difficult for students to detect without well-developed verification skills (Elsayed, 2024; Fulsher et al., 2025). The widespread accessibility of generative AI tools has intensified this issue, rendering AI hallucination an urgent challenge within contemporary educational ecosystems.

Within educational contexts, students had already experienced substantial difficulties in evaluating the credibility of online information prior to the emergence of generative AI. Research by Kiili et al. (2022) demonstrates that students at this level often struggle to provide robust justifications when assessing the credibility of digital texts. The introduction of generative AI further amplifies this risk, as students may perceive AI systems as expert sources without recognising that such systems can also generate false or misleading information. As generative AI tools are now actively used for academic assignments, students are increasingly exposed to AI hallucination risks. This situation creates a clear gap between the education systems aspiration to cultivate digitally fluent learners and the reality of student's vulnerability to misinformation.

A critical gap that necessitates this concept paper is the absence of a structured approach for assessing and developing digital immunity against AI hallucination among secondary school students. Although general AI literacy instruments exist, no specific assessment framework currently focuses on AI hallucination risk and students' verification behaviours (Zhang, Perry, and Lee, 2025). This limitation hinders the systematic evaluation of intervention effectiveness and constrains the precise targeting of strategies to reduce misinformation risks generated by AI, particularly in educational environment.

Theoretical Framework

The theoretical framework of this concept paper integrates four key theories that collectively explain how digital immunity against AI hallucination risks can be developed and assessed within educational contexts.

(i) AI Literacy Framework

The AI literacy framework emphasises an understanding of fundamental AI concepts, ethical awareness, and the ability to evaluate and interpret AI generated outputs. In this concept paper, AI literacy functions as a foundational element because users must first recognise that AI systems are fallible, understand their limitations, and comprehend how AI generates responses before they can assess hallucination risks. This framework therefore supports the development of constructs related to hallucination risk awareness and information accuracy evaluation, consistent with the work of Zhang et al. (2025) on structuring AI literacy components for student populations.

(ii) Dual-Process Theory

Dual Process Theory introduced by Kahneman (2011) distinguishes between two modes of thinking. System 1 represents fast, intuitive, and automatic processes, while System 2 involves slower, analytical, and reflective reasoning. In the context of AI hallucination, this theory explains users' tendency to trust fluent and persuasive AI generated outputs automatically rather than engaging in analytical evaluation of evidence. This theory helps explain blind trust in AI and supports the measurement of users' tendencies to accept or question AI generated information as part of digital immunity.

(iii) Risk Perception Theory

Risk Perception Theory, proposed by Slovic (1987), explain how individual's perceptions of threats influence their attitudes, emotions, and behaviours, sometimes more strongly than objective risk levels. In this concept paper, the theory is applied to conceptualise awareness of AI hallucination risk, referring to how users assess the severity and likelihood of negative consequences resulting from reliance on inaccurate AI-generated information. Clear and realistic risk perception is viewed as a critical driver of behavioural change, including increased verification practices, reduced blind dependence on AI, and more cautious and responsible AI use.

(iv) Information Quality Framework

The Information Quality Framework developed by Wang and Strong (1996) highlights key aspects of information quality such as accuracy, reliability, relevance, and usability. This framework guides the assessment of users' ability to detect false information by focusing on how they evaluate factual correctness, contextual relevance, consistency between text and visuals, and the validity of references generated by AI. It supports the identification of unreliable information in digital environments.

Synthesis of Theoretical Perspectives

These four theories are synthesised to support the development of a digital immunity framework for assessing AI hallucination risks within AI literacy. Collectively, the theories provide a foundation for knowledge development and risk awareness. Dual Process Theory explains cognitive tendencies to trust AI outputs, Risk Perception Theory clarifies how awareness of threats shapes behaviour, and the Information Quality Framework guides the operational definition of false information detection. The integration of these perspectives results in a coherent framework for understanding and measuring how educational users can become more resilient to AI hallucination risks, as illustrated in Figure 1;

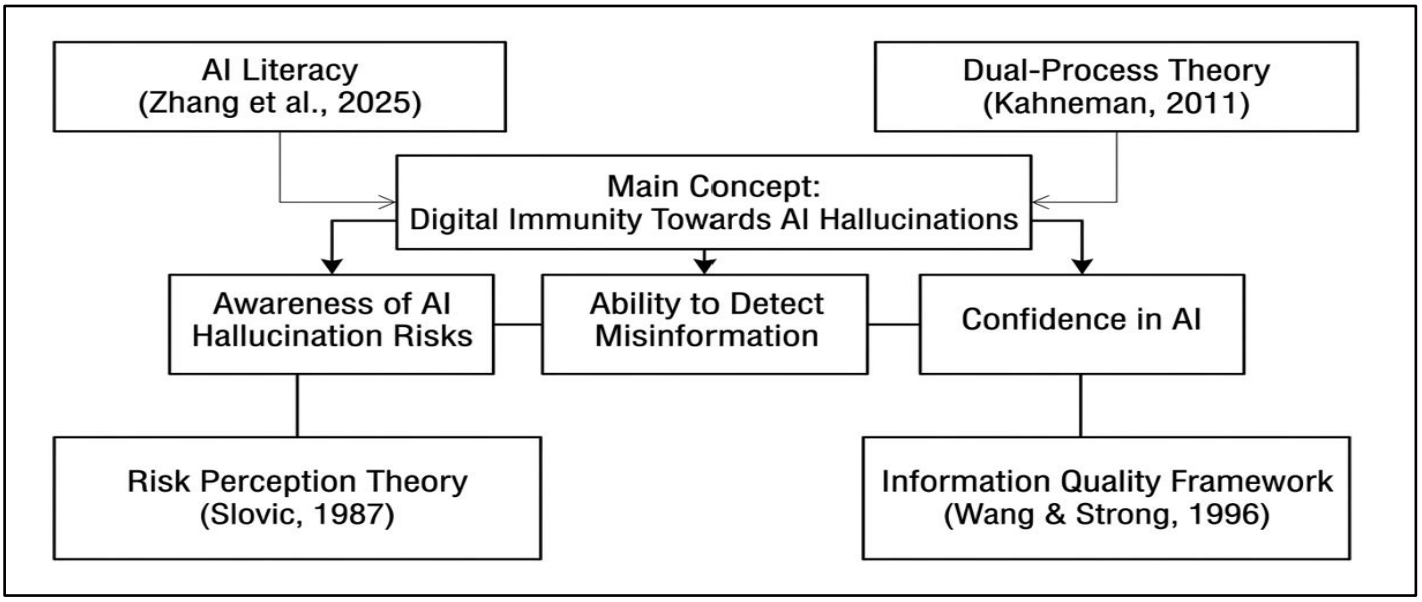

Figure 1: Theoretical framework

LITERATURE REVIEW

AI Hallucination Phenomenon and Real-World Cases

Technical literature indicates that AI hallucination is not an isolated issue, but a systemic problem in large language models and other generative systems (Ji et al., 2023). Mainstream media have reported notable incidents, such as Google Bard providing incorrect information about the James Webb Space Telescope and Gemini generating implausible historical images (Milmo and Hern, 2024; Sparkes, 2023). Empirical studies further show that AI hallucination undermines accuracy and trust in AI-based learning environments. Ciubotaru (2025) reports that hallucinated outputs reduce the reliability of digital learning. Similarly, Erümit and Sarıalioğlu (2025) point out specific risks in science education, where students may receive unsafe or inaccurate experimental information from AI systems. Together, these findings underscore the need for AI literacy that goes beyond conceptual knowledge to include the ability to detect and correct AI hallucinations.

Secondary School Students as a Population at Risk

Information literacy research shows that adolescents already face difficulties in evaluating the credibility of online sources and arguments, even before the widespread adoption of generative AI. Kiili et al., (2022) found that secondary school students often fail to provide strong justifications when judging the credibility of online texts. Instead, they tend to rely on surface features such as website design or the presence of images. With the emergence of generative AI that produces more fluent and persuasive text than typical web sources, the risk intensifies. Students may perceive AI outputs as more authoritative than other sources. This perception increases vulnerability to factual, contextual, multimodal, and logical hallucinations. As a result, secondary school students represent a priority population for efforts to develop digital immunity against AI hallucination risks.

AI Literacy Instruments and Measurement Gaps

Within the AI literacy domain, Zhang et al., (2025) developed the Artificial Intelligence Literacy Concept Inventory (AI-CI). This psychometric instrument measures secondary school students' knowledge and understanding of AI concepts. However, the scale focuses on general AI literacy and does not address AI hallucination risks or students' verification behaviours when interacting with generative AI outputs. Furthermore, Fulsher et al., (2025) in their studies discuss opportunities and challenges but do not provide specific tools to assess student's vulnerability to AI-generated false information. This gap highlights the need for a targeted framework. The present concept paper addresses this gap by proposing a four-construct framework, alongside relevant moderator factors, as a foundation for developing an AI hallucination risk assessment scale.

Malaysia's Digital Education Policy and Global Guidelines

At the policy level, Malaysia's Digital Education Policy emphasises the development of digitally fluent students who are not only technologically competent but also capable of using data analytically and ethically (Ministry Of Education (MOE) Malaysia, 2023). At the global level, UNESCO, (2023) highlights the importance of safe, ethical, and human-centred use of generative AI in education and research. These guidelines stress the need for effective protective mechanisms to prevent learners from exposure to false, misleading, or manipulative information. Despite these national and global efforts, digital immunity against AI hallucination has yet to be operationalised through specific assessment instruments or frameworks tailored to school-level learners.

Conceptual Framework for Risk Assessment Instrument

To address AI hallucination effectively, attention must shift from informal observation to the collection of measurable evidence. This approach is necessary to guide systematic instrument development. The framework serves as the foundation for a diagnostic tool that provides a clear picture of students' risk levels.

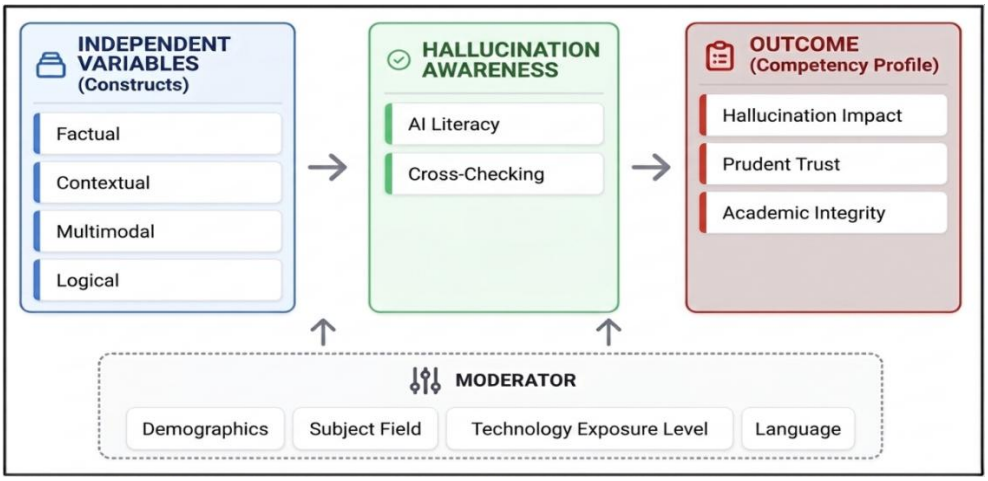

Figure 2: Conceptual Framework

Figure 2 illustrates the conceptual framework of the proposed risk assessment instrument. The framework comprises four independent variables, namely factual, multimodal, contextual, and logical hallucinations, which are shaped by moderator factors such as student demographics, subject domain, level of technology exposure, and language of interaction with AI. Technology exposure reflects both the frequency and diversity of AI use, while interaction language may influence the accuracy of generated outputs. The central mediating process in the framework is students' awareness of AI hallucination risks. This awareness promotes the development of AI literacy and systematic information cross checking. Ultimately, the framework aims to build a comprehensive student competency profile that supports the detection of false information, responsible AI use, and the maintenance of academic integrity through authentic, ethical work and accurate citation.

Psychometric Diagnostic Tool

The proposed instrument is designed as a psychometric diagnostic tool rather than a simple checklist. Its main purpose is to objectively identify students who are most vulnerable to AI hallucination. The instrument also aims to explain the underlying factors contributing to this vulnerability. Through this approach, educators can obtain actionable data to support targeted and evidence-based intervention planning.

Four Core Assessment Constructs

Based on a needs analysis, the proposed instrument adapts the hallucination taxonomy outlined by Ji et al. (2023). The instrument measures AI hallucination risk through four core constructs.

- **Factual Hallucination** measures students' ability to identify factually incorrect information generated by AI. This includes incorrect dates, fabricated statistics, and false citations.
- **Contextual Hallucination** assesses students' ability to detect information that is factually correct but irrelevant or inappropriate to the task or question.
- **Multimodal Hallucination** measures students' ability to evaluate the accuracy of outputs that combine text with images or visual data. These outputs may be misinterpreted or inaccurately generated by AI.
- **Logical Hallucination** assesses students' ability to identify flaws in reasoning, incoherent arguments, or illogical conclusions.

The instrument functions as a starting point for meaningful corrective action. The data collected enable educators to identify whether student risk is driven by weak fact checking skills, uncritical trust in technological authority, or difficulty distinguishing inaccurate outputs. The focus extends beyond measuring usage frequency or general literacy. Instead, the instrument evaluates the effectiveness of students' verification behaviours as a key defense against misinformation generated by AI systems.

Implementation Challenges

Efforts to foster digital immunity against AI hallucination face several implementation challenges. One key challenge concerns teacher's capacity and readiness. Many teachers remain in the early stages of adapting to generative AI and have limited understanding of how these systems operate. Time constraints, existing workload pressures, and the lack of targeted training in AI literacy have led some educators to avoid using AI altogether. This issue is critical, as the development of digital immunity requires structured exposure and systematic guidance from teachers.

Negative perceptions of AI may also arise if communication is not handled carefully. Overemphasising hallucination risks without explaining the benefits of AI and safe usage practices may create fear or resistance toward AI adoption (Valeri, Nilsson, and Cederqvist, 2025). Such perceptions contradict policy goals that promote AI literacy and digital competence. Digital immunity is not intended to restrict AI use, but rather to encourage more critical and ethical engagement with the technology.

Finally, challenges related to policy implementation and sustainability remain. The effective implementation requires adequate support resources. These include training materials, instructional modules, dedicated time for teacher and student training, and collaboration with research institutions. The rapid pace of AI development also demands continuous updates to hallucination examples, teaching strategies, and assessment instruments. Without regular review mechanisms and sustained research support, there is a risk that the digital immunity framework becomes outdated or remains a policy document without meaningful classroom practice.

Positive Implication

The primary aim of this initiative is to develop what can be described as digital immunity. This concept is based on the vaccination analogy introduced by UNESCO (2023). It is not intended to restrict or limit the use of AI technologies. Instead, the approach seeks to strengthen students' capacities by equipping them with critical skills, healthy scepticism, and cognitive strategies to counter potential risks arising from AI hallucinations. The vaccination analogy refers to controlled and guided exposure, not unrestricted use of AI errors. In school settings, especially in science subjects, a clear safety net is needed. This prevents students from accepting incorrect procedures, false data, or misleading explanations. Teachers can label hallucinated outputs as wrong, provide trusted references, and guide discussion with correction. This approach helps students practise verification while keeping learning accurate and safe. Students with digital immunity naturally question, verify, and validate AI-generated information. These practices become habitual rather than task-based. As a result, technology remains a tool that enhances human intellect rather than diminishing critical thinking.

This conceptual framework provides a foundation for the development of interactive and contextualised learning modules. These modules can be designed to expose students, in a controlled manner, to examples of AI hallucinations. Such exposure allows students to practise identifying errors in a safe learning environment. The goal is to improve students' understanding of how AI systems function, including their limitations and potential inaccuracies. The approach also supports the development of practical verification skills. This direction aligns with the recommendations of Dwivedi et al. (2023), who emphasise the urgent need to rethink educational strategies by strengthening critical thinking in response to generative technologies. The modules are intended to promote consistent and evidence-based verification practices among students.

The implications of this initiative extend beyond the classroom. The data generated highlight the need for sustained teacher professional development. Such development is essential to equip educators with the knowledge and skills required to guide students in responsible AI use. At a broader level, empirical data from this initiative can inform policy development. These data provide an evidence-based foundation for establishing guidelines that protect students from false or manipulative information, in line with concerns raised by Elsayed (2024). By integrating findings from this instrument into curricula and teacher training, education systems can better align with the rapid evolution of AI technologies.

IMPLEMENTATION RECOMMENDATIONS

Student Level: Learning-Based Digital Immunity Training

At the student level, the implementation of digital immunity should focus on building routines for questioning, verifying, and filtering AI-generated outputs within authentic task contexts, rather than relying solely on abstract warnings (Ji et al., 2023). One practical strategy involves tasks in which students use AI to generate responses and then systematically identify potential factual, contextual, multimodal, and logical hallucinations by consulting textbooks, academic journals, or other credible sources. This approach familiarises students with verification processes and aligns with established principles of information literacy and AI literacy.

Students can also be trained to apply strategies such as SIFT, which stands for stopping, investigating the source, finding better coverage, and tracing claims, when evaluating information suggested by AI systems (Fulsher et al., 2025). Practical activities may include verifying the existence of cited journals or tracing statements back to their original sources. Written reflection activities, such as learning journals or guided

group discussions, may further support students in articulating their experiences with AI hallucinations. These practices reinforce awareness and foster a cautious mindset that forms the foundation of digital immunity.

Teacher Level: Hallucination-Informed Pedagogical Design

Teachers play a critical role in translating the concept of digital immunity into classroom practice through pedagogical designs that expose students to AI hallucinations within controlled learning environments (Dwivedi et al., 2023) Teachers may design comparative activities in which students evaluate multiple AI responses to the same question and identify factual, logical, or contextual errors. These activities can be complemented by discussions on the causes of such errors and the methods used to detect them, drawing on real-world cases reported in science and technology media.

Targeted AI literacy training for teachers is also essential to ensure they understand different forms of AI hallucination, technological limitations, and effective ways to explain these issues to students using accessible language. Professional development workshops that incorporate case examples, AI usage simulations, and guidance on constructing assessment rubrics for verification skills can support teachers in integrating digital immunity into instruction without adding unrealistic workload demands.

To reduce teacher workload and promote scalable adoption, future implementation efforts may focus on the development of "plug-and-play" pedagogical modules that can be readily integrated into existing lesson plans without extensive preparation. These modules may include structured comparative activities in which students analyse and compare multiple AI-generated responses to the same prompt, identify hallucination types, and justify their evaluations using authoritative sources. Such ready-to-use designs support teachers who are still developing confidence in AI-integrated instruction while ensuring pedagogical consistency across classrooms.

School Level: Policy Culture and Structural Support

At the school level, implementation efforts should prioritise the development of policies, institutional culture, and structural support that align with the aspirations of the Digital Education Policy. Educational institutions may establish AI usage guidelines that clearly define permissible purposes, verification expectations, principles of academic integrity, and the respective roles of students and teachers in ensuring responsible AI use. These guidelines are best developed through collaborative discussions involving teachers from multiple disciplines, school administrators, and where appropriate, student representatives, to promote shared understanding and ownership.

Schools are also encouraged to establish communities of practice or small teams, such as AI literacy committees, to collect examples of AI hallucination cases, share best practices, and coordinate digital education activities throughout the academic year. Structural support, including reliable internet access, appropriate software, and dedicated time within school schedules for digital literacy initiatives, can further strengthen equitable and comprehensive implementation across the school community.

CONCLUSION

Digital immunity against artificial intelligence hallucination risks has become an urgent priority in secondary education, as generative AI increasingly shapes students' learning practices. Without the ability to question, verify, and filter AI-generated information, students are vulnerable to incorporating inaccurate content into academic work and conceptual understanding, thereby undermining academic integrity and knowledge reliability.

This concept paper has articulated the research problem, theoretical grounding, conceptual framework, and educational and policy implications of digital immunity. Drawing on contemporary literature on AI hallucination, AI literacy, and digital education policy, the paper positions digital immunity as a critical complement to existing AI literacy efforts. In alignment with national Digital Education Policy goals and UNESCO guidelines, digital immunity supports the development of learners who are not only competent AI users, but also critical and ethically informed digital citizens.

This study highlights the need to move from speculative concerns toward evidence-based interventions supported by valid and reliable measurement. The absence of empirical validation remains a key limitation of the proposed framework. Future research should therefore prioritise the development and validation of a psychometric instrument to assess AI hallucination risk and digital immunity. Such an instrument would enable systematic identification of vulnerabilities, support targeted pedagogical interventions, and strengthen the evaluation of AI-related educational practices. The goal is not to discourage AI use, but to promote a reflective and critical relationship with the technology. By equipping educators and institutions with appropriate conceptual and diagnostic tools, education systems can better prepare students to navigate AI-mediated environments responsibly.

Given that this study is conceptual in nature, a key limitation lies in the absence of empirical validation of the proposed framework. Future research should therefore prioritise conducting a pilot study of the AI hallucination risk assessment instrument among secondary school students. Such a pilot test would enable the examination of item clarity, internal consistency, and construct validity, and provide preliminary evidence of the instrument's reliability and suitability for the targeted population. Findings from the pilot phase would inform scale refinement prior to large-scale implementation and support the development of evidence-based interventions aimed at strengthening digital immunity.

Corresponding Author

Mohd Effendi @ Ewan Mohd Matore (Ph. D)

Faculty of Education, National University of Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

Email: effendi@ukm.edu.my

ACKNOWLEDGMENT

We gratefully acknowledge financial support from Dana Penyelidikan SDG FPEND 2024 (GG-2024-044) by Faculty of Education, UKM. We thank everyone who is provided with insight and expertise that greatly assisted the research. We thank all my friends in Measurement and Evaluation course for assistance with constructive comments that greatly improved the manuscript.

REFERENCES

1. Ciubotaru, B. I. (2025). The hallucination problem in Generative Artificial Intelligence: accuracy and trust in digital learning. Proceedings of the International Conference on Virtual Learning, 20, 35–45. https://doi.org/10.58503/icvl-v20y202503
2. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., … Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71(March). https://doi.org/10.1016/j.ijinfomgt.2023.102642
3. Elsayed, H. (2024). The Impact of Hallucinated Information in Large Language Models on Student Learning Outcomes: A Critical Examination of Misinformation Risks in AI-Assisted Education. Northern Reviews on Algorithmic Research, Theoretical Computation, and Complexity, 9(8), 1–13. Retrieved from https://northernreviews.com/index.php/NRATCC/article/view/2024-08-07
4. Erümit, A. K., and Sarıalioğlu, R. Ö. (2025). Artificial intelligence in science and chemistry education: a systematic review. Discover Education, 4(1). https://doi.org/10.1007/s44217-025-00622-3
5. Fulsher, A., Pagkratidou, M., and Kendeou, P. (2025). GenAI and misinformation in education: a systematic scoping review of opportunities and challenges. AI and Society. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00146-025-02536-y
6. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., … Fung, P. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12). https://doi.org/10.1145/3571730
7. Kahneman, D. (2011). Thinking, Fast and Slow (1st Editio). United States: Farrar, Straus and Giroux (FGS).

8. Kiili, C., Bråten, I., Strømsø, H. I., Hagerman, M. S., Räikkönen, E., and Jyrkiäinen, A. (2022). Adolescents' credibility justifications when evaluating online texts. Education and Information Technologies, 27(6), 7421–7450. https://doi.org/10.1007/s10639-022-10907-x
9. Milmo, D., and Hern, A. (2024, March 8). 'We definitely messed up': why did Google AI tool make offensive historical images? The Guardian, p. 1. Retrieved from https://www.theguardian.com/technology/2024/mar/08/we-definitely-messed-up-why-did-google-ai-tool-make-offensive-historical-images
10. Ministry Of Education (MOE) Malaysia. (2023). Dasar Pendidikan Digital. Ministry Of Education (MOE) Malaysia, pp. 1–82. Kuala Lumpur: Ministry of Education, Malaysia (MoE). Retrieved from https://www.moe.gov.my/dasarmenu/dasar-pendidikan-digital
11. Slovic, P. (1987). Perception of Risk. Science, 236(4799), 280–285. https://doi.org/https://doi.org/10.1126/science.3563507
12. Sparkes, M. (2023, February). Google Bard advert shows new AI search tool making a factual error. New Scientist, 1. Retrieved from https://www.newscientist.com/article/2358426-google-bard-advert-shows-new-ai-search-tool-making-a-factual-error/
13. UNESCO. (2023). Guidance for generative AI in education and research. In Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/ewzm9535
14. Valeri, F., Nilsson, P., and Cederqvist, A. M. (2025). Exploring students' experience of ChatGPT in STEM education. Computers and Education: Artificial Intelligence, 8, 1–15. https://doi.org/10.1016/j.caeai.2024.100360
15. Zhang, H., Perry, A., and Lee, I. (2025). Developing and Validating the Artificial Intelligence Literacy Concept Inventory: an Instrument to Assess Artificial Intelligence Literacy among Middle School Students. International Journal of Artificial Intelligence in Education, 35(1), 398–438. https://doi.org/10.1007/s40593-024-00398-x