Jurnal Pengabdian kepada Masyarakat
Volume 4, Issue 3, 290-305
e_ISSN: 2987 - 6052
https://ejournal.gomit.id/index.php/aktual
DOI:10.58723/aktual.v4i3.921

From AI Literacy to Editorial Capability: Evaluating Source-Grounded Generative AI Training for Campus Journalism

Gushevinalti*
Universitas Bengkulu, INDONESIA

Novia Putri Raudah
Universitas Bengkulu, INDONESIA

Muhammad Satria Colossal Utama
Universitas Bengkulu, INDONESIA

Article Info

Article history:

Received: September 10, 2026
Revised: September 22, 2026
Accepted: October 17, 2026

Keywords:

Campus Journalism; Generative Artificial Intelligence; NotebookLM; Source-Grounded AI; Student Press

Abstract

Background: Generative artificial intelligence (GenAI) is increasingly used in journalism, but evidence is limited on whether source-grounded training can strengthen student press capability while preserving verification and editorial judgment.
Aims: To evaluate a hands-on NotebookLM training program for campus press representatives and assess immediate changes in knowledge, operational skills, concrete output production, quality, satisfaction, and implementation intentions.
Methods: A one-day, single-group pre–post program evaluation involved 50 students from HIMA media units and university-level student press organizations at Universitas Bengkulu. Aggregate pre–post percentages, percentage-point changes, normalized gains, participation indicators, product completion, quality ratings, satisfaction, and implementation intentions were analyzed descriptively.
Results: Mean knowledge-related indicator percentages increased from 47.0% to 88.4% (+41.4 percentage points), and operational skills increased from 23.0% to 88.4% (+65.4 points). All participants completed an editorial or research-oriented product; mean product quality was 4.33/5, satisfaction was 4.67/5, and 96% intended to use NotebookLM in routine research.
Conclusion: Higher immediate assessment indicators and complete product production were observed after training. Uncertain measurement properties, the single-group design, and absent follow-up preclude causal conclusions and evidence of sustained newsroom productivity.

To cite this article: Gushevinalti., Raudah, N.P., & Utama, M.S.C. (2026). From AI Literacy to Editorial Capability: Evaluating Source-Grounded Generative AI Training for Campus Journalism. *Aktual: Jurnal Pengabdian kepada Masyarakat, 4*(3), 290-305. https://doi.org/10.58723/aktual.v4i3.921
This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License ©2026 by author/s

INTRODUCTION

Generative artificial intelligence (GenAI) is now part of routine media work, with journalists increasingly anticipating hybrid human-AI workflows rather than replacement (Moravec et al., 2026). Across journalism settings, reported benefits include faster production, automation, creative support, and audience engagement, while concerns remain about technical limitations, authenticity, and ethics (Hanh et al., 2026; Galily, 2026). Because AI outcomes also depend on how work is organized and governed, capability and professional judgment are central to responsible adoption (Frenkel, 2026).

These issues are especially relevant to campus press organizations, where student journalists serve real audiences under time, staffing, training, and resource constraints. Indonesian scholarship highlights both the organizational role of student press and its need for professional standards and institutional protection (Rosman et al., 2021; Komala & S. Noeswa, 2025). Similar pressures in other

*** Corresponding author:**
Gushevinalti, Universitas Bengkulu, INDONESIA. ✉ gushevinalti@unib.ac.id

resource-constrained journalism contexts show that AI-related productivity cannot be separated from organizational viability and editorial responsibility (Stanton & Johnson, 2026).

Journalism and communication education therefore needs structured, context-sensitive AI training. Studies report demand for participatory and ethically grounded integration of AI (Asadi & Sultana, 2026), while communication students often view GenAI as a supplementary rather than substitutive tool (Chung et al., 2026). Core capabilities such as critical thinking, ethical reporting, listening, storytelling, and consequential questioning remain essential (Silcock et al., 2026), consistent with digital-literacy research emphasizing evaluative competence (Reddy et al., 2023). AI literacy is also shaped by institutional rules and learning opportunities (Mekheimer & Tohamy, 2026), and journalism educators increasingly emphasize ethics, fact-checking, and information validation (Jaakkola & Wiik, 2025).

The educational value of GenAI depends on task design and retained human agency. Reviews show that creative and problem-solving benefits vary by tool, task, and learner characteristics (Guo, 2026), and that creative confidence is more favorable when users understand their own contribution (Habib & Gardiner, 2026). NotebookLM-oriented studies likewise argue that AI-supported learning is most useful when it deepens engagement with source material rather than replacing it (Brown & Saurez, 2025; Yeo et al., 2025).

NotebookLM is relevant because it is organized around user-provided sources. This study operationalizes a source-grounded workflow in which users assemble a bounded source set, interrogate it, synthesize information, and trace outputs back to source material. Document-intensive evaluations emphasize the value of retrieval-augmented, source-linked outputs when tasks depend on a defined evidence base (Shor et al., 2025; Tozuka et al., 2025; Tufino, 2025), and related work shows that structured prompting can support document transformation when the source base is explicit (Kamińska & Wyciślik, 2026). Source grounding does not guarantee truth, but it makes provenance more inspectable and therefore better aligned with journalistic verification.

Verification remains a central constraint on AI-assisted journalism. Research identifies persistent methodological gaps in fact-checking evaluation (Baptista et al., 2026), limited professional trust even when GenAI is used for efficiency-oriented tasks (Dierickx & van Dalen, 2026), and continuing governance gaps around transparency, fairness, and responsibility (Porlezza et al., 2026). Information-disorder and AI-search studies further highlight ambiguity, misinformation, epistemic opacity, and platform power (Rahimian & Hasanzadeh, 2026; Weinbrand et al., 2026), reinforcing the need to teach provenance inspection and source-bounded verification.

The risks intensify as AI outputs move closer to publication. AI-assisted news can show standardization, contextual variation, journalistic error, and uneven disclosure (Mera-Fernández et al., 2026), while audience research links automated news to perceptions of accuracy, bias, and trust (Li, 2026). Governance and oversight are also associated with different organizational and audience outcomes (Lavich & Ivanishvili, 2026), and public discussion continues to show salient distrust of GenAI among journalists (Pessianzadeh et al., 2026). Productivity should therefore be treated as quality-adjusted rather than as content volume alone.

The same principle applies to multimodal production. AI-generated news imagery can reproduce bias and alter framing (Elhosary, 2026), while AI-supported creative systems can expand interactive possibilities but still require human evaluation (Ji et al., 2026). Heravi, 2026) similarly frames reliable information as a reusable resource for multiple narratives and formats, suggesting that source-grounded AI is most defensible when it supports evidence curation while editorial control remains human-led.

At Universitas Bengkulu, campus press partners produced approximately five articles per month before the intervention, and manual processing of interviews and documents required two to three

days per article. These pressures are consistent with wider evidence on time-constrained journalism and changing professional relationships (de Jong & Dijkstra, 2026) and with research linking work pressure to professional well-being (Mendiguren Galdospin et al., 2026). The study therefore evaluates a sequence of immediate capabilities from AI understanding and source handling to verification and concrete output rather than assuming long-term productivity gains.

The study addresses four gaps. First, much existing work examines attitudes, framing, trust, governance, or anticipated consequences of GenAI, while fewer studies evaluate a structured intervention through observable task performance. Second, AI literacy is often measured generically, even though journalism requires domain-specific competencies in provenance, verification, transparency, and editorial accountability. Third, training evaluations can over-rely on satisfaction or self-confidence without requiring participants to produce an artifact. Fourth, evidence on micro-level AI impact in journalism is still developing. Cristòfol, 2026), for example, proposes a multidimensional measurement architecture encompassing use patterns, perceived productivity, autonomy and editorial decision-making, ethical risk, audience orientation, and cognitive load, but calls for empirical application and validation. The present study provides one such applied setting, albeit with a narrower and more descriptive dataset.

Accordingly, the study asks three research questions: (RQ1) To what extent did participants' knowledge, understanding, and operational skill change from pre-training to the post-training assessment? (RQ2) Did participants translate the acquired competence into concrete journalistic or research-oriented products of acceptable quality? (RQ3) Do the combined process, performance, and transfer-intention indicators support an immediate editorial-capability pathway after source-grounded AI training? For RQ3, process indicators establish intervention exposure and engagement, performance indicators capture learning and completed output, and transfer-intention indicators capture planned continuation beyond the session. The working hypothesis is that structured NotebookLM training strengthens immediate productive capacity by improving source organization, synthesis, prompt formulation, verification, and conversion of source material into editorially usable outputs. This proposition is intentionally narrower than claiming that the training increases long-term newsroom productivity. The available data permit evaluation of competence and immediate output, but not causal or longitudinal claims about publication frequency, production time, audience engagement, or sustained editorial performance.

Figure 1 summarizes the proposed mechanism and explicitly separates outcomes measured in this study from downstream organizational outcomes that remain untested. The contribution is twofold. Empirically, the study evaluates a complete hands-on intervention with 50 campus press participants and links pre-post learning indicators to concrete output production. Conceptually, it advances a source-grounded AI workflow model for student journalism in which AI literacy is treated as a verification-oriented editorial capability rather than as tool familiarity alone. This framing is consistent with ethical arguments that AI should augment rather than displace human judgment (Duymaz & Tekin, 2023; Silva-Atencio, 2025) and with emerging newsroom evidence favoring hybrid human-AI collaboration (Moravec et al., 2026). The analysis therefore asks not whether GenAI can write, but whether a source-grounded training model can help student journalists work more effectively while remaining responsible for what is ultimately published.



Figure 1. Revised source-grounded AI workflow model. Solid elements indicate outcomes measured in this study; dashed elements indicate proposed downstream newsroom outcomes requiring longitudinal evaluation.

METHOD

Research Design

The study used a single-group pre–post program-evaluation design embedded in a university community-engagement training activity. The intervention was conducted on 2 July 2026 in the Multimedia Laboratory of the Faculty of Social and Political Sciences, Universitas Bengkulu, Indonesia. The setting provided individual computer access, stable internet connectivity, projection facilities, and audio equipment. The intensive session ran from 08:00 to 16:00 and combined conceptual instruction, live demonstration, individual and paired practice, discussion, product development, and mentoring.

The design was pragmatic rather than experimental. Every participant received the intervention, and no untreated or active-comparison group was available. Consequently, observed changes are described as training-associated differences rather than isolated causal effects. This distinction is important in light of recent methodological work arguing that the impact of GenAI on journalistic routines is multidimensional and should be operationalized explicitly rather than inferred from simple technology adoption (Cristòfol, 2026). The present evaluation concentrates on immediate competence and output indicators that were available in the program dataset.

Participants and Sampling

The program involved 50 students representing campus media organizations across Universitas Bengkulu. Thirty participants (60%) represented media or documentation functions associated with Himpunan Mahasiswa (HIMA) at the study-program level, and 20 (40%) represented university-level student press organizations. The cohort included students from communication, sociology, political science, and programs outside the Faculty of Social and Political Sciences. Prior individual experience with GenAI or NotebookLM was not collected.

Recruitment was organized through representation from relevant HIMA and student press units. Because the laboratory could accommodate 50 participants, enrollment was capped at that number. All 50 registered participants attended the intervention, corresponding to 100% attendance. The program was thus implemented with the full planned cohort, but the sample should be interpreted as a purposive training cohort rather than a probability sample of Indonesian student journalists.

Training Intervention

The intervention followed a hands-on, project-based format. This pedagogical choice was consistent with the practical objective of moving participants from awareness of GenAI to demonstrable source-processing and content-production skills. The program was organized around five linked domains: (1) conceptual understanding of generative AI and NotebookLM; (2) source management, including the addition of PDF documents, Google Docs, URLs, and other supported materials; (3) query and prompt formulation for structured retrieval and synthesis; (4) verification through citation tracing and comparison with original source material; and (5) transformation of synthesized information into journalistic or multimedia outputs.

The training began with an introduction to digital literacy, the role of GenAI in communication work, and ethical responsibilities when using AI-generated assistance. The principal trainer, affiliated with Mafindo Bengkulu, then demonstrated the workflow of creating a notebook, adding sources, using source overviews, generating summaries, asking targeted questions, and examining supporting citations. This integration of technical and ethical instruction reflects the broader argument in journalism education that AI should be embedded in participatory learning while maintaining an explicit ethical grounding (Asadi & Sultana, 2026; Silcock et al., 2026).

Practical work formed the core of the intervention. Participants worked with prepared and self-selected sources, tested questions of increasing specificity, compared information across sources,

and inspected the relation between generated outputs and the underlying material. Trainers and assistants circulated during practice to address technical difficulties and to reinforce the rule that machine-generated text was not publication-ready by default. Participants were instructed to verify factual claims against the source set, identify limitations or missing context, and exercise human editorial judgment before using any AI-assisted material.

The final practical task required each participant to produce a concrete work product. Permitted outputs included news or opinion articles, research summaries or reports, feature or reportage scripts, and multimedia or infographic content. This output requirement was important analytically because it converted the intervention from a knowledge-only exercise into a performance-oriented task. The program also included mentoring and feedback intended to help participants translate the tool into their ongoing campus-media routines.

Evaluation Framework and Measures

The evaluation framework comprised four dimensions: input, process, output, and outcome. Input indicators captured whether the program had the resources and participation necessary for implementation, including recruitment target, participant composition, and facility readiness. Process indicators captured how participants engaged with the intervention, including attendance, question participation, quality of interaction, and schedule adherence. Output indicators captured immediate learning and performance through pre–post scores, product completion, and quality ratings. Outcome indicators captured satisfaction, broader competency gains, operational skill acquisition, and intended future implementation. These indicator groups served distinct purposes: knowledge indicators captured conceptual understanding; operational-skill indicators captured task execution; broader-competency indicators summarized cross-cutting literacy and analytical capability; product-quality ratings evaluated completed outputs; satisfaction measured perceived training quality; and implementation intentions measured planned, rather than observed, post-training use. Pre–post percentages, participant-reported satisfaction and intentions, and assessments of completed products represent different measurement sources and are analyzed separately. Because the knowledge category includes practical ability, it should not be interpreted as a pure knowledge scale. The administration mode and item wording were unspecified, so the operational and broader-competency percentages cannot be assumed to represent independently observed performance.

Knowledge-related assessment covered five domains: understanding of AI concepts, understanding of NotebookLM, practical ability, understanding of digital ethics, and understanding of verification. Operational skill assessment covered access to NotebookLM, uploading source materials, formulation of effective questions, analysis of generated outputs, and production of research or journalistic work. A broader competency set assessed digital and AI literacy, digital ethics and verification, and research and information analysis.

Product quality was rated on a five-point scale across six dimensions: correct use of NotebookLM, quality of analysis and synthesis, accuracy of information and citations, structure and writing, creativity and originality, and application of digital ethics. Satisfaction was assessed using eight items on a five-point scale covering relevance of material, trainer delivery, training materials, practical and mentoring sessions, facilities, benefit for self-development and the press organization, and willingness to join future training. Implementation intentions recorded whether participants planned to use NotebookLM in routine research, teach peers, use the tool for academic work, develop research-based content, and apply digital ethics. The identity, number, qualifications, and independence of product assessors, scoring anchors, weighting of dimensions, timing of scoring, and handling of disagreements were unspecified. Consequently, neither an independent quality standard nor a threshold for acceptable publication quality can be established from the overall rating.

The measurement architecture overlaps with dimensions proposed in newer frameworks for studying GenAI in journalism, particularly patterns of use, productivity, ethical risk, and editorial decision-making (Cristòfol, 2026). However, the evaluation instruments were designed for program monitoring rather than formal psychometric validation. The dataset therefore does not support claims about latent constructs, scale reliability, or validated composite indices. Measurement documentation does not establish the number of test items, confirmation that identical items were used at pre- and post-test, the maximum raw score, item-level scoring rules, formal content-validation procedures, rater independence, or reliability coefficients. Accordingly, the percentages and five-point ratings are treated as program-monitoring metrics rather than validated psychometric measures.

Analysis Plan

The analysis was descriptive because the dataset contained aggregate percentages and aggregate mean ratings rather than participant-level observations. For each pre-post indicator, absolute percentage-point change was calculated as post-training percentage minus pre-training percentage. Normalized gain was calculated as (post - pre)/(100 - pre), indicating the proportion of the maximum possible improvement achieved from the observed baseline. Normalized gain was included to express observed improvement relative to the remaining possible improvement from each baseline value; it was used for descriptive context only and not as an inferential or standardized effect size. Domain means are unweighted arithmetic averages of the listed indicator percentages; they are not participant-level mean test scores. Normalized gains assume that each percentage uses a comparable scale with a maximum of 100; uncertainty about the original scoring limits their interpretation.

Normalized gain was interpreted only as a descriptive transformation of baseline-to-post improvement and not as an inferential effect size. Participant-level paired scores, standard deviations, item-response distributions, and paired identifiers were unavailable. Paired-samples t tests, Wilcoxon signed-rank tests, confidence intervals around mean score change, and standardized mean-difference measures such as Cohen's d could therefore not be calculated validly.

For categorical indicators with observed counts, frequencies and percentages were reported. Product-quality and satisfaction ratings were summarized using the aggregate arithmetic means recorded in the dataset. Implementation intentions were expressed as numbers and percentages of the 50 participants.

Analytical Scope

Interpretation followed a staged mechanism: training exposure → source-grounded AI competence → improved source handling and verification → concrete content production → potential enhancement of campus press productivity. Evidence was treated as more persuasive when it converged across multiple stages. This prevents a common analytical shortcut in which increased post-test knowledge is equated automatically with productivity.

The model also distinguishes immediate productive capacity from sustained organizational productivity. Immediate productive capacity is evidenced by the ability to perform source-grounded tasks and generate an assessable work product during the intervention. Sustained organizational productivity would require post-training evidence such as publication frequency, time per article, editor revision burden, source-verification rates, content diversity, or audience engagement. The latter outcomes were not available in the current dataset and are therefore treated as future evaluation targets rather than observed effects.

The analysis also adopted a conservative interpretive standard. It did not treat source grounding as a guarantee of truth, does not infer that participants became professional fact-checkers, and does not claim long-term productivity improvement from immediate output data. This approach is consistent

with literature emphasizing that AI use in communication requires transparency, accountability, verification, and continuing human oversight (Porlezza et al., 2026; Dierickx & van Dalen, 2026).

RESULTS AND DISCUSSION

Participation and Training Process

The intervention achieved the planned recruitment target and complete attendance. All 50 registered participants attended, consisting of 30 HIMA representatives and 20 university-level student press representatives. The planned 60:40 participant composition was therefore preserved. Facilities and training resources covered seven facility categories, including the laboratory, 50 participant computers, internet connectivity, projection, audio support, training materials, and 20 source files for practice.

Participation during the session was high. Forty-seven of 50 participants (94%) asked at least one question. A total of 124 questions were recorded, corresponding to 2.48 questions per participant. Fifteen participants (30%) were categorized as very active, asking more than three questions; 20 (40%) were categorized as active, asking two or three; 12 (24%) asked one question; and three (6%) asked none. The mean interaction-quality score was 4.64/5 across question relevance, depth, trainer responsiveness, discussion participation, and learning atmosphere. Schedule adherence was reported as seven of eight sessions (87.5%). The simultaneous count of two five-minute delays is inconsistent with that denominator; the timing figures are therefore not used as evidence of implementation fidelity. Attendance and question participation indicate engagement with the intervention but do not independently establish learning, uniform exposure duration, or implementation fidelity across all activities.

Knowledge and Operational Capability

All five knowledge-related indicators increased from pre-training to post-training assessment. The mean across the five indicators rose from 47.0% to 88.4%, corresponding to an absolute gain of 41.4 percentage points. The normalized gain at the mean level was 0.78, indicating that approximately 78% of the maximum possible improvement from the observed baseline was realized.

Understanding of NotebookLM and practical ability showed the largest absolute increases. NotebookLM understanding rose from 40% to 90%, a 50-point gain and normalized gain of 0.83. Practical ability rose from 35% to 85%, also a 50-point gain, with normalized gain of 0.77. Verification understanding increased from 45% to 87%, a 42-point gain and normalized gain of 0.76. Understanding of AI concepts rose from 55% to 88% (+33 points; normalized gain 0.73), while understanding of digital ethics increased from 60% to 92% (+32 points; normalized gain 0.80).

Operational skills showed even larger descriptive changes. The mean across five operational indicators increased from 23.0% before training to 88.4% after training, an absolute gain of 65.4 percentage points. The normalized gain was 0.85.

The ability to produce research or journalistic work increased from 15% to 88%, the largest absolute gain at 73 percentage points, with normalized gain of 0.86. Ability to upload source data rose from 25% to 92% (+67 points; normalized gain 0.89). Ability to access NotebookLM increased from 30% to 95% (+65 points; normalized gain 0.93). Ability to construct effective questions rose from 20% to 85% (+65 points; normalized gain 0.81). Ability to analyze outputs increased from 25% to 82% (+57 points; normalized gain 0.76).

Table 1. Aggregate Pre-Post Changes Across Knowledge, Skill, and Competency Indicators

Domain	Indicator	Pre (%)	Post (%)	Gain (pp)	Normalized gain
Knowledge	AI concept understanding	55	88	33	0.73
Knowledge	NotebookLM understanding	40	90	50	0.83
Knowledge	Practical ability	35	85	50	0.77
Knowledge	Digital ethics understanding	60	92	32	0.80
Knowledge	Verification understanding	45	87	42	0.76
Operational skill	Access NotebookLM	30	95	65	0.93
Operational skill	Upload sources	25	92	67	0.89
Operational skill	Construct effective questions	20	85	65	0.81
Operational skill	Analyze outputs	25	82	57	0.76
Operational skill	Produce research/journalistic work	15	88	73	0.86
Broader competence	Digital & AI literacy	55	90	35	0.78
Broader competence	Digital ethics & verification	50	88	38	0.76
Broader competence	Research & information analysis	45	85	40	0.73

Note. Values are aggregate program-monitoring metrics. Gain (pp) = percentage-point difference; normalized gain = (post - pre)/(100 - pre). Participant-level distributions were unavailable; normalized gain is reported only to contextualize proportional improvement from baseline and is not an effect-size estimate.

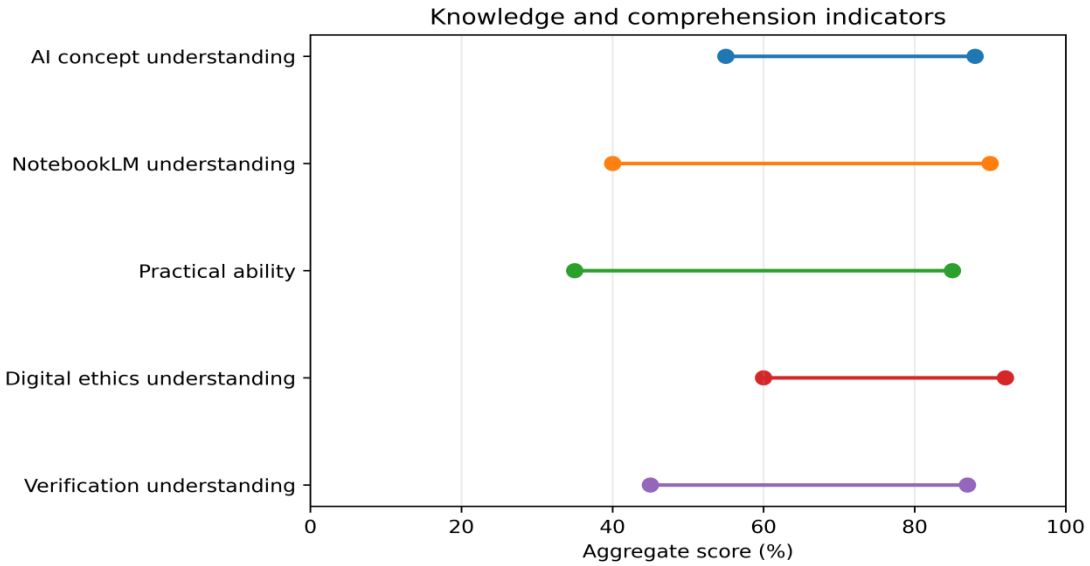


Figure 2. Aggregate pre-post knowledge and comprehension indicators. Values are program-level percentages and do not represent participant-level score distributions.

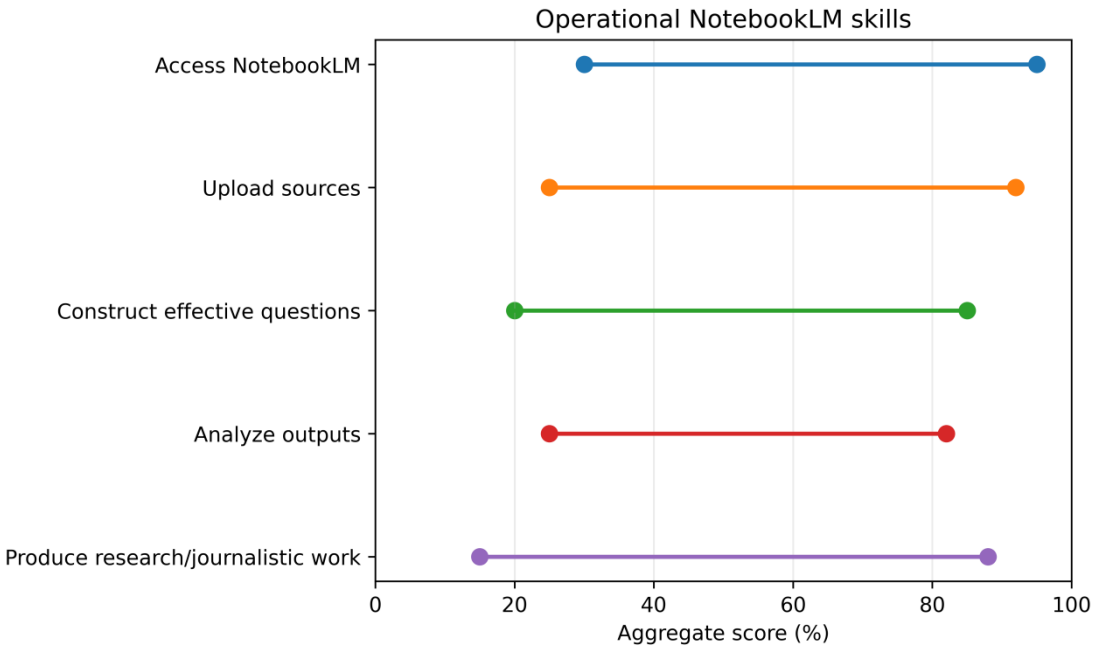


Figure 3. Aggregate pre-post operational NotebookLM skill indicators. Values are program-level percentages and do not represent participant-level score distributions.

The broader competency indicators moved in the same direction. Digital and AI literacy increased from 55% to 90%, digital ethics and verification from 50% to 88%, and research and information analysis from 45% to 85%. The three-item aggregate therefore rose from 50.0% to 87.7%, a 37.7-point gain. Table 1 presents all knowledge, skill, and broader competency indicators together, and Figure 3 visualizes the operational-skill changes.

Editorial Outputs and Product Quality

All 50 participants completed a concrete product during the practical session. This completion rate provides a task-based indicator distinct from assessment percentages and participant-reported perceptions. Completion alone does not establish independent competence or publication readiness.

Twenty-two participants (44%) produced a news or opinion article, 15 (30%) produced a research summary or report, eight (16%) produced a feature or reportage product, and five (10%) produced multimedia or infographic content. Figure 4 shows the distribution. The outputs therefore spanned both conventional text genres and more diversified media forms. This heterogeneity indicates that the training workflow was not restricted to one fixed writing template. The overall product-quality mean was 4.33/5.

Satisfaction and Intended Implementation

Participant satisfaction was high across all eight assessed dimensions, with an overall mean of 4.67/5. Willingness to attend future training received the highest score (4.9), followed by trainer quality and benefit for self-development (both 4.8). Relevance of the material to participant needs and usefulness for the press organization each scored 4.7. The practical and mentoring component scored 4.6, training materials 4.5, and facilities 4.4. The eight displayed rounded item means average 4.675. The overall reported value of 4.67 is retained because unrounded item values and aggregation weights are unavailable; its exact calculation cannot be reconstructed.

Implementation intentions were also strong. Forty-eight participants (96%) intended to use NotebookLM in routine research, 47 (94%) intended to apply digital ethics, 45 (90%) intended to use the tool for academic writing, 42 (84%) intended to teach it to organizational peers, and 40 (80%) intended to develop research-based content. The mean endorsement across these five intended actions was 88.8%.

These measures should be interpreted as intentions rather than verified behavior. They cannot demonstrate that participants subsequently changed organizational routines. Nevertheless, they indicate that the intervention was perceived as transferable beyond the training room. The 84% intention to teach peers is especially relevant to campus press organizations because annual membership turnover can erode organizational knowledge unless competencies are deliberately transferred.

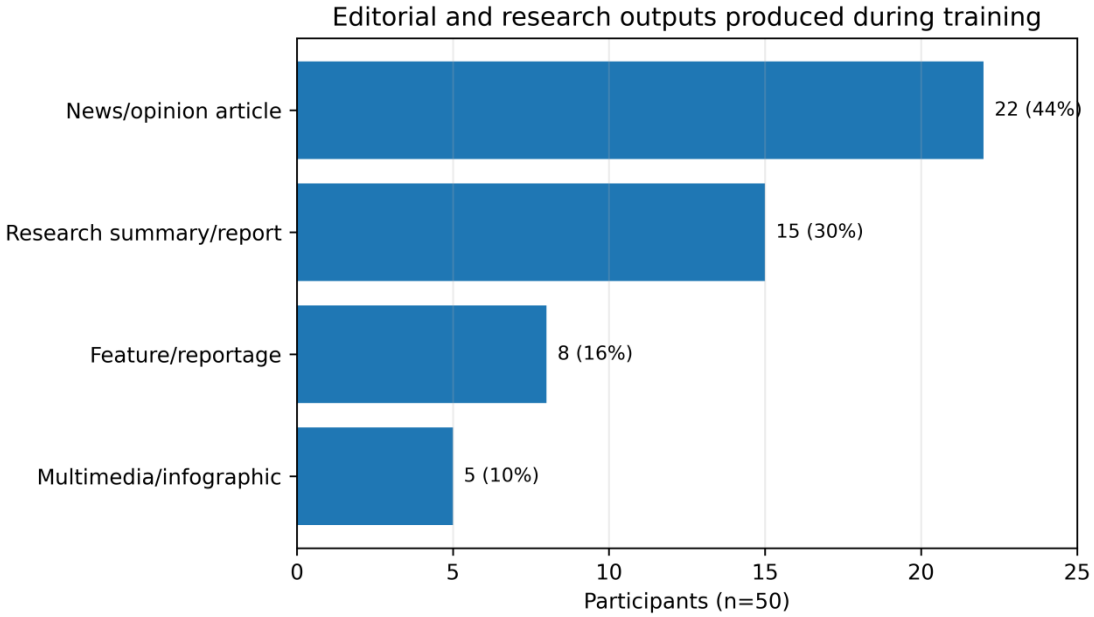


Figure 4. Distribution of the 50 concrete editorial or research outputs completed during the practical training session.

Table 2. Convergent Process, Output, Quality, and Implementation Indicators.

Indicator	Count	Value	Interpretation
Attendance	50/50	100%	Complete intervention exposure
Participants asking >=1 question	47/50	94%	High active engagement
Total questions	124	2.48/person	Sustained interaction
Concrete work produced	50/50	100%	Immediate behavioral output
Mean product quality	-	4.33/5	Positive program quality rating
Mean satisfaction	-	4.67/5	Very high satisfaction
Routine NotebookLM research intention	48/50	96%	High transfer intention
Intent to teach peers	42/50	84%	Potential organizational diffusion
Intent for research-based content	40/50	80%	Potential editorial application

Integrated Descriptive Evidence

Table 2 brings together different types of immediate evidence: attendance and questions describe participation; pre–post indicator percentages describe assessment change; completed products describe task completion; quality ratings describe product appraisal; and satisfaction and intentions

describe participant perceptions. Their coexistence is consistent with the proposed framework but does not test temporal mediation or validate a capability pathway.

The descriptive pattern is compatible with immediate learning and application during training. It does not establish an intervention effect, verified publication quality, or sustained organizational productivity. Publication frequency, production time, editorial revision burden, factual-error rate, audience engagement, and longitudinal quality were not measured.

Discussion

Source-grounded AI as productive editorial capability

The principal finding is the coexistence of higher post-training assessment indicators and complete task production. This pattern extends the evaluation beyond technology acceptance, although it does not establish a causal progression. The intervention required participants to organize evidence, formulate questions, inspect model outputs, and convert the resulting synthesis into editorial products. This interpretation is consistent with evidence that communication students view GenAI as a supplementary tool and that journalists anticipate hybrid rather than replacement models of human-AI work (Chung et al., 2026; Moravec et al., 2026). It also aligns with research showing that efficiency benefits depend on skilled users, organizational participation, and ethical safeguards (Hanh et al., 2026; Frenkel, 2026). For campus press management, the implication is that tool access should be paired with source protocols, verification routines, and editorial oversight rather than treated as a stand-alone productivity intervention.

The larger absolute operational change (65.4 percentage points) than knowledge-related change (41.4 points) is partly consistent with lower operational baselines and greater room for improvement. Repeated hands-on practice and immediate access to mentoring offer plausible explanations, but these mechanisms were not tested. Familiarity with assessment tasks, expectancy, and instructor support may also contribute. Because comparable item scoring and paired observations are unavailable, the two changes should not be interpreted as statistically different effects.

The overall product rating of 4.33/5 describes favorable appraisal within the program. Dimension-specific scores are unavailable, so relative strengths in accuracy, tool use, creativity, or originality cannot be inferred. The division of labor between AI assistance and human editorial judgment is therefore a pedagogical implication rather than a demonstrated quality profile. GenAI can support divergent and convergent thinking, problem solving, and iterative refinement, but its effects remain conditional on task, learner, and design (Guo, 2026). Creative confidence is also more favorable when users retain agency and can identify their own contribution (Habib & Gardiner, 2026). For journalism, the implication is that source extraction, comparison, and synthesis may be partially delegated to AI, whereas framing, interviewing, interpretation, narrative voice, and final publication responsibility should remain human-led. This is especially important because AI-assisted news can become standardized or exhibit journalistic errors when governance and editorial control are weak (Mera-Fernández et al., 2026). Newsroom interviews likewise show that journalists value GenAI for efficiency and data handling while remaining concerned about accuracy, credibility, bias, and responsible use (Cools & Diakopoulos, 2026).

Verification as a constraint on productivity

Source grounding should be understood as a verification affordance, not a guarantee of truth. Heravi, 2026) argues that AI-mediated journalism increasingly shifts value toward the stewardship of reliable information artifacts from which multiple narratives can be assembled. That perspective matches the training design: participants built a bounded evidence base, interrogated it, and converted its contents into different output forms. Work on NotebookLM and retrieval-augmented workflows likewise emphasizes the utility of bounded source collections for traceable document

work (Shor et al., 2025; Tozuka et al., 2025; Tufino, 2025), while Brown & Saurez, 2025; Yeo et al., 2025) caution against replacing engagement with source texts by passive consumption of generated summaries.

Verification therefore functions as a productivity condition rather than as a competing objective. European fact-checkers use GenAI mainly for secondary, efficiency-oriented tasks while maintaining limited trust in model outputs (Dierickx & van Dalen, 2026), and fact-checking scholarship continues to identify methodological gaps in evaluating verification effectiveness (Baptista et al., 2026). Governance research similarly shows persistent gaps around transparency, fairness, and responsibility (Porlezza et al., 2026), while studies of information disorder and AI-generated search summaries emphasize ambiguity, misinformation, epistemic opacity, and platform power (Rahimian & Hasanzadeh, 2026; Weinbrand et al., 2026). Trust and distrust in GenAI coexist among journalists and wider publics, and quality news media frame the technology through simultaneous opportunity and threat (Pessianzadeh et al., 2026; Hendrickx & Coppenolle, 2026). Evidence from automated journalism and AI-mediated summaries further shows that transparency, disclosure, factuality, and oversight shape perceived credibility and organizational legitimacy (Li, 2026; Lavich & Ivanishvili, 2026; Shi et al., 2026). The defensible target is therefore verified output, not output volume alone.

From individual capability to organizational newsroom adoption

The intervention was project based, interactive, and output oriented. Ninety-four percent of participants asked at least one question, and interaction quality averaged 4.64/5. This active format is consistent with journalism-education research recommending contextualized, participatory, ethically grounded AI integration while preserving critical thinking, ethical reporting, storytelling, and consequential questioning (Asadi & Sultana, 2026; Silcock et al., 2026). Because AI literacy is unevenly distributed and institutionally recognized in different ways, campus press organizations should pair tool training with peer mentoring, shared protocols, and repeated practice (Mekheimer & Tohamy, 2026). A recent journalism-assessment intervention also emphasizes authentic, scaffolded GenAI activities that make ethical decision-making, prompt engineering, and verification explicit learning outcomes (Ngu & Weller, 2026).

The present findings describe immediate assessment change and supported task completion; sustained newsroom productivity remains untested. This distinction aligns with Cristòfol's (2026) multidimensional framework for measuring GenAI impact through use patterns, productivity, autonomy, ethical risk, audience orientation, and cognitive load. It also reflects scholarship showing that journalism operates under time pressure, changing professional relationships, and work-context pressures that affect quality and well-being (de Jong & Dijkstra, 2026; Mendiguren Galdospin et al., 2026). Media automation creates value only when embedded in a broader production system rather than treated as a stand-alone tool (Galily, 2026). For campus press, the next analytical level is therefore organizational: whether source-grounded workflows alter actual production cycles, editorial revision, division of labor, and audience engagement over time.

Implications

For campus press organizations, the practical model is straightforward: build a source pack; use AI for bounded extraction, comparison, chronology, and question generation; trace consequential claims back to sources; conduct human verification and additional reporting where the source pack is insufficient; and retain human responsibility for framing, authorship, and publication. Multimodal production should be governed with the same caution because AI-generated visual content can reproduce bias and alter framing (Elhosary, 2026), even as AI-supported creative systems expand interactive possibilities (Ji et al., 2026). In the Global South, this workflow should be contextualized to local language, infrastructure, institutional policy, and training access rather than imported as a universal technological solution.

Research contribution

This study contributes a source-grounded AI workflow model for student journalism and proposes verified productivity as a quality-adjusted research target. Verified productivity is defined here as editorially usable, source-traceable, human-checked output generated with an efficiency gain that does not relax verification or editorial responsibility. This construct is intended to move future research beyond simple measures of tool adoption, satisfaction, or content volume.

Limitations

The study remains an exploratory program evaluation. It used a single-group design with aggregate rather than participant-level data; the monitoring instruments were not formally psychometrically validated; product-quality ratings lacked documented rater independence and inter-rater reliability; post-training outcomes were immediate; implementation intentions were not observed behavior and cannot be treated as evidence of actual implementation; and the study did not measure publication frequency, production time, error rate, editor revision burden, or audience engagement. The cohort came from one university, which limits generalizability, and prior individual experience with GenAI or NotebookLM was not collected. The purposive sample of 50 participants was determined by laboratory capacity, without a reported precision or power calculation. Selection, testing, expectancy, and social-desirability effects remain possible. The lack of paired distributions prevents assessment of individual variation, and product completion during mentored practice cannot establish independent performance. Future evaluations should verify tool use and editorial practice several weeks and months after training rather than relying on immediate intentions alone.

Suggestions

Future studies should retain anonymous paired raw scores, validate instruments, use multiple blinded raters, and compare source-grounded AI training with generic chatbot training and conventional workflows. Key outcomes should include time to first publishable draft, factual-error rate, citation traceability, source diversity, editor revision time, and one- and three-month newsroom metrics. A composite verified-productivity index could reward useful output while penalizing unsupported claims and excessive revision. Qualitative interviews or usage diaries could examine autonomy and cognitive load, and transparency experiments could test how audiences respond to different forms of AI disclosure.

CONCLUSION

The single-group evaluation of 50 campus press participants showed higher post-training aggregate knowledge-related and operational indicators and completion of a journalistic or research-oriented product by every participant. These findings describe immediate assessment change and application during mentored training. The overall product rating and favorable implementation intentions complement these results but do not establish independent publication quality or subsequent use. The design, uncertain measurement properties, and absence of longitudinal data preclude causal conclusions and claims of sustained newsroom productivity. Comparative studies with documented instruments, independent product assessment, and follow-up are needed to evaluate whether these immediate outcomes persist.

ACKNOWLEDGMENT

The authors acknowledge LPPM Universitas Bengkulu, Mafindo Bengkulu, the Communication Science Faculty of Social and Political Sciences at Universitas Bengkulu, and the participating HIMA and campus press organizations for supporting the implementation of the training and evaluation activities.

AUTHOR CONTRIBUTION STATEMENT

GV: conceptualization, supervision, methodology, project administration, and manuscript review. NPR and MSCU: training assistance, documentation, data organization, and initial analysis. All authors reviewed and approved the manuscript.

AI DISCLOSURE STATEMENT

The authors used ChatGPT (OpenAI) during manuscript preparation for structural refinement, language editing, and formatting support. The authors reviewed and edited the resulting text and take full responsibility for the content, interpretation, citations, and final publication.

CONFLICTS OF INTEREST

The authors declare that they have no financial, institutional, or personal conflicts of interest that could have influenced the conduct of this activity, the analysis of the data, the preparation of the manuscript, or its publication.

REFERENCES

Asadi, S., & Sultana, T. (2026). Caught in the AI Current: J-School Educators' Disruption, Adaptation and Future Readiness in the Global South Insights from Bangladesh. *Journalism and Mass Communication Educator*, 81(3), 247–268. https://doi.org/10.1177/10776958251407392

Baptista, J. P., Rivas-de-Roca, R., Ribeiro, F., & Battista, D. (2026). Shadows and Lights on Fact-Checking Research: Trends, Methodologies, and Empirical Approaches. *Communication and Society*, 39(1), 522–540. https://doi.org/10.15581/003.39.1.034

Brown, V., & Saurez, D. (2025). NotebookLM: Revolutionizing Learning for Students with Neurodivert Challenges using AI and Universal Design Principles. *FDLA Journal*. https://nsuworks.nova.edu/fdla-journal/vol9/iss1/22

Chung, A., Chon, M.-G., & Kinsey, D. F. (2026). Generative AI Through the Eyes of Aspiring PR Professionals in the United States: A Q-Methodology Study. *Journalism and Mass Communication Educator*, 81(2), 143–164. https://doi.org/10.1177/10776958251382567

Cools, H., & Diakopoulos, N. (2026). Uses of Generative AI in the Newsroom: Mapping Journalists' Perceptions of Perils and Possibilities. *Journalism Practice*, 20(3), 878–896. https://doi.org/10.1080/17512786.2024.2394558

Cristòfol, F. J. (2026). Measuring the Micro-Level Impact of Generative AI on Science Journalists' Professional Routines: A Methodological Model for Science Communication Contexts. *Iberoamerican Journal of Science Measurement and Communication*, 6(2). https://doi.org/10.47909/ijsmc.340

de Jong, A., & Dijkstra, A. M. (2026). Current Changes in the Relationship Between Scientists and Journalists. A Systematic Literature Review. *Bulletin of Science, Technology and Society*, 46(2), 63–79. https://doi.org/10.1177/02704676261452144

Dierickx, L., & van Dalen, A. (2026). Using It but Not Trusting It: Technology Acceptance and Generative AI in European Fact-Checking. Emerging Media, 4(2 Theme: Generative AI on Fact-checking and Media Reporting), 248–269. https://doi.org/10.1177/27523543261449974

Duymaz, Y. K., & Tekin, A. M. (2023). Harnessing Artificial Intelligence in Academic Writing: Potential, Ethics, and Responsible Use. *European Journal of Therapeutics*, 30(1), 87–88. https://doi.org/10.58600/eurjther1755

Elhosary, M. (2026). Pixels of Prejudice: Decoding Embedded Biases in AI-Generated News Imagery and Their Implications for Visual Journalism Toward an Algorithmic-Mediated Visual Framing. *Journalism and Mass Communication Quarterly*, 103(2), 430–461. https://doi.org/10.1177/10776990251360690

Frenkel, S. J. (2026). New Technology and Work: AI and the Coming Struggle to Democratise Work Organisation. *Sociology Compass*, 20(6). https://doi.org/10.1111/soc4.70212

Galily, Y. (2026). The Next Frontier in Sports Media: Intelligent Automation, Generative Artificial Intelligence, and the Transformation of Audience Engagement. Emerging Media, 4(2 Theme: Generative AI on Fact-checking and Media Reporting), 233–247. https://doi.org/10.1177/27523543261446668

Guo, Q. (2026). How Generative Artificial Intelligence Influences the Creative Thinking of STEAM Learners in Higher Education: A Systematic Literature Review. *Thinking Skills and Creativity*, 62. https://doi.org/10.1016/j.tsc.2026.102336

Habib, S., & Gardiner, Z. (2026). Generative AI and Creative Self-Efficacy: A Systematic Review. *Thinking Skills and Creativity*, 62. https://doi.org/10.1016/j.tsc.2026.102234

Hanh, K. M., Ming, O., & Vincent, M. (2026). Framing Artificial Intelligence Adoption in State-Controlled Media: An Analysis in Vietnamese Journalism. *Social Sciences and Humanities Open*, 13. https://doi.org/10.1016/j.ssaho.2025.102425

Hendrickx, J., & Coppenolle, I. V. (2026). A Tool, a Threat, or Both? How Quality News Media Frame Generative AI. Emerging Media, 4(2 Theme: Generative AI on Fact-checking and Media Reporting), 292–307. https://doi.org/10.1177/27523543261422926

Heravi, B. (2026). AI-Mediated Storytelling in A Liquid Information Ecosystem. *Journalism and Media*, 7(2). https://doi.org/10.3390/journalmedia7020114

Lavich, M., & Ivanishvili, T. (2026). Trust Dynamics and Economic Implications of Generative AI Adoption in Digital Journalism. *Journalism and Media*, 7(2). https://doi.org/10.3390/journalmedia7020102

Jaakkola, M., & Wiik, J. (2025). No-Panic Pragmatism: Nordic Journalism Educators' Preparedness to Teach Artificial Intelligence (AI) in 2024. *Journalistica*, 19(1), 198–224. https://doi.org/10.7146/journalistica.v19i1.151947

Ji, D., Ju, R., & Li, Z. (2026). The Intersection of Artificial Intelligence and Digital Media Art: Exploring New Frontiers of Creativity. *Humanities and Social Sciences Communications*, 13(1). https://doi.org/10.1057/s41599-026-07132-z

Kamińska, A. M., & Wyciślik, Ł. (2026). Towards Objective Abstracts and Keywords: The Helpful Hand of GenAI. *Journal of Scholarly Publishing*, 57(3), 470–491. https://doi.org/10.3138/jsp-2025-0029

Komala, R., & S. Noeswa, W. (2025). Protection of Student Press in Journalism: Student Press Jurnis.Id-UNIS, Tangerang Post-Collaboration Between the Press Council and the Directorate General of Higher Education. *Jurnal Ranah Komunikasi (JRK)*, 9(2). https://doi.org/10.25077/rk.9.2.91-105.2025

Li, Y. (2026). The Impact of Automated Journalism on Media Bias, Accuracy, and Public Trust: Evidence from Young Chinese News Consumers. *Humanities and Social Sciences Communications*, 13(1). https://doi.org/10.1057/s41599-026-06612-6

Mekheimer, M., & Tohamy, G. (2026). AI Literacy Functions As Cultural Capital in Stratified Egyptian Schools. *Discover Education*, 5(1). https://doi.org/10.1007/s44217-026-01695-4

Mendiguren Galdospin, T., Prados Rodríguez, R., Morales I Gras, J., & Ganzabal Learreta, M. (2026). Journalists and Well-Being on TikTok: A Comparison Between Spain and the UK. *Index. Comunicacion*, 16(2), 197–217. https://doi.org/10.62008/ixc/16/02Period

Mera-Fernández, M., Moreno-Gil, V., & Morata-Santos, M. (2026). AI-Generated Content in Spanish Media: Transparency, New Uses, and Defined Strategies. *Journalism and Media*, 7(2). https://doi.org/10.3390/journalmedia7020113

Moravec, V., Gavurova, B., & Rigelsky, M. (2026). New Reality of Public Service Media Journalists: How Generative AI Redefines Journalism and Work Practices. *Telecommunications Policy*, 50(6). https://doi.org/10.1016/j.telpol.2026.103210

Ngu, I. Y., & Weller, D. (2026). From Classroom Design to Newsroom Practice: Designing Genai Assessment Intervention. *Journal of University Teaching and Learning Practice*. https://doi.org/10.53761/4281ht04

Pessianzadeh, A., Sultana, N., Van den Bulck, H., Gefen, D., Jabbari, S., & Rezapour, R. (2026). In Generative AI We (Dis) Trust? Computational Analysis of Trust and Distrust in Reddit Discussions. *Transactions of the Association for Computational Linguistics*, 14, 1612–1638. https://doi.org/10.1162/TACL.a.744

Porlezza, C., Amigo, L., Eberwein, T., Juurik, M., Milosavljević, M., Parder, M.-L., Rathmann, M., & Rozgonyi, K. (2026). Tackling the Challenges of Artificial Intelligence for Public Communication: Governance and Accountability Discourses About the Responsible Use of AI. *New Media and Society*, 28(7), 2886–2911. https://doi.org/10.1177/14614448261445061

Rahimian, S., & Hasanzadeh, M. (2026). Extraction and Analysis of Information Disorder Components in the Formation of Extra-Informationalism. *International Journal of Knowledge Processing Studies*, 6(1), 50–64. https://doi.org/10.22034/kps.2026.569161.1266

Reddy, P., Chaudhary, K., & Hussein, S. (2023). A Digital Literacy Model to Narrow the Digital Literacy Skills Gap. *Heliyon*, 9(4). https://doi.org/10.1016/j.heliyon.2023.e14878

Rosman, A., Zakiah Prodi Jurnalistik, K., & Ilmu Komunikasi, F. (2021). Strategi Komunikasi Massa Lembaga Pers Mahasiswa. *Prosiding Jurnalistik*, 7. https://doi.org/10.29313/.v7i1.25659

Shi, S., Zhang, Y., Xu, X., Filieri, R., & Jia, R. (2026). Biased or Credible? Investigating Perceived Persuasiveness of AI-Generated Review Summaries in Online Travel Platforms. *Tourism Management*, 117. https://doi.org/10.1016/j.tourman.2026.105446

Shor, R., Greene, E. A., Sumberg, L., & Weingrad, A. B. (2025). AI Tools in Academia: Evaluating NotebookLM as a Tool for Conducting Literature Reviews. *Psychiatry* (New York). https://doi.org/10.1080/00332747.2025.2541531

Silcock, B. W., Davie, W. R., & Guo, L. (2026). The Heart of the Interview: Al Tompkins on Teaching, Trust, and the Next Era of Broadcast Journalism Education. *Journalism Education Trends and Techniques in Broadcast and Mobile Journalism*, 20(2 Special Issue), 78–90. https://doi.org/10.1177/19312431261430063

Silva-Atencio, G. (2025). The Challenges and Opportunities for Ethics in Generative Artificial Intelligence in the Digital Age. *DYNA* (Colombia), 92(236). https://doi.org/10.15446/dyna.v92n236.117144

Stanton, R., & Johnson, M. R. (2026). "You Can't Work for Somebody Who Doesn't Exist": the Decline of Games Journalism and the Rise of Independent Gaming News. Games and Culture, 21(4), 704–723. https://doi.org/10.1177/15554120241273332

Tozuka, R., Johno, H., Amakawa, A., Sato, J., Muto, M., Seki, S., Komaba, A., & Onishi, H. (2025). Application of NotebookLM, A Large Language Model with Retrieval-Augmented Generation, for Lung Cancer Staging. *Japanese Journal of Radiology*, 43(4). https://doi.org/10.1007/s11604-024-01705-1

Tufino, E. (2025). NotebookLM: An LLM with RAG for Active Learning and Collaborative Tutoring. *arXiv*, 2. https://doi.org/10.48550/arXiv.2504.09720

Weinbrand, S., Nagappa, A., & Angus, D. (2026). Framing the Future of Search: A discourse analysis of Google's AI Overviews. *Discourse and Communication*, 20(4), 728–747. https://doi.org/10.1177/17504813261447684

Yeo, M. A., Moorhouse, B. L., & Wan, Y. (2025). From Academic Text to Talk-Show: Deepening Engagement and Understanding with Google NotebookLM. *TESL-EJ*, 28(4). https://doi.org/10.55593/ej.28112int