

JURNAL MERPATI

Media Publikasi Pengabdian Kepada Masyarakat Universitas Logistik dan Bisnis Internasional

<https://ejurnal.ulbi.ac.id/index.php/merpati>

THE ROLE OF NOTEBOOKLM IN ENHANCING ARTIFICIAL INTELLIGENCE LITERACY AMONG VOCATIONAL STUDENTS

Rafli Mochamad Ramadhan¹, Mochammad Fauzi Nur Fadilah², Nanda Septiana Ramadhani³, Richard Firmansyah⁴, Muhammad Bagus Tri Atmaja⁵, Nawal Haromaen⁶, Wa Ode Nur Alia⁷, Syafrial Fachri Pane⁸, Mohamad Nurkamal Fauzan⁹, Muhamad Saepudin¹⁰

¹School of Information Technology, Applied Bachelor Program of Informatics Engineering
¹raflimramadhan.204@gmail.com

² School of Information Technology, Applied Bachelor Program of Informatics Engineering
²nurfadilahmfauzi@gmail.com

³ School of Information Technology, Applied Bachelor Program of Informatics Engineering
³nandasr.24@gmail.com

⁴ School of Information Technology, Applied Bachelor Program of Informatics Engineering
⁴richardfirmansyah57@gmail.com

⁵ School of Information Technology, Applied Bachelor Program of Informatics Engineering
⁵mbagus0111@gmail.com

⁶ School of Information Technology, Applied Bachelor Program of Informatics Engineering
⁶nawalharomaen@gmail.com

⁷ School of Information Technology, Applied Bachelor Program of Informatics Engineering
⁷nuraliaaa2109@gmail.com

⁸ School of Information Technology, Applied Bachelor Program of Informatics Engineering
⁸syafrial.fachri@ulbi.ac.id

⁹ School of Information Technology, Applied Bachelor Program of Informatics Engineering
⁹m.nurkamal.f@ulbi.ac.id

¹⁰Software and Game Development, SMK Bakti Nusantara 666
¹⁰msaepudin04@gmail.com

ABSTRACT

The Society 5.0 era requires vocational students to use artificial intelligence productively and ethically, yet many still rely on unverified, open-domain generative AI. This community service program introduced Google NotebookLM, a Retrieval-Augmented Generation-based learning assistant, to Software Engineering students at SMK Bakti Nusantara 666. Through a Project-Based Learning approach, a single-day workshop combined concept delivery, feature demonstration, and hands-on practice building a personal knowledge base from participants' own materials. A pre-test on general AI literacy was given to 26 participants, and a post-test on NotebookLM-specific knowledge to 17. Baseline literacy was already strong (mean 89.7%), though half the participants had never heard of NotebookLM. After the workshop, 16 of 17 post-test respondents (94.1%) scored at least 75, exceeding the target, with strongest mastery in source-grounded answering, citation, and ethical use. The only discriminating item showed participants over-extending the Study Guide feature to a summarisation task meant for Summary, indicating overgeneralisation rather than lack of exposure. Because the two instruments measured related but distinct constructs, a normalised gain could not be validly computed, and results are reported descriptively. A hands-on, document-grounded workshop can strengthen vocational students' operational AI literacy, provided adjacent features are demonstrated contrastively.

Keywords: artificial intelligence literacy, NotebookLM, Retrieval-Augmented Generation, vocational education, project-based learning

1. INTRODUCTION

The Society 5.0 era places artificial intelligence at the centre of the way people learn, work, and manage information, and mastery of this technology has consequently become a basic competency rather than a complementary skill (Shahidi Hamedani et al., 2024). In the education sector, artificial intelligence has developed from intelligent tutoring systems into generative models that are able to personalise learning on a large scale, but at the same time it demands institutional readiness in terms of ethics and instructional design (Holmes et al., 2023). Vocational high schools (Sekolah Menengah Kejuruan, SMK) are the level of education that is most directly affected, because their graduates enter the workforce immediately after graduation (Hasbi et al., 2025). Vocational students therefore need not only technical skills, but also the ability to use artificial intelligence productively, independently, and ethically (Ranuharja et al., 2025).

However, exposure to generative artificial intelligence among vocational students has so far occurred largely informally, through personal use for social media or school assignments, rather than through structured academic guidance. This creates a specific urgency for vocational high schools: because their graduates enter the workforce immediately after graduation, any habit of using artificial intelligence uncritically is likely to be carried directly into professional practice rather than corrected during further study. This gap is not primarily a gap in awareness, since general familiarity with generative artificial intelligence is already common among these students, but a gap in guided academic practice: students are rarely taught how to use such tools productively, how to verify AI-generated information against a trusted source, or how to work independently rather than simply copying instant answers. This gap between the demand for productive, independent artificial intelligence use and the actual practice in schools is the main reason for carrying out this community service program.

The partner of this community service program is SMK Bakti Nusantara 666, a private vocational school in Cileunyi District, Bandung Regency, which is located approximately 20 km, or about 40 minutes of travel time, from Universitas Logistik dan Bisnis Internasional (ULBI). The partner was chosen for two reasons. First, the school runs a Software Engineering (RPL) expertise program whose competency profile is closely related to the field of the service team, which allows the transfer of knowledge to be carried out optimally. Second, the distance between the two institutions is close enough to allow mentoring to be continued after the activity ends. Based on preliminary observation and interviews with the teachers of the productive subjects, the students already possess basic programming skills, the school owns computer laboratories, and the internet connection is stable; however, the school has never held any structured training on the academic use of artificial intelligence.

The situation analysis produced six priority problems of the partner, namely (1) general familiarity with generative artificial intelligence that has not been converted into productive and ethical academic use; (2) NotebookLM has never been introduced as a personal learning assistant; (3) difficulty in understanding and synthesising dense vocational learning materials; (4) weak self-regulated learning outside school hours; (5) the absence of artificial-intelligence-based portfolios that can support industrial internships and further study; and (6) the tendency to use technology for entertainment rather than for academic productivity. These problems are urgent because students who are already familiar with generative artificial intelligence tend to use it instantly to copy answers without curating information or constructing knowledge (Hasbi et al., 2025), while empirical studies show that learners rarely use it to plan, monitor, and evaluate their own learning (Spirgi & Seufert, 2025). If this condition is left unaddressed, the risk of unintentional plagiarism and the decline of critical thinking will increase, because the information that is obtained cannot be traced back to its original source.

To address these gaps, the service team designed a training program centred on Google NotebookLM, a platform that works only on documents uploaded by the user through Retrieval-Augmented Generation, meaning that its answers are grounded in a trusted knowledge base rather than the entire internet (Lewis et al., 2020). Reviews of the application of this technology in education show that such grounding improves factual accuracy and suppresses hallucination, which is the main weakness of large language models in learning contexts (Li et al., 2025; Swacha & Gracel, 2025). Every answer is equipped with in-line citations that point to the source passage, which allows students to verify

information and maintain academic integrity (Google, 2025; Reyna, 2025), and its use as a document-based tutor has been shown to keep the discussion consistent with the teaching material provided by the teacher (Tufino, 2025), while early empirical evaluation of NotebookLM indicates that the quality of the uploaded sources, rather than the platform itself, governs the learning value that students perceive (Çallı & Alma Çallı, 2026). Features such as instant summarisation, study guides, mind maps, and Audio Overview can be accessed free of charge through a browser, making this solution realistic for the infrastructure and budget of the partner school.

To ensure that the training goes beyond simply introducing an application, the activity was designed around a Project-Based Learning approach, in which the participants work on a real project in the form of a personal knowledge base that is built from their own vocational learning materials (Sánchez-García & Reyes-de-Cózar, 2025). This scheme is directed at strengthening self-regulated learning, namely the habit of planning, monitoring, and evaluating one's own learning process; recent experimental evidence shows that learners who receive explicit metacognitive scaffolding in generative artificial intelligence environments improve their planning and self-evaluation, whereas learners working without such scaffolding decline on the same dimensions (Xu et al., 2025). Accordingly, the plan for solving the partner's problems is arranged in stages, namely the measurement of the initial condition, the delivery of the concepts of artificial intelligence and academic ethics, the guided practice of the NotebookLM features, the development of a group mini project, the presentation of the results, and the measurement of the final condition.

Taken together, these conditions define three aims for this community service program, each of which corresponds directly to the priority problems identified in the situation analysis. The first aim, to improve the productive and ethical artificial intelligence literacy of the students of SMK Bakti Nusantara 666, addresses Problems (1) and (6), namely general AI familiarity that has not been converted into productive academic use and the tendency to use technology mainly for entertainment. The second aim, to train the skill of using NotebookLM to summarise, synthesise, and verify vocational learning materials, addresses Problems (2) and (3), namely the school's lack of exposure to NotebookLM and students' difficulty synthesising dense vocational materials. The third aim, to produce a NotebookLM-based mini project that can serve as an initial academic portfolio for the participants, addresses Problem (5), the absence of artificial-intelligence-based portfolios to support internships and further study. Problem (4), weak self-regulated learning outside school hours, is addressed indirectly through the Project-Based Learning design described in the following paragraph, rather than through a separate standalone aim.

2. IMPLEMENTATION METHOD

This community service activity was designed with a quantitative approach using a one-group pretest-posttest pre-experimental design, in which a knowledge test functioned as the main data collection instrument and a participant satisfaction questionnaire functioned as the supporting instrument (Creswell & Creswell, 2023). The quantitative approach was selected because the achievement of the activity had to be measured numerically and objectively, allowing the outcome of the training to be reported in the form of scores and percentages (Sugiyono, 2022). The implementation of the activity is described through five components, namely the place and time of the activity, the target audience, the community service method, the success indicators, and the evaluation method.

2.1 Place and Time

The activity was carried out at SMK Bakti Nusantara 666, which is located at Jalan Percobaan No. 65, Cileunyi Wetan, Cileunyi District, Bandung Regency, West Java Province. The school is approximately 20 km, or about 40 minutes of travel time, from Universitas Logistik dan Bisnis Internasional (ULBI), which allowed coordination between the service team and the partner school to be conducted intensively before and during the activity. The activity was held on Tuesday, 11 August 2026, in a single session that started at 10.00 WIB (Western Indonesian Time) and took place in the computer laboratory of the partner school. The laboratory was chosen because it provides computer units, a projector, and an internet connection that is adequate for accessing NotebookLM directly through a web browser without any additional installation.

2.2 Target Audience

The target audience of this activity was the male and female students of SMK Bakti Nusantara 666, specifically students of the Software Engineering (RPL) expertise program, with a target of 30 participants in one class. The participants were determined purposively with three criteria, namely being actively enrolled at the partner school, possessing basic computer and internet operation skills, and owning a Google account that can be used to access NotebookLM. Demographically, the participants were adolescents between 15 and 18 years of age who come from the area of Cileunyi District, Bandung Regency, and who are being prepared either to enter the world of work or to continue to higher education after graduation. This group was selected because vocational students are the group that most needs the ability to process learning resources independently, yet they have not received structured guidance on the productive use of artificial intelligence.

2.3 Community Service Method

The community service method applied was training or a workshop that combined guided tutorials, hands-on practice, and group mini projects within a Project-Based Learning scheme (Sánchez-García & Reyes-de-Cózar, 2025). The workshop was implemented in five stages. First, preparation and opening, which included registration, an explanation of the objectives of the activity, and the administration of the pre-test. Second, the delivery of material on the basic concepts of artificial intelligence, the working principle of Retrieval-Augmented Generation, and the ethics of the academic use of artificial intelligence. Third, a guided demonstration of the main features of NotebookLM, covering the uploading of sources, instant summarisation, question and answer with in-line citations, mind maps, study guides, and Audio Overview. Fourth, hands-on practice in small groups, in which every group built a personal knowledge base from their own vocational learning materials. Fifth, the presentation of the mini projects, the administration of the post-test and the questionnaire, and the closing of the activity. Throughout the practice sessions, the service team acted as facilitators, accompanying the participants group by group to resolve technical obstacles as they arose.

2.4 Success Indicators

The main indicator of success of this activity is the ability of the participants to master and to implement the NotebookLM material that has been delivered. This indicator was operationalised into four measurable criteria. First, at least 75% of the participants obtain a post-test score of at least 75 on a scale of 100. Second, there is an increase in the average score from the pre-test to the post-test with a normalised gain that falls at least in the medium category. Third, at least 80% of the participant groups are able to produce a mini project in the form of a NotebookLM notebook that contains at least three of their own learning sources, together with a summary, a mind map, and a question-and-answer session that is complete with citations. Fourth, the average participant response to the implementation of the activity reaches a minimum of 4.00 on a five-point Likert scale. The knowledge and skill criteria in these indicators refer to the dimensions of artificial intelligence literacy, which cover conceptual understanding, practical application, and the ethics of use (Ng et al., 2024; Yang et al., 2025; Zhang et al., 2025).

2.5 Evaluation Method

Evaluation was based on a knowledge test administered twice, namely a pre-test before the material was delivered and a post-test after the mini project presentation. The pre-test consisted of nine scored multiple-choice items on general artificial intelligence concepts, ethics, and document-based AI, together with one non-scored item on prior familiarity with NotebookLM, while the post-test consisted of ten scored multiple-choice items on NotebookLM-specific operational knowledge. The original design intended an identical item grid, which would have made the two scores directly comparable; in the field, however, the pre-test and post-test that were actually administered used non-identical item sets covering related but distinct constructs, a deviation that is treated as a limitation and addressed further in Section 3.4. Item wording was reviewed by the facilitators for face validity before use, although a formal reliability statistic, such as Cronbach's alpha, was not computed for this cycle. Where a matched pre-post sample was available, the improvement was additionally examined with the normalised gain formula, that is, the difference between the post-test and the pre-test score divided by the difference between the maximum score and the pre-test score, with the categories of high for a value

of at least 0.70, medium for a value between 0.30 and 0.69, and low for a value below 0.30 (Hake, 1998), and applicable only where both administrations use the same item grid; because that condition was not met in the field, no normalised gain and no inferential comparison between the two administrations are reported, and each instrument is analysed separately (Section 3.4). Qualitative evaluation of the mini project outputs and the participant satisfaction questionnaire was also planned, following the reaction and learning levels of the training evaluation model (Kirkpatrick & Kirkpatrick, 2016). The mini project stage was carried out as planned (Figure 2), but was not scored against a documented rubric during the activity, and the satisfaction questionnaire results had not been compiled at the time of writing; both are therefore reported in Section 3.4 as pending rather than completed indicators.

3. RESULTS AND DISCUSSION

3.1 Baseline AI and NotebookLM Literacy (Pre-Test Results)

Before the material was delivered, 26 of the 30 targeted students of the Software Engineering (RPL) class completed the pre-test. The instrument consisted of nine scored multiple-choice items covering general concepts of artificial intelligence, the working principle of document-based AI, the ethics of AI use in an academic setting, and the rationale for verifying AI-generated information, together with one additional non-scored item asking whether the participant had previously heard of or used Google NotebookLM. Across the nine scored items, participants achieved a mean score of 89.7% (SD = 19.2; range 22.2–100.0%), and 22 of the 26 participants (84.6%) scored at least 75. Item-level results show that participants already answered most conceptual items correctly, ranging from 80.8% on the item about document-based AI to 96.2% on the item about the general definition of AI, which indicates that general AI literacy was not the partner's primary gap, a pattern consistent with AI-literacy frameworks that treat general conceptual understanding and tool-specific operational skill as separate, only loosely correlated dimensions (Ng et al., 2024; Zhang et al., 2025).

A different picture emerged from the non-scored familiarity item. Thirteen of the 26 respondents (50.0%) reported that they had never heard of or used Google NotebookLM, eight (30.8%) had heard the name but never tried it, four (15.4%) reported using it regularly, and one respondent claimed to already be proficient with its features. This distribution confirms the second priority problem identified in the situation analysis, namely that NotebookLM had never been introduced at SMK Bakti Nusantara 666. This gap mirrors the broader pattern reported for Indonesian vocational education, where access to generative AI tools has outpaced structured guidance on how to use them for academic work (Hasbi et al., 2025), and it justifies the workshop's emphasis on guided, hands-on practice rather than on conceptual explanation alone.

3.2 Implementation of the NotebookLM Workshop

The workshop was carried out in a single session on Tuesday, 11 August 2026, in the computer laboratory of SMK Bakti Nusantara 666, following the five stages described in the implementation method (Section 2.3): registration and the administration of the pre-test; delivery of material on the basic concepts of artificial intelligence, the working principle of Retrieval-Augmented Generation, and the ethics of academic AI use; a guided demonstration of the main NotebookLM features, comprising source upload, instant summarisation, citation-based question and answer, mind maps, study guides, and Audio Overview; hands-on group practice, in which each group built a personal knowledge base from their own vocational learning materials; and a closing stage consisting of mini project presentations, the administration of the post-test, and the participant satisfaction questionnaire. The service team acted as facilitators who moved between groups, resolving technical obstacles such as document formatting and account access as they occurred.



Figure 1. Service team and student participants of the NotebookLM workshop, SMK Bakti Nusantara 666

Figure 1 shows the service team together with the students of the Software Engineering (RPL) class at the opening of the workshop in the computer laboratory of SMK Bakti Nusantara 666, held on 11 August 2026. The photo illustrates the scale of participation and the collaborative atmosphere in which the training was conducted.



Figure 2. Hands-on practice session in the computer laboratory

Figure 2 captures two participants working together at a workstation during the hands-on practice stage, applying the NotebookLM features that had just been demonstrated to a set of their own vocational learning materials. This stage produced the personal knowledge bases that were later presented as each group's mini project.

Of the 26 students who completed the pre-test, 17 completed the post-test, a completion rate of 65.4% relative to the pre-test cohort; of these, 16 could be matched to an individual pre-test record, so the matched set is smaller than either administration. This attrition is most likely attributable to the time constraints of a single-day session, in which the later stages of group practice, presentation, and closing left less time for some participants to complete the post-test form before the session ended. A second and more consequential factor is that the post-test form remained open rather than being administered under supervision: the recorded submissions span seventeen hours and thirty-three minutes, with twelve responses arriving within the hour after the session, four later the same day, and one on the following morning. Both conditions are taken into account when interpreting the results reported in Sections 3.3 and 3.4.

3.3 Post-Test Results and NotebookLM Competency Achievement

The post-test consisted of ten scored multiple-choice items covering the definition of NotebookLM, its Retrieval-Augmented Generation mechanism, the procedure for uploading source documents, the Summary and Audio Overview features, the Study Guide feature for generating practice questions, the function of in-line citation, the ethics of using NotebookLM, the perceived benefit of the platform for learning effectiveness, and the function of the Chat/Q&A feature. Of the 17 participants who completed the post-test, the mean score was 88.8% (SD = 9.6; range 60.0–100.0%), and 16 of the 17 participants (94.1%) reached a score of at least 75. This result exceeds the first success indicator set out in Section 2.4, namely that at least 75% of participants obtain a post-test score of at least 75, a level consistent with the accuracy gains reported for document-grounded, RAG-based learning assistants relative to open-domain generative AI (Li et al., 2025; Swacha & Gracel, 2025). A side-by-side summary of the two instruments is presented in Table 1.

Table 1. Comparison between the pre-test and post-test results

Metric	Pre-Test (Baseline)	Post-Test (NotebookLM Competency)
Instrument focus	General AI concepts, ethics, and document-based AI (9 scored items)	NotebookLM-specific operational knowledge (10 scored items)
N (respondents)	26	17
Mean score	89.7%	88.8%
SD	19.2	9.6
Score range (min–max)	22.2–100.0%	60.0–100.0%
Participants scoring ≥ 75	22/26 (84.6%)	16/17 (94.1%)

An item-level analysis, presented in Figure 3, shows that participants performed at ceiling (100% correct) on the items concerning the definition of NotebookLM, its document-grounded RAG mechanism, the procedure for uploading sources, and the function of Audio Overview, and performed strongly (94.1%) on the items concerning citation, ethical use, the platform’s learning benefit, and the Chat/Q&A function. A further item, which asked which feature generates practice questions from a source, was answered correctly by 14 of the 17 participants (82.4%), confirming that the Study Guide feature had been demonstrated and was recognised. Eight of the ten items were therefore answered correctly by at least 94% of participants and the observed scores collapsed onto only four distinct values, so the instrument discriminated weakly at the upper end and its mean is better read as evidence of successful exposure than as a fine-grained measure of proficiency. These results indicate that participants internalised the pedagogically central points emphasised throughout the training, namely that NotebookLM answers are grounded in the participant’s own documents, that in-line citation exists to prevent hallucination and to support verification, and that the platform is meant to support rather than replace independent understanding.

The one discriminating item asked which feature is most effective for turning a long document into a set of summary points. Only 5 of the 17 participants (29.4%) answered “Summary”; seven (41.2%) selected “Study Guide” and five (29.4%) selected “Audio Overview.” Read together with the 82.4% accuracy on the Study Guide item, this pattern is not a failure of exposure: the same cohort could identify what Study Guide is for, yet extended it to an adjacent task for which Summary is the intended tool. The error is therefore better characterised as over-generalisation of a demonstrated feature than as unfamiliarity with it, which is consistent with the fact that the same respondents answered the surrounding items on citation, ethics, and source verification correctly. It must also be acknowledged that the item itself is imperfect: because Study Guide does produce a condensed overview of a source, it is partially defensible as an answer, and overlapping, insufficiently distinguished response options are a recognised source of construct-irrelevant error in multiple-choice assessment (Haladyna et al., 2002). The accuracy reported for this item should therefore be read with that caveat. The finding nevertheless yields a concrete, actionable input for the next training cycle: the boundary between Summary and Study Guide should be demonstrated side by side on the same source document, and the evaluation item should be rewritten to leave exactly one option defensible.

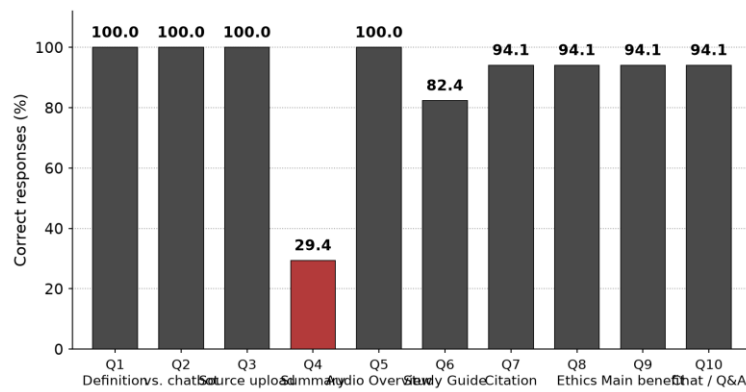


Figure 3. Item-level accuracy of the post-test (N = 17)

3.4 Achievement of Success Indicators

Measured against the four success indicators defined in Section 2.4, the first indicator, that at least 75% of participants obtain a post-test score of at least 75, was achieved and exceeded, with 16 of the 17 post-test respondents (94.1%) reaching that threshold. The second indicator, an increase from the pre-test to the post-test with a normalised gain of at least the medium category, could not be validly computed from the two instruments actually administered. The method planned an identical item grid for both tests (Section 2.5), but the pre-test that was administered assessed general AI concepts and ethics (9 scored items, mean 89.7%), while the post-test assessed NotebookLM-specific operational knowledge (10 scored items, mean 88.8%); the two instruments measure related but distinct constructs rather than an identical item set. For this reason neither a matched gain score nor an inferential test of the difference between the two administrations is reported here: a change score computed across instruments that measure different constructs cannot be interpreted as learning, and the high, narrowly distributed pre-test scores would in any case make such a score unreliable (Cronbach & Furby, 1970). The two administrations are consequently reported side by side as independent descriptive results rather than as a before-and-after comparison. For the next cycle, the service team recommends administering a single, identical item grid before and after the intervention, which would allow a valid normalised gain to be reported (Hake, 1998); where a single grid is not feasible within one session, a retrospective pre-post item format administered at the same sitting is a practical alternative, provided that the documented effects of item format on that design are taken into account (Hwalek et al., 2024).

This recommendation concerns only a future cycle, however, and does not affect the results reported here: because the pre-test and post-test analysed in this article were objective knowledge-test items rather than a retrospective self-report instrument, the self-assessment subjectivity associated with retrospective pre-post designs, in which participants rate their own change after training has ended, does not apply to the present evaluation. As a further, action-based alternative to any self-report measure, the mini-project stage already provides a direct test of applied skill rather than of self-perceived change: each group must build a working knowledge base from its own uploaded sources and produce a source-grounded summary, mind map, and cited question-and-answer session (Figure 2), so that competency is demonstrated through an observable artefact rather than reported by the participant. To make this direct evidence usable for evaluation, the project rubric recommended below is scored from the artefact itself, for example by awarding points for the number and diversity of sources uploaded, the traceability of the summary and question-and-answer session to a cited source, and the correct rather than over-extended use of the Study Guide, Summary, and Chat features, so that the outcome reflects demonstrated action rather than subjective self-judgement.

The remaining two indicators correspond to the product and reaction levels of standard training-evaluation frameworks, which are conceptually distinct from the learning-level gains reported above (Kirkpatrick & Kirkpatrick, 2016). The third indicator, that at least 80% of groups produce a NotebookLM mini project containing at least three of their own learning sources with a summary, a mind map, and a cited question-and-answer session, cannot be reported quantitatively in this article: the hands-on group practice and mini-project stage was carried out as planned, with each group building

and presenting its own NotebookLM knowledge base (Figure 2), but the outputs were not scored against a documented rubric or otherwise recorded at the time of the activity, so no dataset is available for this indicator. The fourth indicator, that the average participant response on the satisfaction questionnaire reaches at least 4.00 on a five-point Likert scale, is likewise not covered by the pre-test and post-test response data analysed in this section, since both files record only knowledge-test answers, and the questionnaire results had not been compiled at the time of writing. Both indicators are reported here as a limitation rather than as an achieved or failed outcome, and are recommended for systematic documentation, using a simple project rubric and a compiled satisfaction questionnaire, respectively, in the next training cycle.

Beyond the indicators themselves, four limitations qualify the results reported in this section. First, the design is a single-group pre-experimental one without a control group, so the observed competency level cannot be attributed to the intervention with certainty. Second, the post-test was completed outside a supervised window, so the possibility that some responses were produced with the learning material, or with the platform itself, open cannot be excluded; this is the most serious threat to the validity of the reported accuracy figures and is straightforward to remove in the next cycle by administering the test in the laboratory before participants leave. Third, one respondent was enrolled in a different year group from the target class, so the post-test cohort is not perfectly homogeneous with the population described in Section 2.2. Fourth, no reliability coefficient was computed for either instrument and one post-test item admits more than one defensible answer, so the item-level percentages should be read as indicative rather than as psychometrically established. These limitations do not affect the descriptive finding that the first success indicator was met, but they do constrain any causal or generalising interpretation of the results.

4. CONCLUSION

This community service program successfully introduced Google NotebookLM as a document-grounded, Retrieval-Augmented Generation-based learning assistant to students of the Software Engineering (RPL) expertise program at SMK Bakti Nusantara 666. The pre-test results indicated that participants already possessed reasonably strong general AI literacy (mean 89.7%), yet the activity confirmed the partner school's core problem, namely that half of the respondents had never heard of or used NotebookLM prior to the workshop. Following the five-stage, Project-Based Learning workshop, 16 of the 17 participants who completed the post-test (94.1%) scored at least 75 on NotebookLM-specific operational knowledge, exceeding the first success indicator and demonstrating that participants were able to apply core features such as source upload, citation-based question and answer, and Audio Overview. Item-level analysis further showed that participants strongly internalised the platform's central pedagogical value, namely document-grounded answers, in-line citation for verification, and the ethical use of artificial intelligence as a support for rather than a replacement of independent learning. The single area requiring refinement was the boundary between the Summary and the Study Guide features: participants correctly identified the primary function of Study Guide but over-extended it to a summarisation task, which points to a need for contrastive rather than sequential demonstration. Because the pre-test and post-test instruments assessed related but distinct constructs rather than an identical item grid, the normalised-gain indicator could not be validly computed in this cycle, and the two administrations are reported as independent descriptive results rather than as a before-and-after comparison. The group mini-project stage was carried out as planned but was not scored against a documented rubric, and the satisfaction-questionnaire results had not been compiled at the time of writing, so both remain to be reported in a future cycle; the post-test was also completed outside a supervised window, which constrains the strength of the reported accuracy figures. Overall, the training met its main objective of improving productive and ethical artificial intelligence literacy among vocational students and produced a reusable training module for the partner school, with a recommendation to administer an identical pre-test and post-test item grid under supervised conditions, to document the mini project with a simple rubric, and to demonstrate the Summary–Study Guide distinction contrastively in future cycles.

5. ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to SMK Bakti Nusantara 666, particularly the school principal, the teachers of the Software Engineering (RPL) expertise program, and all student participants, for their cooperation, enthusiasm, and support throughout the implementation of this community service program. The authors also thank Universitas Logistik dan Bisnis Internasional (ULBI), especially the School of Vocation and the Applied Bachelor Program of Informatics Engineering, for facilitating and funding this activity. Appreciation is also extended to all members of the service team who contributed to the preparation, facilitation, and evaluation of the workshop.

6. REFERENCES

- Çallı, L., & Alma Çallı, B. (2026). Sustainable adoption of AI-generated instructional videos: An empirical evaluation of the LBUC model via NotebookLM. *Systems*, 14(6), 631. <https://doi.org/10.3390/systems14060631>
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
- Cronbach, L. J., & Furby, L. (1970). How we should measure “change”—or should we? *Psychological Bulletin*, 74(1), 68–80. <https://doi.org/10.1037/h0029382>
- Google. (2025). *NotebookLM: Your personalized AI-powered collaborator*. <https://notebooklm.google/>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5
- Hasbi, M., Pramuditya, S. A., & Fauzi, I. (2025). Pemanfaatan kecerdasan buatan dalam pembelajaran di era Society 5.0. *Jurnal Pendidikan Teknologi dan Kejuruan*, 30(2), 112–124. <https://doi.org/10.21831/jptk.v30i2.70021>
- Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial intelligence in education. In *Data ethics: Building trust* (pp. 621–653). Globethics. <https://doi.org/10.58863/20.500.12424/4276068>
- Hwalek, M., Pierce, K., & Straub, V. (2024). Designing a questionnaire with retrospective pre-post items: Format matters. *Evaluation and Program Planning*, 103, 102411. <https://doi.org/10.1016/j.evalprogplan.2024.102411>
- Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. ATD Press.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
- Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., & Wang, F. L. (2025). Retrieval-augmented generation for educational application: A systematic survey. *Computers and Education: Artificial Intelligence*, 8, 100417. <https://doi.org/10.1016/j.caeai.2025.100417>
- Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, 55(3), 1082–1104. <https://doi.org/10.1111/bjet.13411>
- Ranuharja, F., Ganefri, Rizal, F., & Langeveldt, D. (2025). Relevance and impact of generative AI in vocational instructional material design: A systematic literature review. *Salud, Ciencia y Tecnología*, 5. <https://sct.ageditor.ar/index.php/sct/article/view/1336>
- Reyna, J. (2025). The potential of Google NotebookLM for teaching and learning. In *Proceedings of E-Learn 2025: World Conference on E-Learning*. Association for the Advancement of Computing in Education (AACE).
- Sánchez-García, R., & Reyes-de-Cózar, S. (2025). Enhancing project-based learning: A framework for optimizing structural design and implementation—A systematic review with a sustainable focus. *Sustainability*, 17(11), 4978. <https://doi.org/10.3390/su17114978>
- Shahidi Hamedani, S., Aslam, S., Mundher Oraibi, B. A., Wah, Y. B., & Shahidi Hamedani, S. (2024). Transitioning towards tomorrow's workforce: Education 5.0 in the landscape of Society 5.0: A

- systematic literature review. *Education Sciences*, 14(10), 1041. <https://doi.org/10.3390/educsci14101041>
- Spirgi, L., & Seufert, S. (2025). GenAI as a learning assistant, an empirical study in higher education. In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025)*. SCITEPRESS. <https://www.scitepress.org/Papers/2025/131993/131993.pdf>
- Sugiyono. (2022). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.
- Swacha, J., & Gracel, M. (2025). Retrieval-augmented generation (RAG) chatbots for education: A survey of applications. *Applied Sciences*, 15(8), 4234. <https://doi.org/10.3390/app15084234>
- Tufino, E. (2025). *NotebookLM: An LLM with RAG for active learning and collaborative tutoring*. arXiv. <https://arxiv.org/abs/2504.09720>
- Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, 56(5), 1842–1863. <https://doi.org/10.1111/bjet.13599>
- Yang, Y., Zhang, Y., Sun, D., He, W., & Wei, Y. (2025). Navigating the landscape of AI literacy education: Insights from a decade of research (2014–2024). *Humanities and Social Sciences Communications*, 12, 374. <https://doi.org/10.1057/s41599-025-04583-8>
- Zhang, H., Perry, A., & Lee, I. (2025). Developing and validating the Artificial Intelligence Literacy Concept Inventory: An instrument to assess artificial intelligence literacy among middle school students. *International Journal of Artificial Intelligence in Education*, 35, 398–438. <https://doi.org/10.1007/s40593-024-00398-x>