

Generative AI as a Learning Scaffold for Indonesia's National Criminal Code

Fitri¹, Bambang Afriadi²

Universitas Negeri Jakarta¹, Universitas Islam Syekh Yusuf²

fitri@unj.ac.id

Abstract

Indonesia's National Criminal Code (Kitab Undang-Undang Hukum Pidana, KUHP), enacted through Law No. 1 of 2023, became effective on 2 January 2026, creating a new learning environment in which students must master recently operative statutory rules while generative artificial intelligence (GenAI) is already embedded in everyday academic practice. This study examines the perceived pedagogical utility of GenAI for learning the KUHP among 82 university students studying in Jakarta who were surveyed during January-February 2026. A cross-sectional mixed-methods design combined ten five-point Likert items with three open-ended questions. Seven items measured AI-supported learning utility, two measured source-verification literacy, and one measured overall perceived effectiveness. The seven-item learning-utility scale showed high internal consistency (Cronbach's alpha = .943) and a mean of 3.68 (SD = .76, 95% CI [3.51, 3.85]). Students rated AI most positively for explaining previously difficult KUHP provisions (M = 3.84), while ratings were lower for identifying the legally relevant provision in a case (M = 3.56) and strengthening legal argumentation (M = 3.57). Learning utility correlated strongly with overall perceived effectiveness (Spearman rho = .721, $p < .001$), as did verification literacy (rho = .755, $p < .001$). In an exploratory ordered-logit model, both standardized learning utility (OR = 6.54, $p < .001$) and verification literacy (OR = 8.84, $p < .001$) independently predicted higher overall effectiveness ratings after adjustment for semester group. Open-ended responses emphasized simplified understanding, case application, and faster information access, but also indicated the need for official-source verification. The findings support a bounded pedagogical role for GenAI: it is most defensible as a scaffold for explanation, examples, and formative feedback, not as a substitute for statutory reading or independent legal judgment.

Keywords:

generative artificial intelligence; legal education; criminal law; KUHP; Indonesia; legal reasoning; source verification

DOI: <https://doi.org/10.56442/pef.v4i3.1587>

1. Introduction

The entry into force of Indonesia's new National Criminal Code on 2 January 2026 is a major doctrinal and pedagogical transition. Law No. 1 of 2023 replaces the long-standing colonial-era criminal code and restructures the statutory framework that students must learn, interpret, and apply (Republic of Indonesia, 2023). This transition coincides with the rapid normalization of GenAI tools in higher education. Students can now ask conversational systems to restate statutory language, generate examples, construct hypothetical cases, critique draft answers, and propose alternative legal arguments within seconds. These affordances create a plausible instructional benefit, but they also create a methodological and pedagogical problem: fluent output can be mistaken for legally authoritative analysis.

The higher-education literature increasingly treats GenAI neither as an unqualified benefit nor as a technology that can simply be prohibited. Students commonly report benefits in accessibility, speed, idea generation, and personalized explanation, while simultaneously expressing concerns about accuracy, dependency, academic integrity, and reduced critical engagement (Chan & Hu, 2023; Farrokhnia et al., 2024; Tlili et al., 2023). Experimental and quasi-experimental evidence suggests that ChatGPT-supported interventions can improve some learning outcomes, affective-motivational states, and higher-order thinking indicators,

although effects vary by task design and study quality (Deng et al., 2025). Reviews likewise caution that empirical GenAI research in higher education remains theoretically and methodologically heterogeneous (McGrath et al., 2025).

These tensions are sharper in legal education. Legal reasoning is authority-sensitive: a persuasive answer must be grounded in valid legal sources, distinguish current from superseded rules, identify relevant statutory provisions, and explain how a rule applies to facts. Law-school studies show that GenAI may assist with explanation and drafting while performing less reliably on tasks that require detailed issue spotting, jurisdiction-specific analysis, nuanced interpretation, and evaluative reasoning (Alimardani, 2024; Hargreaves, 2023). Legal educators have therefore argued for assessment designs that preserve substantive legal knowledge, AI-system knowledge, and ethical judgment rather than allowing GenAI to displace the reasoning process (Bliss, 2024; Head & Willis, 2024).

The Indonesian context requires specific evidence rather than assumptions imported from other jurisdictions. Research on Indonesian legal education indicates that students and lecturers are already using ChatGPT for legal research, case analysis, writing, and examinations, while institutional policy, curriculum adaptation, and digital-literacy provision remain uneven (Al Asyari et al., 2025). The introduction of a newly operative KUHP intensifies this challenge because students may encounter AI-generated explanations trained on older legal materials or mixed legal regimes. In this setting, the central pedagogical question is not simply whether students use AI, but which learning tasks they perceive it to support, and whether positive perceptions are accompanied by verification against authoritative legal sources.

Accordingly, this study investigates three questions. First, how useful do students perceive GenAI to be for understanding KUHP provisions, applying them to cases, receiving feedback, and developing legal arguments? Second, how are perceived learning utility and legal-source verification associated with students' overall assessment of AI-supported KUHP learning? Third, what benefits do students describe in their own words, and what instructional model follows logically from those patterns? Because the design is cross-sectional and self-reported, the study does not claim causal learning effects. Its contribution is instead to provide an empirically grounded account of perceived utility at the beginning of the 2026 KUHP era and to translate that evidence into a defensible instructional sequence.

Literature Review

a. *GenAI as a learning scaffold in higher education*

The educational value of GenAI is often explained through its capacity to provide immediate, adaptive, conversational support. Large language models can rephrase difficult material, generate examples, provide formative feedback, and sustain iterative questioning. These functions resemble scaffolding when the technology helps learners perform tasks that would otherwise be difficult while preserving the learner's responsibility for interpretation and judgment. Broader scholarship has identified opportunities for personalization, rapid feedback, and lower barriers to accessing explanations (Halaweh, 2023; Kasneci et al., 2023). Student-perception research similarly shows that learners tend to value GenAI for efficiency, accessibility, and idea generation, even while remaining concerned about reliability and over-dependence (Chan & Hu, 2023).

However, the term *scaffold* is important because it implies support rather than substitution. UNESCO's guidance on GenAI recommends a human-centred approach in which technology augments learning while critical judgment, ethical responsibility, and human agency remain central (Miao & Holmes, 2023). The UNESCO student competency framework further

emphasizes that students should understand AI systems, evaluate outputs critically, and use them responsibly rather than treating AI-generated content as self-validating knowledge (Miao et al., 2024). Such principles align with evidence that educational gains depend strongly on how AI is embedded in a task. A meta-analysis of experimental studies found generally positive learning effects but also noted methodological limitations and the need for objective, complex assessments that distinguish genuine learning from improved output quality (Deng et al., 2025).

b. GenAI in legal education

Legal education introduces domain-specific constraints that make uncritical transfer from general higher-education research inappropriate. Legal knowledge is not merely descriptive; it is structured by hierarchy of authority, jurisdiction, temporal validity, interpretive method, and the application of rules to facts. Early legal-education studies found that ChatGPT can generate plausible legal prose but may perform inconsistently across assessment types. Hargreaves (2023) reported stronger performance on general essay-style questions than on problem-style tasks requiring local issue spotting and application. Alimardani (2024), in a criminal-law examination study, likewise found that GenAI performed below the student average on questions requiring detailed legal and critical analysis.

These findings support a distinction between foundational support and advanced legal judgment. GenAI may be useful for clarifying terminology, restating provisions, generating hypothetical facts, or giving preliminary feedback, but those uses do not establish that the system can reliably identify the controlling rule or produce defensible legal reasoning. Recent legal-pedagogy scholarship therefore recommends guided integration rather than unrestricted delegation. Ajevski et al. (2023) argue that legal education must adapt to AI-enabled professional practice while protecting core learning outcomes. Head & Willis (2024) propose a knowledge framework that combines substantive legal knowledge, GenAI ethics, and system knowledge. Bliss (2024) similarly advocates teaching students to evaluate AI-generated research and writing rather than assuming that prohibition will remain workable.

Emerging empirical work strengthens this position. Hyde-Vaamonde & Keller (2025) found that structured critical engagement with AI-generated legal content can help students identify superficial analysis, citation deficiencies, and argumentative weaknesses. Atila-Aydin & Lammasniemi (2026) report that law students' use of GenAI is complex and often combines sophisticated use with concerns about over-reliance, leading them to recommend explicit law-school guidelines. Kanapathy & Shamini (2026) found improved performance when first-year law students used ChatGPT as a study companion, but the gains were more evident in foundational achievement than at the highest levels of legal reasoning. Together, these studies suggest that educational value is task-dependent and that responsible use requires explicit verification and instructional design.

c. Source verification as a condition of educational value

The need for verification is not a peripheral concern. Legal LLM research demonstrates that fluent systems can generate false or unsupported legal propositions. Dahl et al. (2024) documented substantial hallucination rates across legal tasks and showed that models may accept erroneous user premises. Hou et al. (2024) similarly found that machine-generated legal analyses can contain multiple types of gaps and hallucinations that are not easily captured by generic quality metrics. In a law-assignment study, Alimardani (2025) found that some students continued to include fabricated or inaccurate AI-generated material despite explicit instructions to verify outputs. This evidence makes source verification part of legal competence, not merely a technical precaution.

The Indonesian setting adds a temporal dimension to verification. During the January-February 2026 data-collection period, students were learning a code that had only just become legally operative (Republic of Indonesia, 2023). AI-generated answers could therefore mix the former criminal code with Law No. 1 of 2023, provide incorrect article numbers, or rely on outdated commentary. The literature thus supports a bounded hypothesis: GenAI should be most useful where the task is explanatory and least defensible where it replaces rule identification, authority checking, or independent legal judgment. This study tests that proposition through students' perceptions and open-ended accounts.

2. Research Method

a. Design and participants

The study used a cross-sectional mixed-methods design. Data were collected during January-February 2026 from 82 university students studying in Jakarta. The study team described recruitment as random; however, because the dataset does not contain a complete sampling frame or documentation of the randomization procedure, the sample is treated analytically as a cross-sectional student sample rather than as a probability sample from all university students in Jakarta. All 82 respondents reported prior use of AI for learning. The sample comprised 72 women (87.8%) and 10 men (12.2%). Forty-three respondents (52.4%) were in semesters 1-2 and 39 (47.6%) in semesters 3-4. Exact ages were available for 81 respondents ($M = 18.91$, $SD = 0.82$, range 17-21); one respondent reported only "<20" and was excluded from the exact-age calculation.

b. Instrument and operationalization

The questionnaire contained ten Likert statements scored from 1 (*strongly disagree*) to 5 (*strongly agree*) and three open-ended prompts. Items 1-7 assessed perceived AI-supported learning utility: understanding article content and purpose, understanding legal concepts and offence elements, explaining difficult provisions, using case examples, identifying relevant KUHP provisions, receiving feedback, and strengthening legal analysis and argumentation. Items 8-9 assessed verification literacy: checking AI answers against official legal sources and recognizing answers that conflict with applicable law. Item 10 captured overall perceived effectiveness of AI when combined with official-source verification.

The seven-item learning-utility composite had high internal consistency (Cronbach's $\alpha = .943$). The two-item verification composite also showed acceptable-to-high internal consistency for an exploratory short scale ($\alpha = .839$). These composites were calculated as the arithmetic mean of their constituent items. Because the instrument was developed for this study and has not undergone independent construct validation, the composites are interpreted as descriptive research measures rather than established psychometric scales.

c. Quantitative analysis

Descriptive statistics included means, standard deviations, 95% confidence intervals, and proportions selecting *agree* or *strongly agree*. Spearman's ρ was used for bivariate associations because individual responses were ordinal. An exploratory proportional-odds ordered-logit model predicted the five-level overall-effectiveness item from standardized learning utility and verification literacy while adjusting for semester group. Odds ratios (ORs) are reported for interpretability. The model is inferentially exploratory and does not imply causal effects.

d. Qualitative analysis

Eighty respondents provided non-empty answers to each open-ended prompt. Responses were analyzed using multi-label thematic content analysis. Initial codes were developed from

repeated reading of the responses and then applied across the full set. A single response could receive more than one code. For the benefit prompt, recurring categories included simplified understanding, case analysis/application, speed of information access, interactive learning/feedback, and access to multiple perspectives. Frequencies are used as descriptive indicators of salience rather than as population estimates. No intercoder-reliability statistic is reported because the available dataset supported one structured analytic coding pass; future work should use independent coding and adjudication.

e. Ethics and data integrity

The analytic dataset contained no direct personal identifiers. However, the material supplied for manuscript preparation did not include an institutional ethics-approval number or a full record of the informed-consent procedure. Before journal submission, the authors must insert the actual ethics review and consent information applicable to the original data collection. No approval status should be inferred or invented.

3. Results

a. Participant profile

Table 1 summarizes the sample. The respondents were concentrated in the first four semesters, and all had prior experience using AI for learning. This characteristic is analytically relevant because the study evaluates perceived use among users rather than adoption among the full student population.

Table 1. Participant characteristics (n = 82).

Characteristic	Result
Women	72 (87.8%)
Men	10 (12.2%)
Semester 1-2	43 (52.4%)
Semester 3-4	39 (47.6%)
Prior AI use for learning	82 (100%)
Exact age, n	81
Age, mean (SD)	18.91 (0.82)
Age range	17-21

b. Item-level perceptions

The overall pattern was positive but differentiated by task. Table 2 shows that the highest mean among the learning-utility items concerned explaining KUHP provisions that were previously difficult to understand (Item 3, M = 3.84, SD = 0.88; 67.1% agree/strongly agree). Students also reported relatively positive perceptions for understanding the content and purpose of provisions (Item 1, M = 3.77) and legal concepts, terminology, and offence elements (Item 2, M = 3.73).

Ratings became more cautious as tasks moved from explanation toward legal judgment. AI support for identifying the relevant KUHP provision in a case received the lowest mean among the learning-utility items (Item 5, M = 3.56; 51.2% agreement), closely followed by strengthening legal analysis and argumentation (Item 7, M = 3.57; 54.9% agreement). Case-based application (Item 4) also showed greater dispersion (M = 3.60, SD = 1.00), indicating less consensus among respondents.

Table 2. Item-level results.

Item	Construct	Mean	SD	Agree/Strongly agree
1	Understanding article content/purpose	3.77	0.86	63.4%
2	Concepts, terms, and offence elements	3.73	0.86	63.4%
3	Explaining previously difficult provisions	3.84	0.88	67.1%
4	Case examples and application	3.60	1.00	57.3%
5	Identifying relevant KUHP provisions	3.56	0.86	51.2%
6	Feedback on legal answers/analysis	3.70	0.83	62.2%
7	Legal analysis and argumentation	3.57	0.86	54.9%
8	Checking AI against official legal sources	4.11	1.05	73.2%
9	Recognizing inconsistency with applicable law	3.77	1.02	62.2%
10	Overall AI + verification effectiveness	3.84	1.02	62.2%

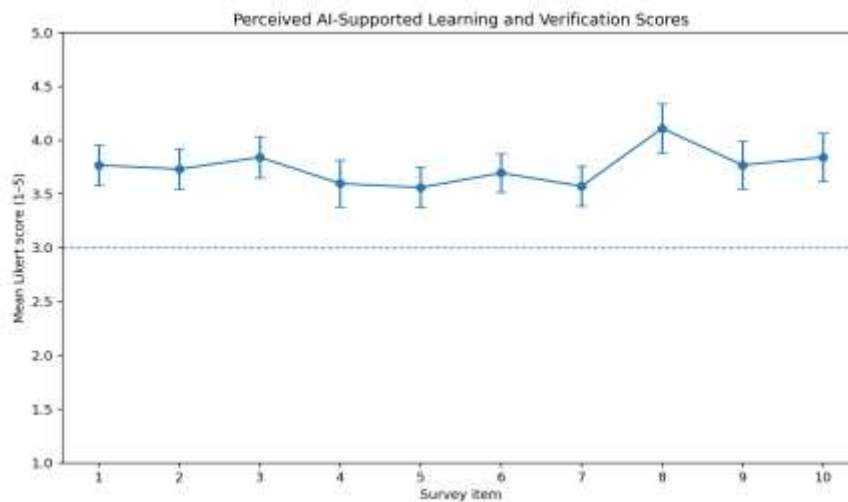


Figure 1. Mean scores for the ten survey items; error bars show approximate 95% confidence intervals.

The seven-item learning-utility composite averaged 3.68 (SD = .76; 95% CI [3.51, 3.85]). The verification composite averaged 3.94 (SD = .96), indicating that respondents reported somewhat stronger source-checking practices than advanced reasoning benefits.

c. Associations with overall perceived effectiveness

Learning utility was strongly associated with Item 10, the overall judgment that AI combined with official-source verification improves KUHP understanding (Spearman rho = .721, $p < .001$). Verification literacy showed an even stronger association with the same outcome (rho = .755, $p < .001$). These correlations indicate that students who perceived greater instructional utility and those who reported stronger verification practices were also more likely to rate AI-supported KUHP learning as effective.

The ordered-logit model produced the same substantive pattern. After standardization and adjustment for semester group, a one-standard-deviation increase in learning utility was associated with 6.54 times higher odds of being in a higher overall-effectiveness category (beta = 1.878, $p < .001$). A one-standard-deviation increase in verification literacy was associated with 8.84 times higher odds (beta = 2.179, $p < .001$). Semester group was not independently significant (OR = 1.64, $p = .350$). Because the predictors and outcome are all self-reported perceptions from a single cross-sectional survey, these estimates should be interpreted as associations, not effects.

Table 3. Exploratory ordered-logit model predicting overall perceived effectiveness.

Predictor	beta	Odds ratio	p
Learning utility (standardized)	1.878	6.54	< .001
Verification literacy (standardized)	2.179	8.84	< .001
Semester 3-4 vs. 1-2	0.494	1.64	.350

d. Student-described benefits

The qualitative responses provide a more concrete account of what “useful” meant to students. Sixty-four of 80 non-empty responses (80.0%) referred to simplified understanding or clearer explanation of legal provisions. Many respondents described AI as a tool for converting formal statutory language into more accessible explanations. Twenty-nine responses (36.2%) referred to case analysis or application, while 25 (31.2%) emphasized speed and efficiency in locating or summarizing information. Interactive learning or feedback appeared in 15 responses (18.8%), and multiple perspectives or comparisons appeared in 13 (16.2%).

These themes are consistent with the quantitative pattern: explanatory functions were more strongly endorsed than functions requiring independent legal classification. The open-ended material therefore reinforces rather than merely repeats the numerical findings.

4. Discussion

a. GenAI appears most credible as an explanatory scaffold

The first major finding is task differentiation. Students were most positive about AI when it helped make difficult statutory material easier to understand and less positive when the task required selecting the controlling provision or developing a legal argument. This pattern is theoretically coherent and consistent with prior legal-education research. Hargreaves (2023) and Alimardani (2024) both found that AI performance weakens when legal tasks demand detailed issue spotting, jurisdiction-specific application, and critical analysis. Kanapathy & Shamini (2026) likewise report stronger benefits for foundational performance than for the highest level of legal achievement.

The implication is not that GenAI lacks pedagogical value, but that its value is bounded by task type. Explanatory scaffolding is a plausible use because students can compare an AI explanation with the statutory text and use the dialogue to identify concepts they do not understand. By contrast, delegating legal classification or final argumentation is more problematic because it shifts the core reasoning task from the learner to the system. This distinction also helps reconcile apparently conflicting educational literature: positive learning effects can coexist with concerns about shallow cognition if AI is used productively for some stages of learning and inappropriately for others (Deng et al., 2025; McGrath et al., 2025).

Verification is not an optional add-on

The strongest analytical result is the independent association between verification literacy and overall perceived effectiveness. Students who reported checking AI outputs against official sources were more likely to regard AI-supported learning as effective even after learning utility was considered simultaneously. Although causality cannot be inferred, this pattern supports the proposition that verification is part of the learning mechanism rather than a separate compliance step.

This interpretation is particularly important in law. Hallucination research shows that LLMs can produce incorrect legal claims with high linguistic confidence (Dahl et al., 2024; Hou et al., 2024). The assignment study by Alimardani (2025) demonstrates that even explicit instructions to verify may not eliminate students’ use of fabricated content. During a statutory

transition such as the 2026 KUHP implementation, temporal validity becomes an additional source of error. Students must therefore be taught to ask not only whether an answer sounds plausible, but also: *What is the controlling source? Is the cited article current? Does the explanation correspond to the enacted text?*

Implications for curriculum and assessment

The findings support a four-stage learning sequence that preserves both efficiency and legal rigor.

- a) **Official-text first.** Students begin with the enacted KUHP provision or another authoritative source, identify the operative wording, and note unfamiliar concepts.
- b) **AI-assisted explanation.** AI is used to restate the provision, define concepts, compare interpretations, or generate examples. The output is treated as provisional.
- c) **Case testing.** Students independently apply the rule to a factual scenario before asking AI to critique omissions, counterarguments, or alternative classifications.
- d) **Verification and reflection.** Students compare material AI claims with official sources, record discrepancies, and revise the analysis in their own words.

This sequence is consistent with human-centred educational guidance (Miao et al., 2024; Miao & Holmes, 2023) and legal-education proposals that preserve substantive legal knowledge and critical judgment (Atilla-Aydin & Lammasniemi, 2026; Bliss, 2024; Head & Willis, 2024). It is also compatible with assessment frameworks that explicitly vary the permitted role of AI rather than treating all use as equivalent (Furze et al., 2024).

For assessment, lecturers should prefer tasks that make the reasoning process visible. Students can be required to identify the authoritative provision, explain its elements, submit an initial human analysis, disclose the purpose for which AI was used, identify at least one potential weakness in the AI response, and provide a final verified conclusion. Such design shifts the assessment target from text production to legal judgment. It also reduces the educational value of simply copying fluent AI output.

b. Evaluation of the evidence and limitations

Several limitations constrain the claims. First, the study measures perceptions rather than objective knowledge gains. A respondent who reports that AI improved understanding may still misunderstand the law. Second, the design is cross-sectional; it cannot establish that AI caused higher KUHP comprehension. Third, the sample is modest, concentrated in semesters 1-4, and strongly gender-imbalanced. Fourth, all respondents were AI users, so the study cannot compare users with non-users. Fifth, the sampling procedure was described as random by the study team but is insufficiently documented to support probability-based generalization. Sixth, the newly developed survey requires further construct validation in larger and more diverse samples.

Future research should therefore use pre/post KUHP knowledge tests, controlled comparisons of AI-assisted and non-AI conditions, direct source-verification tasks, and performance-based case analyses. It should also record the AI platform, model version, prompting strategy, and date of use, because GenAI systems change rapidly. A particularly useful design would compare three conditions: official-source study only, unrestricted AI use, and structured AI use with mandatory verification. Such a design would allow researchers to test whether the four-stage sequence produces measurable gains in legal reasoning rather than merely favorable perceptions.

5. Conclusion

Among 82 university students studying in Jakarta at the beginning of Indonesia's 2026 KUHP era, GenAI was perceived as most useful for making difficult statutory material easier to understand. Perceived benefits were weaker for selecting the legally relevant provision and constructing legal argumentation, indicating that students themselves distinguish between explanation and legal judgment. Learning utility and source-verification literacy were each strongly associated with overall perceived effectiveness, and the open-ended responses similarly emphasized explanation, application, speed, and the importance of checking AI-generated information.

The most defensible interpretation is therefore neither that GenAI should replace conventional doctrinal learning nor that it should be excluded from legal education. It should be positioned as a bounded learning scaffold. An official-text-first, AI-assisted-explanation, case-testing, and verification sequence preserves the efficiency of conversational AI while keeping authority, interpretation, and final legal judgment with the learner. For KUHP education in particular, this arrangement is important because students are learning a newly operative statutory framework in an information environment where fluent AI output may not reflect the current law.

References

- Ajevski, M., Barker, K., Gilbert, A., Hardie, L., & Ryan, F. (2023). ChatGPT and the future of legal education and practice. *The Law Teacher*, 57(3), 352–364. <https://doi.org/10.1080/03069400.2023.2207426>
- Al Asyari, H., Abdul Halim, H., & Tan, R. A. R. N. (2025). ChatGPT and legal education in indonesia: Examining readiness and implications. *Journal of Indonesian Legal Studies*, 10(1), 555–588. <https://doi.org/10.15294/jils.v10i1.19338>
- Alimardani, A. (2024). Generative artificial intelligence vs. Law students: An empirical study on criminal law exam performance. *Law, Innovation and Technology*, 16(2), 777–819. <https://doi.org/10.1080/17579961.2024.2392932>
- Alimardani, A. (2025). Borderline disaster: An empirical study on student usage of GenAI in a law assignment. *IEEE Transactions on Technology and Society*, 6(2), 210–219. <https://doi.org/10.1109/TTS.2025.3540978>
- Atilla-Aydin, S., & Lammasniemi, L. (2026). Preparing guidelines for generative AI in law schools: Student perspectives and educational implications. *The Law Teacher*, 60(1), 181–200. <https://doi.org/10.1080/03069400.2025.2585427>
- Bliss, J. (2024). Teaching law in the age of generative AI. *Jurimetrics*, 64(2), 111–161.
- 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>
- Dahl, M., Magesh, V., Suzgun, M., & Ho, D. E. (2024). Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis*, 16(1), 64–93. <https://doi.org/10.1093/jla/laae003>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- Farrokhnia, M., Banihashemi, 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>
- Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI assessment scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, 40(4), 38–55. <https://doi.org/10.14742/ajet.9434>
- Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), ep421.
<https://doi.org/10.30935/cedtech/13036>
- Hargreaves, S. (2023). “Words are flowing out like endless rain into a paper cup”: ChatGPT & law school assessments. *Legal Education Review*, 33(1), 69–105.
<https://doi.org/10.53300/001c.83297>
- Head, A., & Willis, S. (2024). Assessing law students in a GenAI world to create knowledgeable future lawyers. *International Journal of the Legal Profession*, 31(3), 293–310.
<https://doi.org/10.1080/09695958.2024.2379785>
- Hou, A., Jurayj, W., Holzenberger, N., Blair-Stanek, A., & Van Durme, B. (2024). Gaps or hallucinations? Scrutinizing machine-generated legal analysis for fine-grained text evaluations. *Proceedings of the Natural Legal Language Processing Workshop 2024*, 280–302. <https://doi.org/10.18653/v1/2024.nllp-1.24>
- Hyde-Vaamonde, C., & Keller, A. (2025). Testing the frontier: Generative AI in legal education and beyond. *Law, Technology and Humans*, 7(3), 52–61. <https://doi.org/10.5204/lthj.4037>
- Kanapathy, J., & Shamini, J. A. (2026). Evaluating the role of ChatGPT as a legal educational companion. *Asia Pacific Journal of Educators and Education*, 41(1), 111–131.
<https://doi.org/10.21315/apjee2026.41.1.6>
- Kasneci, E., Sessler, K., Kuechemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Guennemann, S., Huellermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... 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>
- McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2025). Generative AI chatbots in higher education: A review of an emerging research area. *Higher Education*, 89, 1533–1549.
<https://doi.org/10.1007/s10734-024-01288-w>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO.
<https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO.
<https://doi.org/10.54675/JKJB9835>
- Republic of Indonesia. (2023). *Law no. 1 of 2023 on the criminal code (kitab undang-undang hukum pidana)*. <https://peraturan.bpk.go.id/Details/234935/uu-no-1-tahun-2023>
- 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>