Available Online: 20 July 2026 DOI: 10.54254/3049-7248/2026.35571

Task-contingent risk configurations of Generative Artificial Intelligence in education: a URL-clustered analysis of a risk-enriched multi-platform corpus

Na Zhao

Harbin University of Commerce, Harbin, China

zhaona@hit.edu.cn

Abstract. Generative Artificial Intelligence (GenAI) has raised concerns about academic integrity, cognitive dependence, professional roles, educational quality, and equitable participation. Research tends to report these concerns as a unified list of risks, even though their educational significance varies with the task in which AI is envisaged or applied. This study analysed a separately curated, risk-inflated corpus of 3,000 public-text records from six Chinese social media platforms and and 300 source Uniform Resource Locators (URLs). Five non-exclusive risk constellations and five educational task contexts were reconstructed from indexed keyword fields. Binary logistic generalised estimating equations clustered by source URL modelled task associations while adjusting for platform and text length. Cognitive-reliance language formed 35.5% of the corpus, academic-integrity language 22.7%, and teacher-role disruption 18.4%. Assignment/writing and assessment/feedback contexts were strongly associated with academic-integrity language (OR = 3.59, 95% CI [2.86, 4.51]; OR = 3.56 [2.90, 4.37]). Teacher-work contexts were associated with teacher-role disruption (OR = 1.39 [1.10, 1.74]). Variation between platforms was slight for four configurations; teaching-quality concern exhibited modest distinction. The results advocate task-based governance that combines process evidence with assessment, verification with teacher judgment, and access support with institutional policy. Since the corpus was intentionally risk-inflated, its percentages characterise semantic composition rather than platform-user prevalence.

Keywords: generative artificial intelligence, academic integrity, cognitive reliance, teacher digital competence, social media research

1. Introduction

Artificial intelligence has been introduced to higher education as a general-purpose model that is distributed across the institution and students in various ways (the use of the systems), through courses and other informal practices. The pedagogical purpose, human responsibility, data protection, and fair access have now become the focus of international guidance [1, 2]. There are also similar research reviews that record mixed outcomes; GenAI can be used to help with idea generation, explanation, feedback, preparation of resources, and customized learning and unreliable results, lack of clarity on origin, cheating, and unpreparedness are among

the issues that have continued to exist [3-6]. This educational question thus involves the way the activities will be arranged for these systems as well as the technical capacity of the systems.

The existing discussions tend to condense qualitatively different concerns into a single list of risks. The academic-integrity issues are the evidentiary relationship between the submitted work and student learning [7-11]. Cognitive-reliance worries include long-term shifts in attention, memory, reasoning, monitoring, and readiness to deal with challenging material [12, 13]. Professional issues deal with the authority, workload, expertise, and continuity of roles of teachers. The quality issues involve reliability of instruction and adequacy of feedback or evaluation while equity issues concern differential access, support, and family capacity. Every issue has a distinct educational objective. A composite score can hide such differences and generate general recommendations that apply to no specific task.

Task structure is a more discriminative unit of analysis. The written tasks and tests should have proof that the learner is capable of doing certain intellectual work. Classroom interaction needs contingent explanation, diagnosis, and response. Teacher preparation entails curriculum judgment, resource selection, and accountability on learning environment. Student support consists of explanation, practice, and feedback in time. AI use has different educational meaning in these contexts, as the observable product, the process of learning, and the responsible evaluator differ. Assessment reform has therefore focused on program-level assurance, authentic evidence, and designs that allow student reasoning to be reviewed [7, 8, 10].

The interpretation of cognitive support is also very important. Offloading on the cognitive level has been found to be a useful way of minimizing the amount of unnecessary work and freeing up space to do more meaningful work, but this can also lead to offloading not being properly regulated, which will result in poor metacognitive calibration as well as less attention paid to the process that was meant to take place during the task [12, 13]. The issue of dependence among people should be analyzed at the level of relations between tasks and not with regard to the use of technology in general. This rule holds good in the case of professional positions. The competence of teacher agency, moral judgment, pedagogical application, and professional learning are positioned by human-centered AI systems in terms of responsible use [2, 14, 15].

It is revealed that in the recent research of students and teachers, the participants were able to identify brainstorming, drafting, feedback, assessment, and policy expectations during their discussions with GenAI [16, 17]. The wider Artificial Intelligence in Education (AIEd) literature also makes a distinction between learner-oriented, teacher-oriented, and system-oriented applications [18-20]. It can be inferred that there should be a grouping of risk language as it relates to known forms of educational activities. They too imply that the variation at the platform level will not be a primary consideration where the physical arrangement of the work is clear.

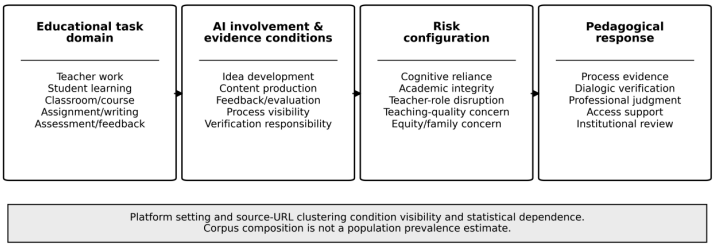

Figure 1. Task-contingent analytical architecture for interpreting GenAI-related risk language

This study investigates a risk-enriched multi-platform corpus provided by the project team. The public language does not provide a direct indication of actual misconduct, cognitive change, employment outcomes, and instructional quality. It does, however, demonstrate how educational anxieties are attached to tasks and

roles. The analysis poses three questions. RQ1 recognizes the composition and co-occurrence of five risk configurations. RQ2 predicts the association between five domains of educational tasks with each configuration when platform and text length are fixed. RQ3 is an assessment of whether there is still significant variation in platforms even after the analysis acknowledges the nesting of records within source URLs. Figure 1 gives the conceptual and analytical sequence.

2. Materials and methods

2.1. Study design and corpus

The risk-enriched subset contained 500 records per platform and was linked to the same 300 source URLs represented in the merged workbook, with 50 URLs per platform. The enrichment process makes category composition observable but prevents inference about the prevalence of risk language among platform users or source items. The analysis therefore treats the 3,000 records as a structured semantic corpus. It does not interpret percentages as population estimates, and it does not use engagement counts, user attributes, or timestamps.

2.2. Corpus audit and semantic construction

The risk configuration of 22 risk keywords was created with the help of five nonexclusive configurations, consistent with the five core concern dimensions identified in existing discourse analysis of GenAI-related educational risks on social media [21] . The reliance on cognition was manifested in the form of overreliance, a lack of independent thought, and inability to think critically, as well as compulsive behavior. Cheating, examination misconduct, plagiarism, ghostwriting, and cheating were among the academic integrity. The teacher role disturbance was also disruption of the teachers, unemployment, reduction of staff, and reassignment. Teaching-quality worry was a combination of general risk vocabulary and the explicit language that the teaching quality is deteriorating. Fairness, educational equity, and family anxiety were part of the concern about equity and family. It was based on the reconstruction of 34 education-related keywords: the task of the teacher, the learning process of the student, the activity of the classroom/course, writing/assignment, and the assessment/feedback. Table 1 provides the frequencies.

2.3. Statistical analysis

The analysis then made a record level and equal-URL category proportion. Pearson chi-square statistics were summarized, as well as Cramers V unadjusted variation of the platforms. The five binary logistic Generalized Estimating Equation (GEE) models subsequently considered each risk arrangement to be an independent result [22]. Models applied working correlation of exchangeable nature and source URL as clustering unit. All specifications consisted of 5 platform indicators, 5 task-domain indicators and standardization of Unicode character length. Bilibili was the one that was taken as the reference platform in terms of coefficient reporting; the substantive findings are based on joint tests of the platforms instead of being based on that choice of the reference.

Equation (1) defines the population-averaged specification. For record i nested in source URL j and risk configuration r, P denotes platform indicators, T denotes task-domain indicators, and Z denotes standardized text length.

$$\operatorname{logit}\{\Pr(Y_{ijr}=1)\}=\beta_{0r}+\sum_{p=1}^{5}\beta_{pr}P_{ijp}+\sum_{q=1}^{5}\gamma_{qr}T_{ijq}+\delta_{r}Z_{ij} \quad (1)$$

2.4. Data protection and ethical safeguards

This research constructs a risk-enriched corpus consisting of 3,000 annotated entries, whose detailed composition covering risk configuration and task domain is illustrated in Table 1. In this research there is no mention of platform text that can be quoted verbatim, username, account identifier, location, date and time, or a source Uniform Resource Locators (URLs). The research does not involve the collection of personal sensitive information including age, sex, race, health status etc. It abides by the principles of data minimization, purpose limitation, limited redistribution and non-re-identification in accordance with the *Personal Information Protection Law* and *Data Security Law of the People's Republic of China* as well as the *General Data Protection Regulation* [23-28].

Table 1. Composition of the risk-enriched corpus (N = 3,000)

Construct	Recorded scope	*n*	% of records
Risk configuration	Cognitive reliance	1,065	35.5
	Academic integrity	680	22.7
	Teacher-role disruption	553	18.4
	Teaching-quality concern	462	15.4
	Equity and family concern	453	15.1
Task domain	Teacher work	2,097	69.9
	Student learning	1,367	45.6
	Classroom/course	738	24.6
	Assignment/writing	701	23.4
	Assessment/feedback	738	24.6

Note: Risk configurations and task domains are nonexclusive. Percentages therefore do not sum to 100%.

3. Results

3.1. Corpus composition

The 3,000 records were divided into six sources of platforms and placed in the 300 source URLs. The mean record length was 61.66 Unicode characters (SD = 22.74). Teacher-work terms occurred in 2,097 records (69.9%), student-learning terms in 1,367 (45.6%), classroom/course terms in 738 (24.6%), assignment/writing terms in 701 (23.4%), and assessment/feedback terms in 738 (24.6%). Task domains were not exclusive, as a single entry might refer to a teacher, an examination, and an assignment.

The highest percentage of the risk configuration was cognitive reliance, which was recorded in 1,065 records (35.5%). The academic integrity language occurred at 680 records (22.7%), teacher-role disruption in 553 (18.4%), teaching-quality concern in 462 (15.4%), and equity/family concern in 453 (15.1%). Most records carried a single configuration: 2,790 records had one category, 207 had two, and three had three. Equal-URL estimates closely reproduced the record-level composition (35.1%, 22.5%, 18.8%, 14.8%, and 15.3%, respectively), which reduces concern that a small number of highly represented source items generated the overall ordering.

3.2. Task-contingent associations

The combined task-domain test was also important to the reliance of cognition, integrity in learning and teacher role disturbance (all $p < 0.001$), whereas teaching-quality concern showed no task-specific differentiation ($p = 0.571$). The equity/family model produced a weaker joint pattern ($p = 0.085$). Table 2 is a table of adjusted odds ratios and Figure 2 are three sets of configurations that have the most well-organized tasks.

Table 2. Adjusted task-domain associations with five risk configurations

Risk configuration	Teacher work	Student learning	Classroom/ course	Assignment/ writing	Assessment/ feedback
Cognitive reliance	0.58 [0.49, 0.68]*	0.85 [0.73, 0.99]	0.74 [0.61, 0.89]*	0.60 [0.50, 0.73]*	0.65 [0.54, 0.78]*
Academic integrity	1.16 [0.97, 1.39]	1.16 [0.97, 1.39]	1.91 [1.54, 2.37]*	3.59 [2.86, 4.51]*	3.56 [2.90, 4.37]*
Teacher-role disruption	1.38 [1.10, 1.74]*	0.89 [0.72, 1.08]	0.80 [0.65, 1.00]	0.60 [0.47, 0.75]*	0.64 [0.51, 0.81]*
Teaching-quality concern	0.96 [0.74, 1.25]	1.02 [0.81, 1.28]	1.01 [0.80, 1.26]	0.81 [0.65, 1.02]	1.00 [0.76, 1.31]
Equity and family concern	0.89 [0.71, 1.12]	0.88 [0.68, 1.12]	0.99 [0.77, 1.27]	0.80 [0.62, 1.02]	0.67 [0.50, 0.89]*

Note: Cells report odds ratio [95% confidence interval] from URL-clustered GEE models controlling for platform and standardized text length. * Benjamini-Hochberg $q < 0.05$ across the 25 task coefficients.

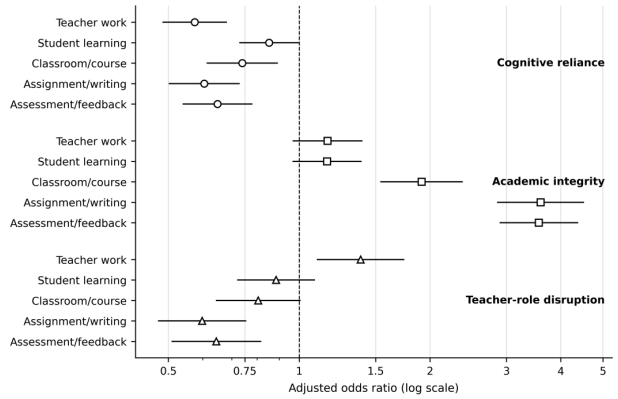

Figure 2. Adjusted task associations from source-URL-clustered GEE models; error bars show 95% confidence intervals

The language of academic integrity was found to be focused on the work in which the results were usually taken as an evidence of learning. The odds of assignment/writing records and assessment/feedback records with 3.59 times higher than the odds of academic-integrity language (95% Confidence Interval (CI) [2.86, 4.51]), and assessment/feedback records had 3.56 times the odds [2.90, 4.37]. Classroom/course records also showed a positive association (Odds Ratio (OR) = 1.91 [1.54, 2.37]). The other task domains had no effect on teacher-work and student-learning indicators.

The teacher-work contexts had a positive effect on the disruption of the teachers in their role (OR = 1.39 [1.10, 1.74]). Assignment/writing and assessment/feedback contexts were inversely associated with that configuration (OR = 0.60 [0.48, 0.75] and 0.65 [0.51, 0.82]). There was a different pattern of cognitive-reliance language. It is worth noting that explicit teacher-work, classroom/course, assignment/writing and assessment/feedback indicators were related to reduced odds of the risk-enriched corpus having cognitive-reliance words. The results are such that it can be assumed that the records with more general educational references would contain the concern of cognitive reliance rather than the one which is fixed to a specific task. The adjustment of false-discovery-rate over the 25 coefficients of tasks maintained the key relations discussed above.

3.3. Platform variation and sensitivity analysis

Unadjusted platform variation was small for cognitive reliance (Cramér's V = 0.017), academic integrity (0.030), teacher-role disruption (0.044), and equity/family concern (0.045). Their joint platform tests were also nonsignificant in the GEE models. Teaching-quality concern showed a modest platform pattern (V = 0.099; joint Wald $\chi^2(5) = 26.96$, $p < 0.001$), with Zhihu showing higher adjusted odds than Bilibili (OR = 1.53 [1.10, 2.12]). Table 3 summarizes the joint tests.

Two sensitivity tests were conducted to test model dependence. The point estimates and confidence intervals were almost identical when an independence working correlation was used. Another analysis limited the corpus to the 2,790 single-configuration records. The intense assignment/assessment correlations with academic integrity and teacher-work correlation with role disruption were also in the same direction and of similar magnitude. General quality and equity models were not as stable under a single-label restriction, which is consistent with their diffuse and partly overlapping lexical content.

Table 3. Joint task and platform tests by risk configuration

Risk configuration	Task Wald $\chi^2(5)$	Task p	Platform Wald $\chi^2(5)$	Platform p	Cramér's V
Cognitive reliance	66.17	< 0.001	2.27	0.810	0.017
Academic integrity	185.70	< 0.001	3.76	0.585	0.030
Teacher-role disruption	44.67	< 0.001	8.63	0.125	0.044
Teaching-quality concern	3.85	0.571	26.96	< 0.001	0.099
Equity and family concern	9.68	0.085	8.30	0.140	0.045

Note: Cramér's V summarizes unadjusted platform variation; Wald tests are from the adjusted URL-clustered GEE models.

4. Discussion

The results indicate that risk language was better structured by educational task than platform. The most prominent trend was related to academic integrity. Assignment and assessment records were significantly more likely to have integrity-related wording, despite the fact that platform, other task domains, text length and source-item clustering had been controlled. This trend can be interpreted as an evidentiary approach to assessment: interest is heightened when a finished product is supposed to be presented as evidence of student ability [7-11]. A pragmatic reply thus demands something more than identification. Course teams are able to gather staged drafts, oral explanation, version histories, annotated decisions or supervised performance in cases where such forms of evidence are relevant to the learning outcome. These actions put reasoning at the disposal of educational judgment and minimize the use of uncertain authorship classifiers [10, 11].

Cognitive-reliance language was the largest proportion of the risk-enriched corpus, but it was not as closely linked to the five explicit task domains. This finding must not be interpreted as proof that specific tasks can prevent learners to rely on cognition. The outcome explains how a language is structured in an enriched corpus. Most records of dependence mentioned thinking, judgment or habitual use in general terms. A helpful interpretation is provided by the literature on cognitive offloading: external support may be educationally useful when learners retain metacognitive control, examine outputs and conduct the reasoning demanded by the curriculum [12, 13]. AI-supported activity can therefore be designed by teachers around explanation, comparison, error diagnosis and reflective account instead of just rapid answer production.

It is a register of professional nature, and teacher-role disruption was created. It had its odds adjusted to the contexts of work of teachers, which raised the chances, whereas in the context of assignments and assessment it lowered them. The corpus put on employment and role words the entire profession such as teaching, preparation, and professional development and not just individual products of students. This register needs to be addressed in the education of the teacher. Modern competence systems focus on people-centered decision-making, AI ethics, pedagogy-based learning, and career education [2, 14, 15]. When training is only an example of how to use tools, then the main problem remains unsolved by the professionals. Instructors must have the practice of making decisions based on contextual information, attention to relationships, disciplinary check, and responsibility in judgment.

There was a lesser degree of task differentiation in the teaching-quality and equity/family issues. They were institutionalized in their language. The quality of instruction was an issue, as well as the access, equity, family power. These problems can not be solved by course-level rules. There should be an infrastructure that is easily accessed, differentiated services, transparent purchase, data management, and review system of the use of AI on different groups. The practice-based and enterprise-related courses have one more requirement of coordination: the teacher and the external mentor need to establish what kind of AI may be used, what evidence needs to be retained by students, and what decisions teachers are supposed to make autonomously.

It was not that much of a platform effect. The concern with the quality of teaching had only moderate differentiation, however, and four risk structures were generally similar on six platforms. This is not an indication that the infrastructure of platforms is irrelevant. That conclusion is restricted by the enriched sampling design and lack of engagement measures. The findings suggest that task relations are more meaningful than platform labels in this collection. The learning can be started as the task, the learning that will be learned, the proof to be made and the one who has to evaluate the task can be taken as a factor of the educational governance process.

There are limitations of the study. The corpus is a risk-enriched corpus and it is not able to make any guesses about the population prevalence or opinions in the populace. It explains the recorded fields of vocabulary, but not the latent attitude, behaviour, learning process, and cause-effect.

5. Conclusion

The research identified five risk situations in a corpus of 3,000 records with risks and their relation to educational activities was simulated and the relationship. The language of academic integrity was grouped around tasks, testing, and classroom/learning activities. It is also common to have language on cognitive reliance, but it is more scattered than most, and issues of quality and equity are only at the institutional level. There were four configurations whose variation was small. These results can be used to advocate for the differentiated governance of GenAI. Evidence of the process should be evident and judgment must be defensible as part of the assessment; learning activity must be metacognitive and verified; professionalism

must be present in the teacher development and technical fluency; institutional policy must be accessible to all people and data practice should be fair. The study cannot determine the frequency of these concerns among the users, but the study will help understand how the discourse about the risk has been related to the actual work organization of education.

References

[1] Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. *UNESCO*. https://doi.org/10.54675/EWZM9535

[2] Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. *UNESCO*. https://doi.org/10.54675/ZJTE2084

[3] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[4] Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, 4. https://doi.org/10.1186/s41239-023-00436-z

[5] Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*, 15. https://doi.org/10.1186/s40561-023-00237-x

[6] Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

[7] Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

[8] Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2), Article 7. https://doi.org/10.53761/1.20.02.07

[9] Eaton, S. E. (2023). Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity, 19*, 23. https://doi.org/10.1007/s40979-023-00144-1

[10] Tertiary Education Quality and Standards Agency. (2023). Assessment reform for the age of artificial intelligence. *Tertiary Education Quality and Standards Agency*. https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence

[11] Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, 26. https://doi.org/10.1007/s40979-023-00146-z

[12] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

[13] Firth, J., Torous, J., Stubbs, B., Firth, J. A., Steiner, G. Z., Smith, L., Alvarez-Jimenez, M., Gleeson, J., Vancampfort, D., Armitage, C. J., & Sarris, J. (2019). The "online brain": How the Internet may be changing our cognition. *World Psychiatry, 18*(2), 119–129. https://doi.org/10.1002/wps.20617

[14] Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education, 32*(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

[15] Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*, 38. https://doi.org/10.1186/s41239-023-00408-3

[16] Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, 43. https://doi.org/10.1186/s41239-023-00411-8

[17] Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education, 20*, 59. https://doi.org/10.1186/s41239-023-00427-0

[18] Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

[19] Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

[20] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, 39. https://doi.org/10.1186/s41239-019-0171-0

[21] Nagar, L. (2024). What is the impact of ChatGPT on education? *International Journal of Scientific Research in Engineering and Management, 8*(4), 1–5. https://doi.org/10.55041/ijsrem32175

[22] Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

[23] Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C. M., & Association of Internet Researchers. (2020). *Internet research: Ethical guidelines* 3.0. https://aoir.org/reports/ethics3.pdf

[24] Fiesler, C., & Proferes, N. (2018). "Participant" perceptions of Twitter research ethics. *Social Media + Society, 4*(1), 1–14. https://doi.org/10.1177/2056305118763366

[25] Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology, 51*(6), 1149–1168. https://doi.org/10.1177/0038038517708140

[26] Standing Committee of the National People's Congress. (2021). *Personal Information Protection Law of the People's Republic of China.* https://en.spp.gov.cn/2021-12/29/c_948419.htm

[27] Standing Committee of the National People's Congress. (2021). *Data Security Law of the People's Republic of China.* https://www.npc.gov.cn/englishnpc/c2759/c23934/202112/t20211209_385109.html

[28] European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). *Official Journal of the European Union, L119*, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng