Auditable AI Literacy Interventions: Embedding Regulatory Principles into Higher Education

Edisy Kin Wai Chan
School of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom
k.w.chan@soton.ac.uk

Beatrice Yan-yan Dang
School of Education and Childhood
University of the West of England
Bristol, United Kingdom
yan2.dang@live.uwe.ac.uk

Abstract

In recent years, artificial intelligence (AI) has become an integral part of education, work, and governance, making AI literacy a critical competency for higher education. Yet, in today's higher education landscape, courses and programmes involving AI literacy tend to focus primarily on teaching knowledge and skills while overlooking a crucial element: *auditability*—the capacity to document, assess, and demonstrate responsible AI use in ways that align with regulatory standards. In this paper, we introduce the concept of *Auditable AI Literacy Interventions*, which incorporate audit instruments into AI literacy education to parallel standard regulatory practices such as conformity assessments, provenance tracking, and oversight structures. We outline a conceptual framework for designing these interventions, propose practical tools for classroom use, and illustrate how they can be integrated into tertiary level course modules. The main contribution of this work is to reconceptualize AI literacy: it should serve not only as an educational objective but also as a means of preparing institutions for regulatory compliance, thereby aligning higher education with emerging standards for regulatable machine learning.

1 Introduction

As AI systems become increasingly integrated into everyday life, their governance has emerged as a pressing societal challenge [18, 21, 35]. Recent regulatory initiatives, such as the EU AI Act [9] and other emerging national policies [4], emphasize the principles of transparency, accountability, and auditability in AI development and deployment. While these requirements are typically directed at industry and research, higher education plays a critical role in preparing future practitioners, policymakers, and citizens to operate in a regulatable AI ecosystem.

AI literacy, defined as the knowledge and critical capacities required to understand, evaluate, and responsibly use AI systems, has become an urgent educational priority [1, 7, 16, 19]. Yet most interventions focus on skill acquisition or ethical awareness without addressing auditability—the ability to generate evidence of conformance, traceability, and oversight [22]. Without embedding such practices, educational interventions risk producing learners who are knowledgeable but unprepared for regulatory realities.

In this paper, we propose the concept of *Auditable AI Literacy Interventions*, in which curriculum designs integrate audit instruments that mirror common regulatory practices such as conformity assessments [11, 29], provenance tracking [14, 30], and oversight structures [13, 33]. We contribute:

Workshop on Regulatable ML at the 39th Conference on Neural Information Processing Systems (NeurIPS 2025).

1. A conceptual framework for defining audit objectives and instruments in AI literacy education;
2. An illustrative example of how these instruments can be embedded into a higher education course module;
3. A discussion of implications for educators, students, and policymakers in aligning AI education with regulatable machine learning.

Our aim is to position AI literacy not only as an educational goal but also as a regulatory readiness strategy, ensuring that higher education contributes directly to the broader challenge of building trustworthy and accountable AI systems.

2 Auditable AI Literacy Intervention Framework

In this section, we present the concept of *Auditable AI Literacy Interventions*, highlighting how curriculum designs can embed systematic auditing mechanisms aligned with standard regulatory requirements, including conformity assessments, provenance tracking, and oversight structures. The framework is structured around four components: audit objectives, audit instruments, curriculum integration, and regulatory alignment.

2.1 Audit Objectives

We define AI literacy auditability across four key dimensions, derived from prior AI literacy frameworks [5, 8, 16, 23, 26, 27, 34]:

1. **Awareness**: Understanding the capabilities, risks, and regulatory context of AI.
2. **Ethics**: Ability to recognize potential harms, bias, and accountability of AI systems.
3. **Evaluation**: Critical capacity to assess the reliability and limits of AI outputs.
4. **Use**: Practical competence in deploying AI tools responsibly.

2.2 Audit Instruments

To operationalize these objectives, we propose the following instruments:

- **Pre/Post Surveys**: Surveys adapted or modified from established instruments—–including MAILS [3, 17], AICOS [20], GLAT [15], ABCE [28], and AIERS [36]—–to measure students' AI literacy growth across the dimensions of awareness, ethics, evaluation, and use (Table 1).
- **Reflective Logs**: Structured templates—drawing inspiration from documentation practices such as Datasheets for Datasets [12], FactSheets [2], and Model Cards [24]—that guide students in documenting their interactions with AI tools, critically identifying observed risks (e.g., bias, hallucination), and outlining corresponding mitigation strategies. These logs act as *provenance records*.
- **Assignment Rubrics**: Derived from the AIAS scale [31] and the Transparency Index Framework [6], these rubrics assess coursework across multiple dimensions, including the quality of content, transparency in the use of AI tools, ethical reflection, and completeness of documentation.
- **Peer/Faculty Audit Checklists**: Structured review forms grounded in principles of AI auditing [10, 25], designed to simulate third-party audits by evaluating whether AI tool usage was transparently disclosed, potential risks were adequately addressed, and regulatory or ethical implications were taken into account [32].

2.3 Curriculum Integration

Audit checkpoints can be embedded within standard course modules:

1. **Pre-course**: Survey to establish baseline AI literacy and regulatory awareness.

Table 1: Representative survey instruments mapped to AI literacy dimensions

Dimensions	Associated Instruments
Awareness	MAILS Understand AI, AICOS Understanding AI, ABCE Cognitive, GLAT Know & Understand
Ethics	MAILS AI Ethics, AICOS Ethics AI, ABCE Ethical, GLAT Ethics, AIERS
Evaluation	MAILS Create AI, AICOS Create AI, ABCE Cognitive, GLAT Evaluate & Create
Use	MAILS Apply AI, AICOS Apply AI, GLAT Use & Apply

2. **Mid-course**: Reflective logs and assignments with disclosure requirements.
3. **End-course**: Post-survey combined with a peer audit exercise, providing traceable evidence of growth and oversight.

2.4 Regulatory Alignment

This framework reflects regulatory mechanisms that are fundamental to regulatable machine learning:

- **Conformity Assessments**: Pre/post surveys function as educational analogues of conformance testing [11].
- **Provenance Tracking**: Reflective logs parallel documentation requirements for high-risk AI systems [30].
- **Oversight Structures**: Peer/faculty checklists simulate external audits and human-in-the-loop oversight [33].

Overall, this framework positions AI literacy interventions as auditable, conformance-aware structures that prepare higher education institutions and learners for participation in a regulatable AI ecosystem.

3 Illustrative Example

To demonstrate feasibility, we sketch the design of a short AI literacy module within an undergraduate social science course. The module lasts two weeks and introduces students to both the technical and regulatory dimensions of generative AI systems.

3.1 Learning Activity

Students are asked to use a large language model (LLM) to generate a short essay on the topic of "Social Mobility in Education". The goal is not only to familiarize students with generative AI, but also to develop their ability to critically evaluate, document, and reflect on the use of such tools.

3.2 Audit Instruments in Practice

The proposed audit instruments are embedded throughout the module:

- **Pre-survey**: Prior to the activity, students complete a brief AI literacy questionnaire, developed in accordance with the illustrative instruments in Table 1, to establish a baseline across the four dimensions of AI literacy: Awareness, Ethics, Evaluation, and Use.
- **Reflective log**: During the task, students complete a structured log template documenting the AI tools they used, the prompts issued, and any risks observed (e.g., biased examples, hallucinated references, or lack of transparency in outputs). The log also includes a section where students align their observations with relevant regulatory categories such as transparency and accountability.
- **Assignment rubric**: The essay is evaluated on both content quality and auditability, requiring students to disclose their AI usage, justify the role of the AI system in their workflow, and critically reflect on its limitations.

- **Peer audit checklist**: Each student's work is reviewed by a peer using a structured checklist, including items such as *"Was AI use properly disclosed?"*, *"Were risks documented?"*, and *"Did the reflection address regulatory implications?"*. This process is designed to simulate a third-party audit.
- **Post-survey**: At the conclusion of the module, students complete the same questionnaire administered at the beginning. Comparing pre- and post-survey results provides evidence of growth across the four AI literacy dimensions and serves as a conformity-style assessment.

3.3 Expected Outcomes

This example illustrates how auditability can be embedded into AI literacy interventions without requiring significant curricular overhaul. The process generates traceable evidence of learning, familiarizes students with conformance-style documentation, and cultivates awareness of AI risks in a manner aligned with regulatory expectations. While conceptual, such a module provides a template for empirical pilot studies in higher education settings.

4 Discussion and Conclusion

This paper introduced the concept of *Auditable AI Literacy Interventions*, arguing that AI education in higher education should cultivate competencies while integrating mechanisms for auditability, transparency, and regulatory alignment. By framing AI literacy development through the lens of regulatable machine learning, we highlight how educational practices can contribute to broader governance ecosystems.

4.1 Implications

The proposed framework suggests several implications:

- **For educators**: Audit instruments provide practical ways to integrate regulatory principles into existing curricula without extensive redesign. They can also serve as scaffolds for assessment, helping instructors move beyond evaluating technical proficiency to evaluating responsible and transparent use of AI.
- **For students**: Exposure to audit-style practices fosters both technical competence and regulatory awareness, preparing learners for future conformance contexts. It also encourages habits of documentation and reflection that are increasingly expected in professional AI development.
- **For policymakers**: Higher education can act as a site of alignment between AI literacy and regulatory readiness, producing evidence that educational interventions can emulate regulatory tools such as conformity assessments and provenance requirements. Universities adopting such practices could serve as testbeds that inform the evolution of standards and certification schemes.

Beyond higher education, these instruments could also inform professional training, certification schemes, and organizational governance, extending the reach of AI literacy into broader regulatory contexts.

4.2 Limitations and Future Work

Our framework is conceptual and has yet to be empirically validated. Future work should pilot these audit instruments in classroom settings, examine their psychometric reliability, and investigate how students engage with conformance-style practices. Longitudinal research could further explore whether fostering auditability in education translates into greater accountability in professional AI use. At the same time, adoption challenges such as faculty workload, student resistance, or concerns about introducing "audit culture" into learning environments need careful consideration.

4.3 Conclusion

Auditable AI literacy interventions bridge the gap between education and regulation by equipping learners with both knowledge and practices of oversight and accountability. Embedding auditability into AI literacy course modules offers a low-cost, high-value strategy for aligning higher education with the principles of regulatable machine learning. By extending existing AI literacy frameworks with an explicit focus on auditability, this work positions universities as educators of future practitioners while simultaneously preparing institutions and learners for a rapidly evolving regulatory landscape.

References

[1] Omaima Almatrafi, Aditya Johri, and Hyuna Lee. A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, 6:100173, Jun 2024. doi:10.1016/j.caeo.2024.100173. URL https://www.sciencedirect.com/science/article/pii/S2666557324000144.

[2] Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development*, 63(4/5):6–1, Sep 2019. doi:10.1147/JRD.2019.2942288. URL https://ieeexplore.ieee.org/document/8843893.

[3] Astrid Carolus, Martin J Koch, Samantha Straka, Marc Erich Latoschik, and Carolin Wienrich. MAILS-Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1(2):100014, Oct 2023. doi:10.1016/j.chbah.2023.100014. URL https://www.sciencedirect.com/science/article/pii/S2949882123000142.

[4] CCIA. Global Round-Up: National AI Policies. Computer & Communications Industry Association Whitepaper, Mar 2025. URL https://ccianet.org/library/global-round-up-national-ai-policies/. Accessed: 2025-08-21.

[5] Cecilia Ka Yuk Chan. A comprehensive AI policy education framework for university teaching and learning. *International journal of educational technology in higher education*, 20(1):38, Jul 2023. doi:10.1111/bjet.13411. URL https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1111/bjet.13411.

[6] Muhammad Ali Chaudhry, Mutlu Cukurova, and Rose Luckin. A transparency index framework for AI in education. arXiv:2206.03220 [cs.CY], May 2022. doi:10.48550/arXiv.2206.03220. URL https://arxiv.org/abs/2206.03220.

[7] Thomas K.F. Chiu, Zubair Ahmad, Murod Ismailov, and Ismaila Temitayo Sanusi. What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6:100171, Jun 2024. doi:10.1016/j.caeo.2024.100171. URL https://www.sciencedirect.com/science/article/pii/S2666557324000120.

[8] Mutlu Cukurova, Fengchun Miao, et al. *AI competency framework for teachers*. UNESCO Publishing, 2024. doi:10.54675/ZJTE2084. URL https://unesdoc.unesco.org/ark:/48223/pf0000391104.

[9] European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). The European Parliament and the Council of the European Union, Jul 2024. URL https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Accessed: 2025-08-21.

[10] Linda Fernsel, Yannick Kalff, and Katharina Simbeck. Assessing the Auditability of AI-integrating Systems: A Framework and Learning Analytics Case Study. arXiv:2411.08906 [cs.CY], Oct 2024. doi:10.48550/arXiv.2411.08906. URL https://arxiv.org/abs/2411.08906.

[11] Luciano Floridi, Matthias Holweg, Mariarosaria Taddeo, Javier Amaya, Jakob Mökander, and Yuni Wen. capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. SSNR, Mar 2022. doi:10.2139/ssrn.4064091. URL https://ssrn.com/abstract=4064091.

[12] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, Nov 2021. doi:10.1145/3458723. URL https://dl.acm.org/doi/10.1145/3458723.

[13] Ellen P. Goodman and Julia Trehu. Algorithmic auditing: Chasing AI accountability. *Santa Clara High Tech. L. J.*, 39:289, May 2022. URL https://digitalcommons.law.scu.edu/chtlj/vol39/iss3/1/.

[14] Melanie Herschel, Ralf Diestelkämper, and Houssem Ben Lahmar. A survey on provenance: What for? What form? What from? *The VLDB Journal*, 26(6):881–906, Oct 2017. doi:10.1007/s00778-017-0486-1. URL https://link.springer.com/article/10.1007/s00778-017-0486-1.

[15] Yueqiao Jin, Roberto Martinez-Maldonado, Dragan Gašević, and Lixiang Yan. GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence*, page 100436, June 2025. dot:10.1016/j.caeai.2025.100436. URL https://www.sciencedirect.com/science/article/pii/S2666920X25000761.

[16] Michelle Kassorla, Maya Georgieva, and Allison Papini. AI Literacy in Teaching and Learning: A Durable Framework for Higher Education. An EDUCAUSE Working Group Paper, Oct 2024. URL https://www.educause.edu/content/2024/ai-literacy-in-teaching-and-learning/executive-summary. Accessed: 2025-08-21.

[17] Martin J Koch, Astrid Carolus, Carolin Wienrich, and Marc E Latoschik. Meta AI literacy scale: Further validation and development of a short version. *Heliyon*, 10(21), Oct 2024. doi:10.1016/j.heliyon.2024.e39686. URL https://www.cell.com/heliyon/fulltext/S2405-8440(24)15717-9.

[18] Alex Krasodomski, Arthur Gwagwa, Brandon Jackson, Elliot Jones, Stacey King, Mira Lane, Micaela Mantegna, Thomas Schneider, Kathleen Siminyu, and Alek Tarkowski. Artificial intelligence and the challenge for global governance: Nine essays on achieving responsible AI. Chatham House - The Royal Institute of International Affairs, Jun 2024. URL https://apo.org.au/node/327142. Accessed: 2025-08-21.

[19] Matthias Carl Laupichler, Alexandra Aster, Jana Schirch, and Tobias Raupach. Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3:100101, Sep 2022. doi:10.1016/j.caeai.2022.100101. URL https://www.sciencedirect.com/science/article/pii/S2666920X2200056X.

[20] Sarah Markus, Astrid Carolus, and Carolin Wienrich. AICOS: Towards an Objective Scale of AI Literacy for Generative AI. arXiv:2503.12921 [cs.CL], Mar 2025. doi:10.48550/arXiv.2503.12921. URL https://arxiv.org/abs/2503.12921.

[21] Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, et al. Artificial intelligence index report 2025. arXiv:2504.07139 [cs.AI], Jul 2025. doi:10.48550/arXiv.2504.07139. URL https://arxiv.org/abs/2504.07139.

[22] Bahar Memarian and Tenzin Doleck. Fairness, accountability, transparency, and ethics (fate) in artificial intelligence (ai) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 5:100152, Jun 2023. doi:10.1016/j.caeai.2023.100152. URL https://www.sciencedirect.com/science/article/pii/S2666920X23000310.

[23] Fengchun Miao, Kelly Shiohira, et al. *AI competency framework for students*. UNESCO Publishing, 2024. doi:10.54675/JKJB9835. URL https://unesdoc.unesco.org/ark:/48223/pf0000391105.

[24] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, Jan 2019. doi:10.1145/3287560.3287596. URL `https://dl.acm.org/doi/10.1145/3287560.3287596`.

[25] Jakob Mökander. Auditing of AI: Legal, ethical and technical approaches. *Digital Society*, 2(3):49, Nov 2023. doi10.1007/s44206-023-00074-y. URL `https://link.springer.com/article/10.1007/s44206-023-00074-y`.

[26] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2:100041, Nov 2021. doi:10.1016/j.caeai.2021.100041. URL `https://www.sciencedirect.com/science/article/pii/S2666920X21000357`.

[27] Davy Tsz Kit Ng, Wenjie Wu, Jac Ka Lok Leung, Thomas Kin Fung Chiu, and Samuel Kai Wah Chu. Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, 55(3):1082–1104, Dec 2024. doi:10.1111/bjet.13411. URL `https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1111/bjet.13411`.

[28] Davy Tsz Kit Ng, Wenjie Wu, Jac Ka Lok Leung, Thomas Kin Fung Chiu, and Samuel Kai Wah Chu. Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, 55(3):1082–1104, May 2024. doi:10.1111/bjet.13411. URL `https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1111/bjet.13411`.

[29] NIST. Conformity Assessment Basics. National Institute of Standards and Technology, Feb 2023. URL `https://www.nist.gov/standardsgov/conformity-assessment-basics`. Accessed: 2025-08-21.

[30] Gabriele Padovani, Valentine Anantharaj, and Sandro Fiore. Provenance Tracking in Large-Scale Machine Learning Systems. arXiv:2507.01075 [cs.LG], July 2025. doi:10.48550/arXiv.2507.01075. URL `https://arxiv.org/abs/2507.01075`.

[31] Mike Perkins, Leon Furze, Jasper Roe, and Jason MacVaugh. The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6):49–66, Apr 2024. doi:10.53761/q3azde36. URL `https://open-publishing.org/journals/index.php/jutlp/article/view/810`.

[32] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 33–44, Jan 2020. doi:10.1145/3351095.3372873. URL `https://dl.acm.org/doi/abs/10.1145/3351095.3372873`.

[33] Chris Schmitz, Jonathan Rystrøm, and Jan Batzner. Oversight Structures for Agentic AI in Public-Sector Organizations. In *Proceedings of the 1st Workshop for Research on Agent Language Models (REALM 2025)*, pages 298–308, Vienna, Austria, Jul 2025. Association for Computational Linguistics. doi:10.18653/v1/2025.realm-1.21. URL `https://aclanthology.org/2025.realm-1.21/`.

[34] Jane Southworth, Kati Migliaccio, Joe Glover, Ja'Net Glover, David Reed, Christopher McCarty, Joel Brendemuhl, and Aaron Thomas. Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, 4:100127, Jan 2023. doi:10.1016/j.caeai.2023.100127. URL `https://www.sciencedirect.com/science/article/pii/S2666920X23000061`.

[35] Araz Taeihagh. Governance of Generative AI. *Policy and Society*, 44(1):1–22, Feb 2025. doi:10.1093/polsoc/puaf001. URL `https://doi.org/10.1093/polsoc/puaf001`.

[36] Ziying Wang, Ching-Sing Chai, Jiajing Li, and Vivian Wing Yan Lee. Assessment of AI ethical reflection: the development and validation of the AI ethical reflection scale (AIERS) for university students. *International Journal of Educational Technology in Higher Education*, 22(1):19, Mar 2025. doi:10.1186/s41239-025-00519-z. URL `https://link.springer.com/article/10.1186/s41239-025-00519-z`.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: **The abstract has clearly captured all the main points made by this work, and the introduction has made a comprehensive outline for the whole paper.**

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

Justification: **Section 4.2 has discussed the limitations of this work.**

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: **This paper is a conceptual paper so it does not include mathematical results.**

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: **This paper is a conceptual paper so it does not include any experiments.**

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

   Justification: **This paper is a conceptual paper so it does not include any experiments requiring code.**

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: **This paper is a conceptual paper so it does not include any experiments.**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: **This paper is a conceptual paper so it does not include any experiments.**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: **This paper is a conceptual paper so it does not include any experiments.**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: **The authors has reviewed the NeurIPS Code of Ethics and confirm that the research conducted in this paper conform the Code of Ethics.**

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: **Section 4.1 has discussed the potential positive societal impacts of this work, and the authors believe that this work is unlikely to have any negative societal impacts.**

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

Justification: **This paper is a conceptual paper so it poses no such risks.**

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: **This paper is a conceptual paper so it does not use any existing assets.**

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: **This paper is a conceptual paper so it does not release any new assets.**

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: **This paper is a conceptual paper so it does not involve any crowdsourcing nor research with human subjects.**

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: **This paper is a conceptual paper so it does not involve any crowdsourcing nor research with human subjects.**

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: **The core method development in this paper does not involve LLMs as any important, original, or non-standard components.**

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.