applied sciences

MDPI

Article

Evaluating Source-Grounded AI Technical Assistants for Higher Education: Technical Accuracy, Knowledge Boundaries, and Repeatability in Printing Technology

Nemanja Kašiković, Sandra Dedijer, Neda Milić Keresteš *, Gojko Vladić and Živko Pavlović

Department of Graphic Engineering and Design, Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia; knemanja@uns.ac.rs (N.K.); dedijer@uns.ac.rs (S.D.); vladicg@uns.ac.rs (G.V.); zivkopvl@uns.ac.rs (Ž.P.)

* Correspondence: milicn@uns.ac.rs

Abstract

Source-grounded generative AI assistants can improve access to disciplinary literature, but citations do not ensure technical correctness. We evaluated two textile- and offset-printing assistants in Google NotebookLM (now Gemini Notebook), using 20 controlled prompts per assistant. One subject-matter expert rated technical accuracy, completeness, source support, and contextualization and manually checked citations. Mean overall scores were 4.73/5 and 4.61/5, respectively; because prompt sets differed, these are descriptive, not comparative. Source support exceeded technical accuracy in 35 of 40 original responses. In eight explicitly prompted knowledge-boundary tests, six responses showed appropriate abstention and two showed abstention with reservations; this evaluates prompted handling, not spontaneous recognition. Of seven repeated pairs with preserved identical wording, six received unchanged overall scores (mean absolute difference, 0.036/5); content comparison found recurring errors and differences across the ten pairs, although three offset pairs were only same-topic comparisons. A recurring textile error labeled source-reported heat-setting temperatures as glass-transition temperatures; a reported fabric-temperature rise applies only to vat-dye all-in cotton printing in saturated steam. With one expert evaluator and 15 original offset prompts preserved only as topic labels, this is an exploratory case study, not a validated benchmark. Student learning outcomes were not assessed.

Keywords: generative artificial intelligence; higher education; source grounding; technical education; engineering education; printing technology; AI literacy; knowledge boundaries; repeatability; NotebookLM

Check for updates

Academic Editor: Dan Sui

Received: 27 September 2026
Revised: 20 October 2026
Accepted: 22 October 2026
Published: 25 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Generative artificial intelligence (GenAI) has rapidly expanded from general-purpose conversational systems to tools embedded in teaching, learning, research, and disciplinary practice. In higher education, the potential benefits include immediate feedback, personalized explanation, improved access to learning resources, and support for self-directed learning. At the same time, the educational value of GenAI depends on how it is integrated, how users evaluate its outputs, and how institutions address risks related to inaccurate information, over-reliance, academic integrity, and uneven AI literacy [1–6]. Recent higher-education research has therefore emphasized both the opportunities and risks of AI-based tools, the need for responsible and pedagogically aligned integration,

and the importance of developing students' and lecturers' capacity to critically evaluate AI-generated information [1–6].

The problem becomes more consequential in specialized engineering and technology domains. A fluent but technically inaccurate answer about process settings, material compatibility, equipment configuration, or testing conditions can mislead a learner even when the language is clear and the response appears authoritative. Conversely, a general-purpose model may not distinguish sufficiently between a textbook principle, a manufacturer-specific setting, an experimental condition, and a normative requirement. These distinctions are central in printing technology, where process performance depends on combinations of equipment, substrates, inks or dyes, environmental conditions, measurement procedures, and technical standards.

One proposed response to this challenge is source grounding: restricting or prioritizing an assistant's responses around a curated set of documents and providing citations that allow users to inspect the evidence. Google's NotebookLM was designed around this principle and explicitly presents responses with links to the supplied sources. Google nevertheless cautions that source-grounded responses can still contain inaccuracies and should be checked against the underlying material [7,8]. This distinction is important for higher education. Source grounding can improve traceability, but the reliability of the resulting explanation depends not only on whether a statement is cited, but also on the quality, currency, scope, and interpretation of the cited source.

The present study builds on earlier work in printing technology. A previous development study described a modular approach for transforming curated professional and educational literature into domain-specific AI technical assistants for offset printing, digital printing, flexography, gravure, screen printing, and textile printing [9]. A subsequent exploratory user study of six printing-technology assistants found high perceived usefulness and acceptance, particularly for rapid access to technical information, while users remained more cautious about unconditional trust and consistency [10]. Related work in graphic engineering has also examined AI-supported packaging and design practice, machine-learning-based print-quality analysis, and conversational support in the ASKING educational platform [11–15]. These studies establish the educational and professional relevance of AI in the field, but they do not answer a more specific question: how technically reliable are source-grounded assistants when they are systematically tested with domain questions designed to expose overgeneralization, unsupported specificity, and knowledge-boundary failures?

This study addresses that gap through an expert evaluation of two source-grounded assistants: a Textile Printing Technical Assistant and an Offset Printing Technical Assistant. The contribution is methodological as well as empirical. The conceptual distinction between faithfulness to a source and factual correctness is established in research on generated text (Section 2); the contribution here is its application as a criterion-level expert audit in an engineering domain. Rather than treating citation presence as a proxy for correctness, the evaluation separates technical accuracy, completeness, source support, and contextualization; explicitly tests whether the assistants abstain when universal or manufacturer-specific parameters cannot be justified; and examines whether strengths and weaknesses recur when selected prompts are repeated. The initial textile test set (T01–T16), conducted before four additional boundary prompts were added, is treated as the first phase of this study and has not been published elsewhere. Its criterion-level ratings are reported in the Supplementary Materials. The subsequent phases comprise four additional textile knowledge-boundary prompts, a 20-prompt evaluation of the offset assistant, repeatability tests of both assistants, and a cross-domain analysis of the four rating dimensions, boundary handling, and response stability.

The study was guided by four research questions: (RQ1) How accurately and completely do source-grounded domain-specific AI assistants answer technical questions in printing technology? (RQ2) How effectively do they recognize knowledge boundaries and avoid unsupported universal or manufacturer-specific recommendations? (RQ3) What recurrent failure modes remain despite source grounding and manually verified citation traceability? (RQ4) To what extent are response-quality patterns stable across repeated prompts?

2. Source-Grounded AI and Disciplinary Learning in Higher Education

Higher-education research increasingly frames AI literacy as more than the ability to operate a tool. It includes understanding limitations, evaluating outputs, applying AI appropriately, and making ethical and epistemically responsible decisions [3,4,6]. Wakeman et al. describe critical thinking and evaluation as core components of AI literacy, while Uijl et al. distinguish learning about AI from learning with AI and emphasize reciprocal development of lecturer and student AI literacy [3,4]. These perspectives are directly relevant to source-grounded assistants: a cited answer may feel more trustworthy than an uncited one, but learners still need to evaluate the source, context, and technical applicability of the claim.

Studies of AI teaching assistants likewise show a tension between perceived usefulness and concerns about accuracy. Hu found positive attitudes toward AI teaching assistants among students in a higher-education setting, alongside explicit concerns about the reliability of AI-generated information and the continuing importance of human educators [5]. Denecke et al. similarly identified both opportunities and risks for AI-based tools in higher education [1]. Schwenke et al., examining chatbot-supported thesis writing, showed why continuous validation of generated information and references remains necessary [6]. Together, this literature suggests that educational adoption should be evaluated not only through usability or acceptance, but also through the epistemic behavior of the system under uncertainty. Broader reviews likewise emphasize both the potential educational benefits of generative AI and persistent concerns about accuracy, over-reliance, and the need for pedagogically grounded integration [16–18].

Research on retrieval-augmented and citation-generating language models points to the same distinction. A generated statement can be faithful to its source and still be factually incorrect, a separation first formalized for abstractive summarization [19]. In an audit of four generative search engines, only 51.5% of generated sentences were fully supported by their citations, and only 74.5% of citations supported the sentence to which they were attached [20]. Expert evaluation of commercial retrieval-augmented legal research tools found hallucinated (incorrect or misgrounded) answers in 17–33% of queries, despite vendor claims that grounding eliminates such errors [21]. Citations also shape trust: in a controlled experiment, the presence of citations increased users' trust in chatbot responses even when the citations were random, whereas checking them decreased trust [22]. In higher education, NotebookLM has already been configured as a source-grounded Socratic tutor for physics that restricts responses to curated course materials [23]. Together, these studies indicate that citation presence is a weak proxy for correctness and that domain-specific assistants need to be evaluated at the level of technical content.

Reliability and validation in AI-based information systems also include system-level properties beyond the factual correctness of individual generated answers. Semenov et al. [24] examine poisoning threats to recommender systems and evaluate a behavioral detection layer through controlled attack scenarios, repeated runs, and parameter-sensitivity analysis. Their focus is the integrity of interaction data and robustness of a recommender-system defense, not the factual accuracy of generated text. We therefore cite this work as a

complementary example of broader AI-system dependability, not as evidence that source grounding improves answer accuracy.

The printing-technology setting provides a useful test bed because it combines conceptual knowledge with process-specific constraints. The earlier user evaluation of these assistants reported an overall mean of 4.36/5 across 17 Likert items, with Practical Usefulness (4.46) and User Acceptance and Satisfaction (4.48) as the highest-rated constructs [10]. These perceptions are encouraging but do not establish technical validity.

For the present study, the central educational proposition is therefore deliberately limited: source-grounded technical assistants may function as complementary tools for clarification, literature navigation, laboratory preparation, and guided problem solving, but their suitability depends on reliable behavior at the boundaries of the knowledge base. The study does not claim that the assistants improve student learning outcomes; that question requires a separate learner-centered experimental design.

3. Materials and Methods

3.1. Assistants and Knowledge-Base Configuration

Two domain-specific assistants were evaluated: the Textile Printing Technical Assistant and the Offset Printing Technical Assistant. Both were implemented in Google NotebookLM, which was renamed Gemini Notebook in July 2026 [8]. Each used a curated corpus of professional literature, scientific publications, educational materials, and technical documentation. Itemized inventories list 14 named sources for Textile Printing and seven for Offset Printing, with their languages (Supplementary File S1, Corpus Inventory sheet; Supplementary File S2, Table S1); the Textile count matches the 14 sources recorded in the evaluation archive. Bibliographic details were completed from the title pages of the source PDFs and the evaluator's records. The sources were selected by the evaluator as core literature for the courses the assistants support; no formal selection criteria were defined in advance, and no dated export of the notebook sources or custom instructions was retained, so the inventories list the sources as recorded by the evaluator rather than a time-stamped copy of each corpus.

The evaluator (N.K.) confirmed that the domain instructions summarized in Supplementary File S2 (Section SA) were the custom instructions used during testing; the summaries were prepared after the evaluation for documentation. The instructions emphasize source-grounded answers, citation traceability, domain boundaries, explicit acknowledgement of insufficient evidence, and avoidance of unsupported process values. Textile-specific guidance addresses fibre, ink/dye, pretreatment, fixation, and troubleshooting context; offset-specific guidance covers press configuration, ink–water balance, materials, measurement, standards, and process-dependent settings. The retained records identify Google NotebookLM (renamed Gemini Notebook in July 2026), but do not identify the underlying model version or model-specific generation settings.

The assistants were designed as supplementary learning materials for courses taught by the evaluator. This course-related purpose informed their design, but the present study evaluated assistant responses rather than student use or learning outcomes.

The public notebook links used in the study are reported in the Data Availability Statement. The knowledge bases and custom instructions were not changed during the repeatability phase. This was essential because changing sources between runs would confound response variation with knowledge-base variation.

3.2. Evaluation Prompt Design

Each assistant was evaluated with 20 controlled prompts. The Offset Printing Assistant used five categories with four prompts each: fundamental knowledge; materials and

process; print quality and testing; troubleshooting; and knowledge boundaries. The textile evaluation began with an initial 16-prompt expert phase covering core technical knowledge, pretreatment and process parameters, troubleshooting, and stress tests/knowledge boundaries. Four additional knowledge-boundary prompts were then added so that both assistants had 20 evaluated prompts. These four textile prompts tested universal steaming conditions, manufacturer-specific processing recommendations, conflicting fixation recommendations, and self-identification of a technically relevant question that could not be reliably answered from the available corpus. The retained records establish this sequence: initial 16-prompt textile evaluation, addition of four knowledge-boundary prompts, and subsequent repeatability testing. The offset assistant was evaluated on its 20-prompt set, followed by repeatability testing. Five selected prompts per assistant were repeated in fresh conversations. All tests were conducted on 22 September 2026: the initial 16-prompt textile phase, the four added textile prompts, the 20-prompt offset evaluation, and the repeat tests. The supplementary workbook's Basic Evaluations and Full Responses sheets retain the prompt wording where it is available: all 20 original Textile prompts and five original Offset prompts are verbatim, whereas Offset T01–T14 and T16 remain topic labels only. The Original Verbatim Archive preserves the 50 response uploads before translation. Missing prompt wording has not been reconstructed.

Prompt-category mapping follows the retained prompt-level dataset: for Textile, T01–T09 are Fundamental knowledge, T10–T16 are Process parameters and materials, and T17–T20 are Knowledge boundaries; for Offset, T01–T04 are Fundamental knowledge, T05–T08 are Materials and process, T09–T12 are Print quality and testing, T13–T16 are Troubleshooting, and T17–T20 are Knowledge boundaries. Textile T14–T16 also contain boundary-related cues but remain in their original process/materials category and are discussed separately from the focused boundary set. That set comprises T17–T20 for each assistant (n = 8); all eight prompts explicitly cue handling of non-universal, missing, or context-dependent information. The study therefore assesses prompt-guided boundary handling and does not estimate spontaneous recognition in ordinary technical prompts.

The prompts were intentionally heterogeneous. Straightforward conceptual questions tested explanatory competence; process questions tested the handling of material- and equipment-dependent parameters; troubleshooting prompts required causal reasoning and diagnostic sequencing; and boundary prompts were designed to elicit overconfident specificity unless the assistant correctly recognized missing information. No web discovery was used by the evaluated assistants during the controlled tests; answers were judged against the fixed notebook sources. The evaluation structure and the analytical purpose of each component are summarized in Table 1.

Table 1. Evaluation structure and analytical purpose.

Component	Textile Assistant	Offset Assistant	Analytical Purpose
Primary prompt set	16 prompts from the initial evaluation + 4 added boundary prompts	20 prompts in five balanced categories	Assess technical response quality across ordinary and difficult tasks
Scoring	Four 1–5 criteria	Four 1–5 criteria	Rate accuracy, completeness, evidence, and context
Citation checking	Manual expert verification	Manual expert verification	Verify claim traceability
Boundary coding	Unsupported claims; appropriate abstention	Unsupported claims; appropriate abstention	Code unsupported claims and abstention
Repeatability	5 repeated prompts	5 repeated prompts	Assess response stability

3.3. Expert Scoring and Source Verification

Each response was evaluated on a five-point scale (1–5) for four criteria: technical accuracy, completeness, source support, and contextualization. Technical accuracy assessed the correctness of the technological explanation. Completeness assessed whether the response addressed the requested elements. Source support assessed whether relevant claims could be traced to the notebook sources cited by the assistant. Contextualization assessed whether the response preserved distinctions among materials, equipment, formulations, process conditions, and experimental or industrial contexts instead of presenting source-specific values as universal rules. An overall score was calculated as the arithmetic mean of the four criteria. Equal weighting was used as a transparent descriptive convention because no empirically validated weighting scheme was available; it does not imply that the four dimensions have equal educational or practical importance. Criterion-level scores are reported alongside the composite so readers can inspect the profile and any trade-offs.

The original evaluation records do not contain a contemporaneous written set of score-level anchors. To make the recorded judgments interpretable, we clarify the evaluator's operational scale retrospectively; this clarification does not constitute a new rating of the responses. For each criterion, 5 indicated that it was fully met without a material shortcoming; 4 indicated strong performance with a minor error, omission, or qualification gap; 3 indicated a material but bounded shortcoming requiring correction; 2 indicated a major defect that made an important part of the answer unreliable; and 1 indicated a largely incorrect, unsupported, or nonresponsive answer. Half-point increments were used as intermediate judgments. Technical accuracy concerned correctness, completeness concerned coverage of requested elements, source support concerned traceability of substantive claims, and contextualization concerned qualification of material-, equipment-, and process-dependent conditions. Examples illustrate the anchors: Offset T10 received 4.0 for technical accuracy because its account mixed standard editions and described five substrate categories, whereas ISO 12647-2:2013 defines eight (PS1–PS8) [25]; this primary standard was not in the assistant's corpus and provides an external technical reference. Textile T17 received 3.0 for technical accuracy and 5.0 for source support because it attached an incorrect technical label (Tg) to source-supported heat-setting values (Section 4.1). These cases show that source support and technical accuracy are related but distinct. Citation checking was performed manually by the subject-matter expert using source passages exposed by the notebook interface. After the expert evaluation, OpenAI ChatGPT (GPT-5.6 Sol) assisted with organization, English translation and synthesis of archived materials, descriptive calculations, comparisons of original and repeated responses, and manuscript drafting. It was not used as an independent subject-matter rater, did not generate or alter the technical ratings, and did not independently verify the original notebook source passages; technical ratings, source verification, and scientific interpretation remained the responsibility of the human investigator.

The supplementary workbook's Scoring Rubric sheet presents criterion-specific descriptions for each score anchor and examples drawn from the recorded evaluations. No criterion score below 3 was observed; levels 1 and 2 are therefore defined anchors without observed response examples. The rubric is a retrospective clarification, and no response was re-scored. It therefore explains how the recorded ratings should be read, but it cannot demonstrate that every original score was assigned under a prospectively standardized procedure; this qualification applies in particular to the high ratings reported in Section 4.1.

The evaluator also recorded whether a response contained unsupported claims and whether abstention was appropriate. 'Appropriate abstention' did not require a refusal to answer the entire question. It referred to the assistant explicitly declining to provide a universal or exact parameter when the knowledge base did not justify one, while still

explaining relevant principles and identifying the information needed for a reliable recommendation. An unsupported claim was coded when a substantive technical statement could not be traced to any notebook source during manual checking. Abstention was coded as Yes, Yes with reservations (an initial abstention later weakened by insufficiently qualified values), or N/A (the prompt was not designed to test abstention); when an affirmative code followed an explicit instruction not to estimate, the supplementary dataset labels it 'Yes, prompt-directed.'

Textile T15 is coded Yes, prompt-directed: its wording explicitly instructed the assistant not to estimate a universal pretreatment pH, and the response stated that no universal value was supported before giving source-specific alkaline conditions. Textile T16 remains N/A for abstention because it asked the assistant to identify a technically important unanswered question rather than supply a technical parameter; this is prompt-guided boundary identification and is described separately. Neither T15 nor T16 is included in the dedicated eight-prompt denominator. This distinction separates compliance with an explicit boundary instruction from spontaneous limitation recognition, which was not measured.

The sole evaluator was one of the authors (N.K.), a Full Professor in Graphic Engineering and Design at the Faculty of Technical Sciences, University of Novi Sad. He teaches and researches the printing-technology subjects addressed here, including textile printing; his graduate and doctoral research addressed textile print quality and monitoring of process parameters, and he has an extensive publication and textbook record in the field. He also participated in developing and configuring the evaluated assistants and their source corpora. This dual role is relevant to interpretation and may introduce expectancy bias. No second qualified expert independently rated the responses; consequently, no inter-rater agreement statistic or disagreement-adjudication procedure is available.

3.4. Repeatability Assessment

Five prompts from each assistant were repeated in fresh conversations with the same knowledge base and custom instructions. The textile repeats included one prior conflicting-recommendations prompt and the four added knowledge-boundary prompts; all five original prompt wordings were available for comparison. The offset repeats covered ink-water balance (T04/R04), ISO 12647-2 (T10/R10), scumming troubleshooting (T13/R13), universal dampening parameters, and knowledge-base limitations. For the first three offset pairs, the original records preserve topic labels but not verbatim prompt wording; these are treated as same-topic comparisons only and are excluded from identical-prompt repeatability statistics. The other seven pairs had documented identical wording. Repeated responses were scored using the same rubric and compared for key technical claims, recurring errors, boundary recognition, and troubleshooting structure. Because only one repeat was performed per selected prompt, the analysis is a limited qualitative repeatability check, not a statistical reliability study.

The supplementary Repeatability sheet reports the four criterion-level ratings for all ten repeated responses. Each repeated overall score is calculated as the mean of the four unrounded criterion ratings and rounded to two decimals only for display; score differences and summary statistics use the unrounded means.

3.5. Data Analysis and Data Provenance

Only descriptive statistics were used. The evaluation was conducted on 22 September 2026, and the criterion ratings represent the subject-matter expert's judgments for that evaluation. All ratings were assigned by the evaluator during the evaluation and recorded in an evaluation spreadsheet; the prompt-level ratings are provided unchanged in the Basic Evaluations sheet of Supplementary File S1. The surviving archive does not retain

separate timestamps for the initial textile evaluation, the four subsequently added textile prompts, and the repeat tests, so their within-day order is documented by the sequence of the retained records rather than by timestamps. ChatGPT (GPT-5.6 Sol) was used after the expert evaluation to assist with organization, English translation and synthesis of the archived response and evaluation materials, descriptive calculations, comparison of original and repeated responses, and manuscript preparation. These later analytical and editorial uses did not generate or alter the expert ratings; the human evaluator retained responsibility for technical judgments and source checking. The supplementary dataset includes scores for all four criteria across 40 original tests (20 per assistant). For textile T01–T16, the original criterion-level ratings yield means for the initial 16-prompt phase of technical accuracy 4.63, completeness 5.00, source support 5.00, contextualization 4.50, and overall 4.78. Complete scores also exist for textile T17–T20 and all 20 offset prompts. Prompt scores average the four criteria; assistant scores average the 20 original prompts. Small differences between exact and displayed means reflect rounding without affecting the two-decimal summaries. The ten repeats were analyzed separately. The workbook contains English working copies and the unedited uploaded archive of all 50 recorded responses. This is an exploratory, expert-audited case study, not a validated benchmark; the descriptive results are limited to these assistants, prompts, sources, and evaluation conditions.

4. Results

4.1. Overall Expert Evaluation

Both assistants received high descriptive expert-evaluation scores (Table 2). These ratings were assigned by a single evaluator who also participated in developing the assistants, using score anchors that were documented retrospectively (Section 3.3). They therefore represent descriptive expert judgments rather than independently validated measures of assistant performance. The near-ceiling source-support means (5.00 and 4.98) should be interpreted with this limitation in mind. Across 20 original prompts, the Textile Printing Assistant's mean overall score was 4.73/5. The Offset Printing Assistant's mean overall score was 4.61/5. Source support was the highest-rated dimension for both assistants, whereas contextualization was lower than completeness and source support. At the prompt level, source support exceeded technical accuracy in 35 of 40 prompts (16 textile and 19 offset) and was never lower, and it exceeded contextualization in 33 of 40. Technical accuracy received the maximum rating in only 5 of 40 prompts. No response was coded as containing unsupported claims, a coding that relied on the expert's live checks of the cited passages. Because the assistants were tested with different prompt sets in different domains, these means are descriptive and should not be interpreted as a direct comparison of assistant performance.

Table 2. Descriptive expert-evaluation results.

Assistant	Technical Accuracy M (SD)	Completeness M (SD)	Source Support M (SD)	Contextualization M (SD)	Overall M (SD)
Textile Printing	4.48 (0.47)	5.00 (0.00)	5.00 (0.00)	4.43 (0.47)	4.73 (0.22)
Offset Printing	4.40 (0.26)	4.80 (0.30)	4.98 (0.11)	4.25 (0.34)	4.61 (0.18)

Note: Each assistant was evaluated on 20 prompts by a single evaluator who also participated in developing the assistants (Section 3.3). SD describes variation across prompts within that assistant; it is not inter-rater variability or an inferential uncertainty estimate. The prompt-level records are provided in the supplementary dataset. Medians (ranges), textile/offset: technical accuracy 4.5 (3.0–5.0)/4.5 (4.0–5.0); completeness 5.0 (5.0–5.0)/5.0 (4.0–5.0); source support 5.0 (5.0–5.0)/5.0 (4.5–5.0); contextualization 4.5 (3.5–5.0)/4.0 (4.0–5.0); overall 4.75 (4.13–5.00)/4.63 (4.25–5.00).

The textile assistant's performance was not uniformly stronger. In the added knowledge-boundary test on universal steaming conditions, it received an overall score of 4.13 and mislabeled source-reported heat-setting temperatures as glass-transition temperatures. The retained source identifies 190–200 °C for polyester and 210–220 °C for nylon 6.6 as practical heat-setting temperatures (§7.4.7, p. 164); a separate chapter of the same handbook lists Tg values of 80–90 °C for polyester and 50 °C for nylon 6.6 (Table 10.1, p. 228) [26,27]. The source evidence and its retrospective verification are documented in the supplementary workbook's Source Error Verification sheet. The 115 °C measured fabric temperature and theoretical value up to 140 °C are discussed for vat-dye all-in printing on cotton in ideal 100 °C saturated steam (§8.3.5, pp. 285–287) [28]. Thus, the Tg finding is an assistant interpretation error, while the steaming-temperature observation is source-supported only under its stated process conditions.

The lowest-scoring offset response concerned ISO 12647-2 (T10; 4.25). The primary standard was not in the notebook corpus, and the response relied on secondary literature. It described five standard paper types, the classification of the 2004 edition and its 2007 amendment, without identifying the edition, whereas the current 2013 edition defines eight printing-substrate categories (PS1–PS8) [25].

4.2. Knowledge-Boundary Awareness and Abstention

Focused knowledge-boundary analysis used eight dedicated prompts (Textile T17–T20 and Offset T17–T20). All eight explicitly asked the assistant to address missing, non-universal, or context-dependent information; six responses were coded Yes and two Yes with reservations. These results measure boundary handling under explicit prompting, not spontaneous recognition. The two T20 self-limitation prompts asked each assistant to identify a technically relevant question that could not be answered from the corpus and to explain the missing evidence; these are prompt-guided identification tasks.

Textile T17 illustrates why the abstention code must remain separate from technical accuracy: although the answer initially rejected universal steaming conditions, it mislabeled heat-setting temperatures in the cited handbook as glass-transition temperatures (Section 4.1). The abstention code records initial handling of the prompted limit and does not imply that the full answer was technically correct or operationally reliable.

4.3. Recurrent Failure Modes

Qualitative analysis identified six types of limitation (Table 3). The clearest recurring category was overgeneralization of source-specific numerical values.

A second, conceptually distinct category was a source-grounded interpretation error: Textile T17 reproduced source-supported heat-setting values but mislabeled them as glass-transition temperatures (Section 4.1). A third category concerned normative-document coverage: when the primary standard was absent, the assistant could provide a broadly correct explanation while blending edition-specific details from secondary literature. Troubleshooting answers sometimes identified plausible causes without prioritizing diagnostic checks. Some responses correctly abstained from giving a universal setting but then weakened that caution by listing highly specific examples. Several English answers also contained Serbian or Croatian technical terms, although the offset instructions asked for English terminology.

Table 3. Qualitative taxonomy of limitations observed across the two assistants.

Limitation Type	Representative Examples	Educational Significance
Overgeneralization of numerical values	Dampening pH/conductivity; water uptake; printing pressure; fixation and washing limits	Learners may mistake a literature example for a universal operating rule.
Assistant misclassification of source values	Textile T17/R17: heat-setting values labeled as Tg	A correctly cited number can still carry a wrong technical label.
Primary-document coverage gap	Offset T10/R10, ISO 12647-2 interpreted from secondary sources	Edition-specific or normative details may be outdated or blended.
Incomplete troubleshooting prioritization	Offset T12–T16; improved in R13	Novices may change multiple variables without isolating the root cause.
Qualification drift after abstention	Offset T17–T19; Textile T18	A correct refusal can be undermined by later overly specific recommendations.
Language inconsistency	Serbian/Croatian terms in at least 12 of 20 original offset responses	Minor professional-style issue; technical meaning largely preserved.

Note: Categories are descriptive, may overlap, and were not counted as mutually exclusive frequencies; the table is not a frequency ranking.

4.4. Repeatability of Strengths and Limitations

Among the seven documented identical-prompt pairs (Table 4), six received exactly the same overall score. The mean absolute score difference, calculated from unrounded composite means, was 0.036 points on the five-point scale, and the largest difference was 0.25 points (Textile T14/R14, 5.00 to 4.75). The three remaining offset pairs (T04/R04, T10/R10, and T13/R13; Table 5) are same-topic comparisons because the original prompt wording was not preserved; they are reported separately and excluded from these statistics. Across all ten pairs, eight composite scores were unchanged; the mean absolute difference was 0.038, with a maximum of 0.25.

Table 4. Documented identical-prompt repetitions (n = 7).

Assistant	Pair	Scores Original/Repeat	Difference	Main Observation
Textile	T17/R17	4.13/4.13	0.00	The Tg category error recurred; the steaming rise is source-supported only for the stated vat-dye cotton process.
Textile	T18/R18	4.25/4.25	0.00	Same overgeneralized recipes and absolute limits.
Textile	T19/R19	4.75/4.75	0.00	Stable context-sensitive interpretation.
Textile	T20/R20	4.88/4.88	0.00	Stable recognition of manufacturer-specific limits.
Textile	T14/R14	5.00/4.75	−0.25	Similar reasoning; repeat wording slightly more categorically.
Offset	T17/R17	4.75/4.75	0.00	Boundary recognition recurred with same conductivity caveat.
Offset	T20/R20	4.88/4.88	0.00	Repeatedly identified source limits without inventing settings.

Table 5. Same-topic offset comparisons with unavailable original prompt wording (excluded from identical-prompt statistics).

Pair	Overall Scores	Difference	Interpretation
T04/R04	4.50/4.50	0.00	Same-topic comparison; original wording unavailable.
T10/R10	4.25/4.25	0.00	Same-topic comparison; original wording unavailable.
T13/R13	4.50/4.63	+0.13	More structured troubleshooting with added numbers; unavailable wording prevents attribution to the repeat.

For the original Textile T17 response, the recorded ratings were 3.0 for technical accuracy, 5.0 for completeness, 5.0 for source support, and 3.5 for contextualization; their unrounded mean is 4.125 (reported as 4.13). The repeated Textile R17 response received the same four criterion ratings and the same mean. Thus, the repeat reproduced the heat-setting/Tg category error alongside high completeness and source-support ratings; the high composite does not negate the technical error. The separate temperature observation is supported under the narrow vat-dye, all-in cotton steaming conditions documented in the supplementary workbook's Source Error Verification sheet [28]. Equal weighting can obscure the terminology error, so the component profile is reported alongside the composite.

The seven documented identical-prompt pairs should not be interpreted as a formal reliability coefficient. The study used only one repeat for these selected prompts and did not estimate inter-rater reliability. Their value is qualitative: the assistants repeated correct boundary behavior as well as inaccurate claims and contextual weaknesses. The three topic-only offset comparisons are not evidence of repeatability under identical wording. For Textile T17/R17, the same Tg mislabeling recurred in the repeat (Section 4.1). Stable behavior is therefore not equivalent to valid behavior.

A claim-level comparison of the ten saved original–repeat pairs shows why score agreement alone is insufficient. In the five identical-prompt textile pairs, the T17/R17 glass-transition/heat-setting misclassification recurred, and the separate fabric-temperature statement shifted from 115–140 °C in T17 to "up to 115 °C" in R17; T18/R18 again began by declining an exact recipe but then supplied multiple highly specific process values; T19/R19 retained the same contextual interpretation of differing fixation recommendations; and T20/R20 both withheld manufacturer-specific waveform settings while changing the illustrative ink formulation. T14/R14 retained the same non-universal reasoning, although the numerical examples and wording varied. In the two identical-prompt offset pairs, T17/R17 again rejected universal pH/conductivity settings but repeated overgeneralized thresholds, whereas T20/R20 consistently identified missing press and blanket data and withheld exact settings. The three offset pairs with unavailable original wording are only same-topic comparisons: R13 was more diagnostically structured than T13 but added equipment-dependent values, while R04 and R10 retained the main topic-level conclusions. The full pair-by-pair comparison is in the supplementary workbook's Repeatability Claim Comparison sheet. The archived English working copies do not preserve NotebookLM source-link metadata showing which retrieved passages were attached to each claim; accordingly, we report content and numerical-claim patterns, not citation-selection repeatability. The T17/R17 source evidence is documented separately as retrospective verification.

5. Discussion

5.1. Source Grounding Supports Traceability but Does Not Ensure Correctness

In the textile steaming case, the response contained technically inaccurate Tg labels for source-supported heat-setting values. The retained source distinguishes Tg from heat setting: §7.4.7 identifies the cited 190–200 °C and 210–220 °C values as practical heat-setting

ranges, while Table 10.1 reports lower Tg values for polyester and nylon 6,6 [26,27]. This evidence establishes an assistant misclassification rather than an error in the cited source; its page-level evidence is documented in the supplementary workbook's Source Error Verification sheet. The separate temperature rise is confined to the specific vat-dye all-in cotton steaming process described by Miles [28].

In the terminology of research on generated text, the assistants were largely faithful to their sources but not always factually correct [19]. The prompt-level pattern reported in Section 4.1, with source support exceeding technical accuracy in 35 of 40 prompts, indicates that this gap was systematic rather than confined to a single response, although it rests on one expert's ratings and was not tested inferentially.

This result is consistent with Google's own description of NotebookLM/Gemini Notebook: grounding is intended to tie responses to supplied sources and facilitate checking, while the product documentation still warns that AI outputs can be inaccurate [7,8]. Citation presence can itself raise user trust, even when the citations are irrelevant [22]. From an educational perspective, this supports a verification-centered model of AI literacy. Students should not be taught that a citation icon means a claim is trustworthy. They should be taught to inspect the source, identify whether the cited passage is primary or secondary, determine whether the source applies to the material/equipment at hand, and recognize when numerical parameters are conditional.

5.2. Implications for AI Literacy in Technical Higher Education

Recent higher-education literature increasingly positions critical evaluation as a central AI-literacy competence [3,4,6]. The present results provide a concrete engineering example of why this matters. A learner using a source-grounded assistant may receive a fluent answer that is well cited but still needs to ask: Is the value universal or formulation-specific? Is this a manufacturer's specification or a textbook example? Does the statement come from a current primary standard? Are the units and measurement conditions comparable? Is the recommended troubleshooting step diagnostic, or merely one of several plausible interventions?

These questions can be incorporated into teaching design. Instead of using a technical assistant only to retrieve answers, instructors can ask students to compare an assistant response with the cited source, classify each numerical recommendation by scope, identify missing boundary conditions, or improve a troubleshooting sequence. Such activities align with the broader move from passive AI use toward critical, responsible, and discipline-specific AI literacy described in recent higher-education research [3,4,6,29]. The assistant then becomes not only a convenience tool but also an object for epistemic training. These are proposed teaching uses based on the response-quality findings; they were not implemented or evaluated in a course, and no student learning or AI-literacy outcomes were measured.

5.3. Knowledge Boundaries as a Design Requirement

The eight dedicated knowledge-boundary prompts show how the assistants handled explicitly signaled uncertainty; they do not establish spontaneous recognition of limits in routine use. Because each prompt identified a universal, missing, or context-dependent condition, the results should be interpreted as performance under explicit cueing. Even with this cueing, technical qualifications sometimes weakened later in the same response. Textile T17 contained a material terminology error despite an initial rejection of universal settings: it mislabeled heat-setting temperatures as Tg. System instructions should therefore require assistants to label numerical values as universal principles, source-specific examples, or manufacturer- and process-specific settings, and to preserve those qualifications throughout the response.

A similar design rule should apply to standards. If a user asks what a standard 'requires', the assistant should distinguish between information taken from the primary normative document and information described in secondary literature. If the primary document is absent, it should state that limitation explicitly. The repeated ISO 12647-2 result shows that this requirement was not automatically satisfied by general source-grounding instructions.

5.4. Repeatability and the Reproduction of Error

The claim-level comparison shows both repeated behavior and meaningful content variation among the ten saved pairs. Recurring patterns included the textile T17/R17 terminology error, the continuation of numerical advice after an initial refusal in T18/R18, and repeated overgeneralization of conductivity thresholds in Offset T17/R17. Variation was also visible: Offset R13 added a clearer diagnostic sequence but introduced further equipment-dependent numerical claims, and the Textile T20/R20 answers changed the illustrative ink formulation while maintaining the same refusal to invent waveform settings. These are within-sample observations from one repeat per selected prompt, not estimates of response stability. The three offset comparisons without original prompt wording remain thematic only. The preserved record does not include NotebookLM source-link metadata showing which retrieved passages were attached to each claim, preventing systematic pairwise citation analysis; the supplementary sheet therefore states this limit and separates content comparison from source verification.

5.5. Relationship to Prior Printing-Technology Research

The current findings extend the authors' earlier work in three directions. The Future-BME development study focused on the architecture and potential applications of multiple domain-specific printing assistants [9]. The subsequent exploratory user study showed that users value rapid access, clarity, and practical usefulness [10]. The present exploratory, expert-audited case study adds a technical accuracy analysis by asking whether the assistants produce contextually appropriate, source-supported answers under deliberately challenging conditions. This progression—from system design, to user perception, to case-based expert audit—provides a basis for later educational trials, rather than a validated benchmark of general assistant performance.

The findings also connect with broader AI developments in graphic engineering. Research has explored AI in packaging, design practice, and machine-learning-based print quality [11–14]. The ASKING AI+ work further illustrates how conversational support can be embedded in a structured graphic-engineering learning environment [15]. The present study differs by focusing on the epistemic behavior of a source-grounded assistant rather than on creative output or user acceptance. That distinction is important for engineering education, where the reliability of technical explanation may be a prerequisite for meaningful classroom integration.

5.6. Limitations

Several limitations should be acknowledged. First, the study involved two assistants from one technical discipline and one primary subject-matter expert; the ratings are expert judgment, not multi-rater consensus. Second, although all 50 recorded responses and complete ratings for 40 original tests are supplied, 15 original offset prompts (T01–T14 and T16) survive only as topic labels, not verified verbatim prompts. Exact cited notebook passages were not systematically archived, although the expert checked sources during evaluation; the T17/R17 source evidence was documented retrospectively from the retained books. In addition, the underlying model version, model-specific generation settings, and a dated export of the notebook sources and custom instructions were not retained

(Section 3.1). Together, these gaps prevent exact independent replication of a substantial part of the experiment, particularly the offset evaluation, and limit broad claim-level citation verification; the study is therefore presented as exploratory. Third, the repeatability check contains one additional generation for each of ten selected cases. Seven pairs have identical prompt wording and three are topic-only offset comparisons. We added a claim-level comparison of the saved texts, but one repeat cannot estimate run-to-run variability or support a formal reliability coefficient. The findings are therefore preliminary and specific to these selected prompts and conditions.

Fourth, the study used no prospectively defined technical answer key or independently validated reference dataset, and included no general-purpose AI or conventional information-retrieval baseline. The expert assessed responses against domain knowledge and available literature, with manual citation checks; this is not equivalent to an independent benchmark. The findings therefore describe these two configured assistants under the tested prompts and sources. They cannot isolate the causal contribution of source grounding or establish that either assistant outperforms a general-purpose model or another retrieval approach. Future work should use a version-controlled reference set based on primary standards and manufacturer documentation, multiple independent raters, and matched baseline systems under the same prompts and conditions.

In addition, the evaluator's involvement in assistant development and the absence of a second independent rater leave potential expectancy bias unquantified. The written score anchors presented in Section 3.3 are a retrospective clarification of the operational judgments, not a prospectively archived rubric; independent validation and inter-rater agreement were not obtained.

Because all eight dedicated boundary prompts explicitly signaled the limitation to be addressed, the study cannot determine how often either assistant would identify boundaries spontaneously in ordinary technical queries. Textile T14–T16 are reported as separate prompt-guided cases and remain outside the focused denominator.

The textile case demonstrates why source support and technical correctness must be evaluated separately. The retained handbook labels the assistant's polyester and nylon figures as heat-setting ranges (§7.4.7), while Table 10.1 reports lower Tg values [26,27]. Future work should incorporate source-quality grading, version control for technical standards, and explicit prioritization of primary manufacturer or normative documentation where relevant.

Finally, although Section 5.2 discusses possible classroom activities, the study included no classroom implementation or student participants and measured no learning or AI-literacy outcomes. These suggestions are prospective pedagogical implications, not empirically tested educational effects.

6. Conclusions

This exploratory, expert-audited case study evaluated two source-grounded AI technical assistants for textile and offset printing using 40 original prompts and ten repeated tests. Both assistants received high descriptive ratings from one subject-matter expert, with mean overall scores of 4.73/5 for textile and 4.61/5 for offset; because their prompt sets differed, these values are descriptive rather than comparative. Source support exceeded technical accuracy in 35 of 40 original prompts. Of the seven repeat pairs with documented identical wording, six received the same overall score (mean absolute difference 0.036/5). Across eight explicitly framed knowledge-boundary prompts, the assistants initially indicated when universal or manufacturer-specific parameters could not be specified, although some answers later included insufficiently qualified examples; this does not establish spontaneous recognition in ordinary prompts. These single-expert ratings are prelimi-

nary, structured expert judgments and remain subjective, not independent validation or generalizable estimates of accuracy; the study is not a validated benchmark, and student learning outcomes were not assessed. The claim-level review of the ten saved pairs also found recurring errors and meaningful content variation; with only one repeat per selected case, these observations remain exploratory and do not estimate run-to-run stability. The teaching activities suggested in Section 5.2 are prospective only and were not implemented or evaluated.

Because the study included neither an independent technical benchmark nor a general-purpose AI or retrieval baseline, it does not show that source grounding improves accuracy or that these assistants outperform alternatives. The findings are limited to the systems, prompts, corpora, and evaluation procedure examined here; future comparative work should use independently validated references and matched baselines.

The results nevertheless show that source grounding should not be equated with technical correctness. The assistants repeatedly overgeneralized source-specific numerical values, occasionally provided troubleshooting information without sufficient diagnostic prioritization, and could misclassify source-supported values or rely on secondary interpretations of standards. The most important case involved the heat-setting/Tg terminology error, which reappeared in the repeated response. The retained handbook identifies those values as heat-setting ranges, not Tg [26,27]. This demonstrates that transparent grounding makes errors easier to audit, but cannot replace expert curation and verification.

For higher education, the practical implication is that source-grounded assistants are best positioned as complementary tools for explanation, guided problem solving, and access to disciplinary literature, not as autonomous technical authorities. Their educational value can be strengthened when students are taught to inspect citations, distinguish primary from secondary sources, identify boundary conditions, and challenge overly universal recommendations. In this sense, reliable use of AI in technical education depends as much on source literacy and contextual judgment as on model capability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app162110557/s1, Supplementary File S1: Evaluation workbook with study setup, prompt-level ratings, full responses, the original verbatim archive, repeatability and claim-comparison sheets, scoring rubric, corpus inventory, source error verification, and documentation check; Supplementary File S2: Source inventories and assistant instruction summaries, including Table S1: Source inventories of the two assistants, and Section SA: Assistant instruction sets used during testing (summaries). Fifteen original offset prompts remain topic labels and are not reconstructed.

Author Contributions: Conceptualization, N.K., G.V. and S.D.; methodology, S.D., Ž.P. and N.M.K.; technical evaluation, N.K.; source verification, N.K.; formal analysis, N.M.K. and S.D.; data curation, N.M.K. and S.D.; supervision, G.V. and Ž.P.; writing—original draft preparation, N.K., S.D. and N.M.K.; writing—review and editing, N.K., S.D. and N.M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-34/2026-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through the project "Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad 2026" (No. 01-3609/1).

Institutional Review Board Statement: Not applicable. The study evaluated AI-system outputs and did not involve human participants, personal data, or an educational intervention with students.

Informed Consent Statement: Not applicable.

Data Availability Statement: The assistants were hosted in Google NotebookLM/Gemini Notebook: Textile Printing Technical Assistant, https://notebook.google.com/notebook/37a236db-59aa-4218-8c71-4b4eae5c33d7 (accessed on 22 September 2026); Offset Printing Technical Assistant, https://notebook.google.com/notebook/b121d6b7-7f98-4b5e-a2a4-c059f72326d2 (accessed on 22 September 2026). Fifteen original offset prompts remain only as topic labels. The sources relevant to Textile T17/R17 were retained in the evaluated corpus, but their page-level evidence was not archived with the original response records. The edition-corrected reference, original technical values, material and process conditions, and source locators are documented in the supplementary workbook's Source Error Verification sheet. The absence of a contemporaneous NotebookLM export remains a limitation; the retrospective source verification does not constitute a dated capture of the interface or a new evaluation.

Acknowledgments: The authors acknowledge the curated educational, scientific, and technical sources incorporated into the assistants' knowledge bases. After the expert evaluation, the authors used OpenAI ChatGPT (GPT-5.6 Sol) to assist with organization, English translation and synthesis of archived evaluation materials, descriptive calculations, comparisons of original and repeated responses and manuscript drafting. The human authors reviewed and edited the AI-assisted output and accept full responsibility for the content of this publication. Use of OpenAI ChatGPT (GPT-5.6 Sol) for post-evaluation organization, English translation and synthesis of archived materials, descriptive calculations, response comparisons, and manuscript preparation is disclosed in Sections 3.3 and 3.5, and Acknowledgments. The tool did not serve as an independent expert evaluator, generate or alter ratings, or replace manual source verification. All technical judgments and interpretations remain the responsibility of the human authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Denecke, K.; Glauser, R.; Reichenpfader, D. Assessing the Potential and Risks of AI-Based Tools in Higher Education: Results from an eSurvey and SWOT Analysis. *Trends High. Educ.* **2023**, *2*, 667–688. [CrossRef]
2. Tillmanns, T.; Salomão Filho, A.; Rudra, S.; Weber, P.; Dawitz, J.; Wiersma, E.; Dudenaite, D.; Reynolds, S. Mapping Tomorrow's Teaching and Learning Spaces: A Systematic Review on GenAI in Higher Education. *Trends High. Educ.* **2025**, *4*, 2. [CrossRef]
3. Wakeman, S.; Johnson, H.; Cary, J.; Endacott, C.; Westine, C.; Liu, Q. AI Literacy: University Students' Perceptions and Practices. *Trends High. Educ.* **2026**, *5*, 44. [CrossRef]
4. Uijl, S.G.; Verhagen, P.; Wiersma, E.M.; Geluk, H.; Oomens, G.; Boor, I. Educating in the Age of GenAI: Exploring AI Literacy Through Mutual Reinforcement in Higher Education. *Trends High. Educ.* **2026**, *5*, 47. [CrossRef]
5. Hu, Z. Exploring Student Acceptance of AI Teaching Assistants in African Higher Education. *Trends High. Educ.* **2026**, *5*, 53. [CrossRef]
6. Schwenke, N.; Söbke, H.; Kraft, E. Potentials and Challenges of Chatbot-Supported Thesis Writing: An Autoethnography. *Trends High. Educ.* **2023**, *2*, 611–635. [CrossRef]
7. Google. Dive Deeper into I/O 2026 with NotebookLM. 27 May 2026. Available online: https://blog.google/innovation-and-ai/products/notebooklm/notebooklm-google-io-2026/ (accessed on 22 September 2026).
8. Google. NotebookLM Is Now Gemini Notebook. 16 July 2026. Available online: https://blog.google/innovation-and-ai/products/gemini-notebook/notebooklm-gemini-notebook/ (accessed on 22 September 2026).
9. Kašiković, N.; Anđelković, A.; Zeljković, Ž.; Popović, L.; Stančić, M.; Dragić, M.; Gazibarić, Z. Digitalization of Expert Knowledge through AI-Based Technical Assistants: A Case Study in Printing Technology. In Proceedings of the 2nd International Conference on Business, Management, and Economics Engineering (Future-BME 2026), Vrdnik, Serbia, 30 September 2026.
10. Kašiković, N.; Anđelković, A.; Zeljković, Ž.; Popović, L.; Stančić, M.; Dragić, M.; Malinović, B. Assessing the Usefulness of AI-Based Technical Assistants in Printing Technology Education and Practice. In Proceedings of the XVI Conference of Chemists, Technologists and Environmentalists of Republic of Srpska, Banja Luka, Bosnia, 16–17 October 2026.
11. Ahmed, K.K. The Power of Artificial Intelligence in Packaging Design. In Proceedings of the 12th International Symposium on Graphic Engineering and Design (GRID 2024), Novi Sad, Serbia, 14 November 2024. [CrossRef]
12. Agung Sudarmanto, J. Exploring the Intersection of AI and Creativity in the Local Indonesian Graphic Designers' Perspective. *J. Graph. Eng. Des.* **2025**, *16*, 27–35. [CrossRef]
13. Ardhianto, P.; Fabian Aquirel, T. Integrating AI Image Generation into the Designer's Creative Process for Mascot Creation. *J. Graph. Eng. Des.* **2026**, *17*, 33–43. [CrossRef]

14. Basak, S.; P.C, S.; Maity, A.; Chandra Paul, K. Prediction of Dot Gain in Flexographic Color Printing Using Machine Learning. *J. Graph. Eng. Des.* **2025**, *16*, 37–49. [CrossRef]
15. Kašiković, N.; Glušac, D.; Premčevski, V.; Anđelković, A.S.; Tasić, N. Artificial Intelligence as a Learning Assistant: Design and Evaluation of the ASKING AI+ Platform for Graphic Engineering Education. In Proceedings of the GRID 2026, Novi Sad, Serbia, 12–14 November 2026.
16. Merino-Campos, C. The Impact of Artificial Intelligence on Personalized Learning in Higher Education: A Systematic Review. *Trends High. Educ.* **2025**, *4*, 17. [CrossRef]
17. Kasneci, E.; Sessler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; et al. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [CrossRef]
18. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ. Sci.* **2023**, *13*, 410. [CrossRef]
19. Maynez, J.; Narayan, S.; Bohnet, B.; McDonald, R. On Faithfulness and Factuality in Abstractive Summarization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 1906–1919. [CrossRef]
20. Liu, N.F.; Zhang, T.; Liang, P. Evaluating Verifiability in Generative Search Engines. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, 6–10 December 2023; pp. 7001–7025. [CrossRef]
21. Magesh, V.; Surani, F.; Dahl, M.; Suzgun, M.; Manning, C.D.; Ho, D.E. Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools. *J. Empir. Leg. Stud.* **2025**, *22*, 216–242. [CrossRef]
22. Ding, Y.; Facciani, M.; Poudel, A.; Joyce, E.; Aguinaga, S.; Veeramani, B.; Bhattacharya, S.; Weninger, T. Citations and Trust in LLM Generated Responses. *AAAI Tech. Track Nat. Lang. Process. I* **2025**, *39*, 23787–23795. [CrossRef]
23. Tufino, E. NotebookLM as a Socratic Physics Tutor: Design and Preliminary Observations of a RAG-Based Tool. *Phys. Educ.* **2025**, *7*, 2550014. [CrossRef]
24. Semenov, S.; Mikhav, V.; Meleshko, Y.; Paranyak, N.; Pochebut, M. Phase-Dynamic Model of User Interactions for Protecting Recommender Systems from Poisoning Attacks. *Appl. Sci.* **2026**, *16*, 3769. [CrossRef]
25. *ISO 12647-2:2013*; Graphic Technology—Process Control for the Production of Half-Tone Colour Separations, Proof and Production Prints—Part 2: Offset Lithographic Processes. International Organization for Standardization: Geneva, Switzerland, 2013.
26. Hall, M.E. Finishing of technical textiles. In *Handbook of Technical Textiles*, 1st ed.; Horrocks, A.R., Anand, S.C., Eds.; Woodhead Publishing: Cambridge, UK, 2000; pp. 152–172.
27. Bajaj, P. Heat and flame protection. In *Handbook of Technical Textiles*, 1st ed.; Horrocks, A.R., Anand, S.C., Eds.; Woodhead Publishing: Cambridge, UK, 2000; pp. 223–263.
28. Miles, L.W.C. Fixation and aftertreatment processes. In *Textile Printing*, 2nd ed.; Miles, L.W.C., Ed.; Society of Dyers and Colourists: Bradford, UK, 2003; pp. 275–291.
29. Anwar, N. The Use of Generative Artificial Intelligence to Develop Student Research, Critical Thinking, and Problem-Solving Skills. *Trends High. Educ.* **2025**, *4*, 34. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.