

Deep Learning and Deeper Learning in Education: A Critical Narrative Review of Evidence and Implications for Indonesia in the Society 5.0 Era

Azizah Dwi Safitri*, Verra Ayu Fatmawati, Iqriana Nafi'a Salsabilah, Rizky Dwi Rachmawati, A.F. Suryaning Ati MZ

Primary School Teacher Education Department, Universitas Muhammadiyah Lamongan, Indonesia

**Corresponding author: dwiazizah825@gmail.com*

Received: 27 February 2025; Accepted: 03 October 2025; Published: 10 October 2025

Abstract: Indonesian education policy links three ideas that are often run together: Society 5.0, neural-network deep learning (DL) and pembelajaran mendalam (deeper learning), a pedagogical approach. This critical narrative review asks what evidence supports AI-enabled learning applications, for whom and under what conditions in Indonesia. Secondary reviews, randomised and field trials, international guidance, Indonesian policy documents and empirical studies were purposively selected through targeted searches in October 2025 and coded thematically. The technical and pedagogical senses of “deep learning” are conceptually independent, and Society 5.0 is a normative Japanese policy vision, not an evidenced stage of development. Evidence is strongest for intelligent tutoring systems and emerging for carefully designed generative AI tutors, whereas unguided chatbot access reduced later performance in one field experiment. Indonesian evidence is thin and mostly self-reported. AI is best positioned as a scaffold for deeper learning, conditional on teacher capacity, equitable access and local evaluation.

Keywords: deep learning; deeper learning; Society 5.0; artificial intelligence in education; Indonesia

INTRODUCTION

Indonesia's Ministry of Primary and Secondary Education announced that a deep learning approach, pembelajaran mendalam, would begin in the 2025/2026 school year without yet being compulsory, and described it as an approach to teaching rather than a new curriculum ([Pusat Penelitian Badan Keahlian DPR RI, 2024](#); [Republika, 2025](#)). By July 2025 the Ministry was presenting coding and AI as part of the same package and training teachers of all subjects. In the same period, national AI policy named education as a priority sector ([OECD.AI, n.d.-a, n.d.-b](#)), and Society 5.0 is routinely invoked as the reason schools must change (for example, [Subandowo, 2022](#)). Three different ideas therefore circulate under overlapping labels. “Deep learning” is at once a family of multi-layer neural-network methods in computer science ([LeCun et al., 2015](#)), a pedagogical aim of deeper understanding as opposed to surface memorisation ([Marton & Säljö, 1976](#); [Fullan & Langworthy, 2014](#)), and the name of a current Indonesian policy. Society 5.0 is a Japanese policy vision that is sometimes treated as a historical stage Indonesia must enter.

AI in education has been reviewed extensively. [Zawacki-Richter et al. \(2019\)](#) screened 2,656 publications and synthesised 146 studies on AI in higher education, and reported little critical reflection on risks and weak links to pedagogical theory. [Chen et al. \(2020\)](#), [Crompton and Burke \(2023\)](#) and [Bond et al. \(2024\)](#) later mapped the field; Bond et al. reviewed 66 earlier reviews, found that adaptive systems, personalisation, and profiling and prediction dominate, and called for more attention to ethics, rigour and underrepresented contexts, including developing countries and non-STEM disciplines. Two things are still missing. No review we located separates the

technical and pedagogical senses of deep learning and asks how the first can serve the second. And the existing reviews focus on higher education in high-income settings, so they say little about what transfers to a school system like Indonesia's, where policy ambition is moving faster than classroom evidence.

This article addresses that gap with a critical narrative review. It is not a systematic review and does not claim to be exhaustive; the Method explains the selection of sources and its limits. Four research questions guide it.

RQ1. How do the technical and the pedagogical meanings of “deep learning” differ, and how are they connected in Indonesian policy?

RQ2. What does Society 5.0 mean in its original source, and how is it used in Indonesian education discourse?

RQ3. What evidence exists on the effects of AI-enabled applications on learning, and under what conditions do effects appear or fail?

RQ4. What risks and contextual constraints, including equity, bias, privacy and teacher readiness, bear on adoption in Indonesia?

The contribution is threefold: a conceptual separation that policy and research can reuse, an evidence-graded synthesis of AI applications that links them to the three elements of pembelajaran mendalam, and a set of implications and research questions specific to Indonesia.

METHOD

Review Design

We conducted a critical narrative review. This design suits the questions because they span concepts, policy and empirical evidence, and the relevant sources are too heterogeneous for a single pooled or PRISMA-style analysis. We therefore do not report a PRISMA flow diagram, record counts or a formal risk-of-bias assessment, since these would imply a systematic search that was not run. No bibliometric or network-mapping analysis was performed.

Source Identification

Sources were identified in October 2025 through targeted queries in English and Indonesian in a general web search engine, followed by checks of publisher, university-repository, government and international-organisation pages to confirm bibliographic details. Query terms combined “deep learning”, “deeper learning”, “pembelajaran mendalam”, “Society 5.0”, “intelligent tutoring”, “generative AI”, “algorithmic bias”, “teacher readiness” and “Indonesia” with education terms. Licensed databases (Scopus, Web of Science, ERIC) were not searched. Reference lists of the retrieved reviews were used to find foundational works.

Eligibility and Selection

A source was included if it was (a) a peer-reviewed review, meta-analysis, randomised trial or field experiment on AI applications in education, published 2011–2025; (b) a foundational source that defines a concept used in the review, of any date; (c) an official policy document, or a reliable record of an official announcement, or an international guidance document; or (d) an Indonesian empirical study with a stated sample and method. We excluded sources without identifiable authors or venue, commercial blog posts used as evidence of effects, and sources whose content we could not verify against the text, abstract or an official record. News reports were used only to document what Indonesian authorities announced, not to support claims about effects. Sources were chosen purposively to cover each research question, and the final set is listed in the References.

Data Extraction and Synthesis

For each evidence source we recorded type, context and level of education, design and sample, outcome, main finding and stated limitations (Table 2). We synthesised thematically using four codes that follow the research questions: concept, application, evidence and risk or context. We graded the strength of evidence for each application with three simple rules. Higher means

meta-analytic or multiple controlled studies. Moderate means a single randomised or field experiment, or a systematic review of heterogeneous studies. Lower means a position paper, a non-probability survey, a preprint or a policy statement. This is an explicit heuristic, not a GRADE assessment.

Scope and Terminology

To avoid conflating the two meanings, we use deep learning (DL) only for multi-layer neural-network methods, and deeper learning or *pembelajaran mendalam* for the pedagogical approach. The review asks how AI, including DL-based systems and generative models, can support deeper learning, and where it cannot.

Research ethics and use of AI. This review analysed published literature and publicly available policy documents. It involved no human participants and no personal data, so ethical clearance and informed consent were not required.

RESULTS & DISCUSSION

The review finds that “deep learning” names two unrelated things, that Society 5.0 is a policy vision rather than an evidenced stage, and that the strongest evidence for AI in education concerns carefully designed tutoring, not general exposure to AI. The sections below follow the four research questions.

Two Meanings of Deep Learning (RQ1)

In computer science, deep learning is a family of methods that learn several levels of representation from data using multi-layer neural networks (LeCun et al., 2015). Deng and Yu (2014) sort deep networks into three classes: unsupervised or generative, supervised, and hybrid. Accounts that list only two types omit supervised learning, which is the class behind most classification tasks. Artificial neural networks are loosely inspired by the brain but are not a model of how the brain learns, so the statement that they are built “in the same way” as the brain should not be used. A common illustration, chess, also needs correction. The Deep Blue system that beat a world champion relied on specialised hardware, search and a hand-built evaluation function (Campbell et al., 2002); deep neural networks entered game playing later, with systems such as AlphaZero that learn by self-play (Silver et al., 2018). Neither example concerns education.

In education, deep learning has a much older meaning. Marton and Säljö (1976) distinguished a deep approach, aimed at the meaning of a text, from a surface approach, aimed at reproducing it. Fullan and Langworthy (2014) later used “deep learning” for new pedagogies that build competencies such as character, citizenship, critical thinking, communication, collaboration and creativity. Indonesia’s *pembelajaran mendalam* belongs to this second family: officials describe it through three elements, mindful, meaningful and joyful learning (Pusat Penelitian Badan Keahlian DPR RI, 2024; Republika, 2025). Table 1 sets the two meanings side by side.

Table 1. Two meanings of “deep learning” and their use in this review

Feature	Deep learning (DL), technical	Deeper learning, pedagogical
Field	Machine learning, computer science	Educational psychology, curriculum
What “learning” means	A model adjusts its parameters from data	A person builds understanding and transferable competence
Core sources	LeCun et al. (2015); Deng & Yu (2014)	Marton & Säljö (1976); Fullan & Langworthy (2014)
Indonesian term	Often <i>pembelajaran mendalam</i> , or left as deep learning	<i>Pembelajaran mendalam</i> (policy approach)
Evidence of effect on students	Only through applications built on it	Depends on how the approach is taught and assessed
Role in this review	A technology that may support teaching	The educational aim that technology should serve

The two meanings are linked only by an instrumental question: can DL-based systems help students learn more deeply? Nothing in the technique implies that they do. The three policy

elements also need a theoretical anchor that the sources we found on them, largely early-childhood and thematic-learning articles (for example, [Diputera et al., 2024](#)), do not supply. *Meaningful* learning corresponds to [Ausubel's \(1968\)](#) account of relating new material to what the learner already knows. *Mindful* learning corresponds to metacognition, the monitoring and regulation of one's own learning ([Flavell, 1979](#)). *Joyful* learning is best read through self-determination theory, in which competence, autonomy and relatedness support intrinsic motivation ([Ryan & Deci, 2000](#)). Anchoring the elements this way makes them testable: each implies learning objectives, activities and assessments, and each can be checked against what an AI tool actually does ([Table 3](#)).

Society 5.0 as a Policy Vision (RQ2)

Society 5.0 originated in Japan's 5th Science and Technology Basic Plan, adopted in January 2016, and not in 2019. The plan describes it as a "super-smart society" that follows the hunter-gatherer, agricultural, industrial and information societies, built by merging cyberspace and physical space for the benefit of people (Cabinet Office, 2016, as cited in [Hitachi-UTokyo Laboratory, 2020](#); [Fukuyama, 2018](#)). Three misstatements should be avoided in Indonesian writing on the topic: dating the concept to 2019, describing features of the information society, such as the spread of the internet, as Society 5.0, and writing "industrial revolution 5.0", since Industry 5.0 is a related but distinct industrial concept (see [Potočan et al., 2021](#), on how the two relate).

For education, the more important point is epistemic. Society 5.0 describes a desired future articulated by a government. The sentence "Indonesia is transitioning to Society 5.0" is therefore a policy claim, not a finding, and the claim that education "will be left behind" if it does not adapt is a technologically deterministic argument resting on a policy slogan. A review can use the vision as a frame while asking what evidence supports specific school practices, whom they serve and who may be excluded. [Potočan et al. \(2021\)](#) themselves present Society 5.0 as a problem of balancing economic advancement with social problems, a balance that Indonesian education discourse often drops.

The literacies attached to Society 5.0 also need correction. Accounts of Society 5.0 literacies are often stated inconsistently. Two published frameworks are relevant. [Aoun \(2017\)](#) proposes three new literacies for an age of AI: data, technological and human literacy. The [World Economic Forum \(2015\)](#) lists six foundational literacies: literacy, numeracy, scientific, ICT, financial, and cultural and civic literacy. A four-item list of data, technological, numeracy and human literacy, sometimes announced as six, mixes Aoun's three with numeracy. We use Aoun's three when discussing AI-era literacy and avoid the phrase "six literacies".

Indonesian Policy and Teacher Readiness (RQ2, RQ4)

Indonesia's policy environment is active but fragmented across agencies. The National AI Strategy (Stranas KA) 2020–2045 names education and research among its focus areas ([OECD.AI, n.d.-a](#)). The National AI Roadmap for 2025–2029, led by the Ministry of Communication and Digital, operationalises the strategy, lists education among its target application areas, and identifies gaps in ethical readiness, infrastructure, data, research funding and the supply and distribution of AI talent ([OECD.AI, n.d.-b](#)). Separately, the Ministry of Primary and Secondary Education introduced *pembelajaran mendalam* as a voluntary approach from 2025/2026 ([Republika, 2025](#)) and, by July 2025, had added a self-paced training programme on deeper learning, coding and AI for teachers of all subjects. In the sources we retrieved, no document specifies how DL-based systems would deliver the three elements of *pembelajaran mendalam*. This is a limit of our source set, since we did not access the underlying ministerial regulations, but it matches the pattern of shared vocabulary without a stated mechanism.

Empirical evidence on readiness is limited and of lower strength. The largest study we found is a preprint survey of 349 K-12 teachers in at least 25 provinces, recruited through non-probability channels in 2025. Teachers most often used AI for developing content knowledge and preparing teaching media, with over 64% reporting frequent use and 66% reporting satisfaction. The most common difficulties were overly generic output (37%), limited access to technology and infrastructure (35.1%) and concern about overreliance (33.1%). Teachers' most requested support

was creating assessment items (72.2%) and syllabi or lesson plans (69.6%). The pattern suggests that current use targets teacher workload, not the student-facing deeper learning that the policy describes. Subgroup differences, such as higher perceived value among the 32 teachers in Eastern Indonesia, rest on small cells and should be treated as hypotheses. A separate survey-based study found that digital leadership, digital readiness and perceived usefulness shape Indonesian teachers' intention to use AI (Harsanti et al., 2025).

What the Evidence Shows about AI Applications (RQ3)

Evidence is strongest for intelligent tutoring systems (ITS), moderate for generative AI tutors built with pedagogical safeguards, and mostly descriptive for broader claims about personalisation and prediction. Table 2 summarises the sources behind this judgement.

Table 2. Key evidence on AI applications in education, with strength of evidence

Source	Type and design	Context	Main finding	Evidence strength
VanLehn (2011)	Meta-analysis of experiments comparing tutoring types	Tutoring studies across settings	Versus no tutoring, human tutoring $d = 0.79$ and step-based ITS $d = 0.76$; answer-based systems about 0.3	Higher
Kulik & Fletcher (2016)	Meta-analysis of controlled ITS evaluations	Controlled evaluations in schools and colleges	Median ITS effect of about 0.66 standard deviations over conventional instruction	Higher
Zawacki-Richter et al. (2019)	Systematic review, 146 of 2,656 records	Higher education	Four application areas; little critical reflection on risks and weak links to pedagogy	Moderate
Bond et al. (2024)	Review of 66 reviews	Higher education	Adaptive systems and personalisation prominent; calls for ethics, rigour and wider contexts	Moderate
Kestin et al. (2025)	Randomised crossover trial, $N = 194$	University physics, United States	Custom AI tutor produced over twice the median learning gain of an active-learning lesson; higher engagement	Moderate
Bastani et al. (2025)	Randomised field experiment, nearly 1,000 students	High school mathematics, Turkey	Unguarded GPT-4 raised practice performance but lowered later unaided results by 17%; guardrails mitigated the loss	Moderate
Baker & Hawn (2022)	Review of algorithmic bias	Education generally	Evidence concentrated on a few demographic groups and on one country	Moderate
Kasneci et al. (2023)	Position paper	All levels	Opportunities and risks of large language models; users need competencies to judge outputs	Lower

In the ITS literature, VanLehn (2011) found that human tutoring raised outcomes by $d = 0.79$ over no tutoring, far below the $d = 2.0$ often quoted, and that step-based systems, which give feedback on each step of a solution, reached $d = 0.76$. Kulik and Fletcher (2016) reported a median effect of about 0.66 standard deviations across controlled evaluations. Two cautions apply. The systems in these meta-analyses were designed before the current generation of DL, so the evidence concerns tutoring design, with feedback at the level of solution steps, more than any particular algorithm. And they were mostly tested in structured domains such as mathematics and physics, so transfer to subjects such as Bahasa Indonesia, civics or religion is untested.

A common illustration of “adaptive learning” is a student who struggles with fractions and receives extra exercises. That picture hides what an adaptive system must do: model what the learner knows, choose the next task, and give feedback that helps the learner correct a specific

error. It also leaves open the teacher's role, which any evaluation should specify: who sets goals, interprets reports and handles what the system cannot. Broader labels such as personalisation, profiling and prediction dominate the review literature (Bond et al., 2024; Zawacki-Richter et al., 2019), but those reviews mainly document prevalence and urge more rigour. We found no controlled Indonesian evaluation of any of these applications.

Generative AI adds new evidence with a clear message: design matters more than access. In the Harvard trial, students learned more from a tutor built on pedagogical principles than from an active-learning class, although this was one course at one elite university, with a purpose-built tool and short topics (Kestin et al., 2025). In the Turkish field experiment, unrestricted access to GPT-4 helped students while it was available but left them worse off once it was removed, whereas a version that gave hints and safeguarded learning largely avoided the loss (Bastani et al., 2025). Both studies came from outside Indonesia and neither measured long-term retention in the way a national policy would need.

Table 3 maps these findings to the three elements of *pembelajaran mendalam*. The middle column is our interpretation of mechanism, and the evidence column states what is directly supported.

Table 3. Mapping AI applications to the three elements of deep learning

Element	Theoretical anchor	Possible AI mechanism (our interpretation)	Direct evidence	Main risk
Meaningful	Ausubel (1968)	Step-level feedback helps learners connect each step to what they already know	Step-based ITS near human tutoring (VanLehn, 2011)	Evidence from structured domains; no Indonesian subject-matter tests
Mindful	Flavell (1979)	Hints in place of answers keep learners monitoring their own reasoning	Safeguarded tutor limited the harm seen with unguarded access (Bastani et al., 2025)	Unguarded use can replace thinking; metacognition not measured directly
Joyful	Ryan & Deci (2000)	Immediate feedback and own-pace practice may support competence and autonomy	Higher self-reported engagement and motivation (Kestin et al., 2025)	Novelty effects and self-report; no Indonesian evidence

Generative AI: How It Works and What Follows for Teaching (RQ3, RQ4)

Large language models such as those behind ChatGPT are first trained on very large collections of unlabelled text, then tuned on instructions and human feedback so that their answers follow what users ask for (Ouyang et al., 2022). Descriptions that stop at feeding unlabelled data to a deep learning algorithm omit instruction tuning, and the claim that the result is quick and accurate overlooks a known failure: these models can produce fluent statements that are false, often called hallucination (Ji et al., 2023). Kasneci et al. (2023) set out opportunities for learners from primary school to university, but argue that teachers and learners need competencies to understand the technology and its limits. UNESCO's guidance on generative AI (Miao & Holmes, 2023) proposes a human-centred approach: protect learner data, set an age limit for independent conversations with generative AI, and validate tools for ethical and pedagogical appropriateness before use.

These points bear directly on Indonesian practice. If teachers mainly use AI to generate lesson plans, items and teaching media, then checking factual accuracy and curriculum alignment becomes part of the teacher's work, and the 37% of teachers who report generic output are describing a quality problem that training alone may not solve. If students use AI directly, the Turkish experiment suggests that tool design and classroom rules matter as much as access (Bastani et al., 2025).

Risks, Equity and Context (RQ4)

Algorithmic bias is documented in educational prediction and decision systems, but the evidence base is narrow: Baker and Hawn (2022) found it concentrated on a small number of

racial and ethnic groups and on sex, largely in the United States. That limit cuts both ways. It means that Indonesian regions, languages, religions, disability groups and school types are almost unstudied, and that a model validated elsewhere cannot be assumed fair in Indonesia. Learner-data privacy, one of the issues UNESCO asks governments to regulate, is an open question for tools adopted school by school (Miao & Holmes, 2023). Academic integrity is the third recurring concern in the literature on large language models (Kasneji et al., 2023).

Access is the most concrete constraint. The national AI roadmap itself lists infrastructure and talent distribution among its gaps (OECD.AI, n.d.-b), and 35.1% of surveyed teachers named limited access to technology. Bond et al. (2024) note that research contexts need broadening to developing countries. A reasonable inference, which we offer as a hypothesis and not a finding, is that technology-led reform will widen existing gaps between well-connected and poorly connected schools unless access, offline-capable tools and teacher support are funded together. The same logic applies to teacher agency. Training programmes address skills, but whether teachers can adapt, reject or redesign AI tools in their own subjects is a separate question that current Indonesian evidence does not answer.

Implications for Practice, Policy and Research

The findings support concrete, evidence-linked recommendations, summarised in Table 4. They follow from the evidence in Table 2 and should be read with its strength ratings.

Table 4. Implications by audience

Audience	Implication	Basis
Teachers	Use generative AI to draft materials but verify accuracy and curriculum fit; for student use, prefer tools that give hints rather than answers and have students attempt tasks first	Ji et al. (2023); Bastani et al. (2025)
School leaders	Build digital readiness and leadership before tool adoption, and plan for offline or low-bandwidth use	Harsanti et al. (2025)
Policymakers	State which meaning of deep learning a policy uses, specify how AI is expected to support each element of pembelajaran mendalam, fund infrastructure with training, and require privacy protection and validation of tools	Miao & Holmes (2023); OECD.AI (n.d.-b)
Developers	Design for pedagogical guardrails, Indonesian-language and regional contexts, and uneven connectivity	Bastani et al. (2025)
Researchers	Run controlled, subject-specific evaluations in Indonesian schools and test for subgroup fairness	Baker & Hawn (2022); Bond et al. (2024)

Five directions for future research follow from the gaps identified.

1. Do AI tutors with pedagogical guardrails improve learning in Indonesian classrooms, in subjects beyond mathematics and science, and for how long?
2. Which teacher practices turn AI-generated materials into deeper learning, and which produce the generic output teachers report?
3. Do AI tools perform equally well across Indonesian regions, languages, school types and disability groups?
4. Can *meaningful*, *mindful* and *joyful* learning be measured, and do AI-supported lessons change those measures?
5. What combination of infrastructure, offline tools and training allows low-connectivity schools to benefit?

Limitations

This is a narrative review, and its conclusions are bounded by its method. Sources were selected purposively through a general web search engine, not through licensed bibliographic databases, so relevant studies may be missing and selection bias is possible; no record counts can be given. Screening and coding were not duplicated by a second coder. The evidence-strength ratings are a heuristic and not a formal quality appraisal. Most sources are in English, and

Indonesian-language research may be under-represented despite searching in both languages. Indonesian policy was documented partly through secondary reports and international policy pages, and we did not access the underlying ministerial regulations. The Indonesian empirical evidence is thin, includes a preprint and a non-probability sample, and cannot support national generalisation. The tutoring meta-analyses predate current generative models, and the two generative AI trials come from the United States and Turkey, each in a single setting. Finally, the field is moving quickly, and results from 2025 may be revised by later studies.

CONCLUSION

This review set out to separate and then connect three ideas that Indonesian policy treats together. Deep learning as a neural-network technique and deeper learning as a pedagogical aim are independent (RQ1). Society 5.0 is a Japanese policy vision from 2016 that can frame discussion but cannot substitute for evidence (RQ2). The strongest evidence supports carefully designed tutoring, with step-level feedback or pedagogical guardrails, and warns that unguided access can reduce later learning (RQ3). Indonesian evidence is limited, shows teachers using AI mainly to ease workload, and points to infrastructure, output quality and overreliance as constraints (RQ4). The practical conclusion is that AI should be positioned in Indonesia as a scaffold for deeper learning, adopted where teacher capacity, equitable access and local evaluation are in place, and that policy should state the mechanism by which each tool is expected to support meaningful, mindful and joyful learning.

DECLARATIONS

Author Contributions: Conceptualization, A.D.S. and A.F.S.A.M.; methodology, A.D.S. and A.F.S.A.M.; investigation, A.D.S., V.A.F., I.N.S., and R.D.R.; data curation, A.D.S., V.A.F., and R.D.R.; formal analysis, V.A.F.; validation, I.N.S.; visualization, I.N.S.; resources, R.D.R.; writing—original draft, A.D.S., V.A.F., and I.N.S.; writing—review and editing, R.D.R. and A.F.S.A.M.; supervision, A.F.S.A.M.; project administration, A.F.S.A.M.

Funding: This research received no external funding.

Ethical Clearance and Informed Consent: Not applicable. This review analysed published literature and public policy documents and involved no human participants.

Conflict of Interest: The authors declare no conflict of interest.

Data Availability Statement: Data available on request from the authors.

Declaration of AI Use: The authors did not use generative AI or AI-assisted tools in the preparation of this manuscript.

REFERENCES

- Aoun, J. E. (2017). *Robot-proof: Higher education in the age of artificial intelligence*. MIT Press. <https://doi.org/10.7551/mitpress/11456.001.0001>
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.
- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21, Article 4. <https://doi.org/10.1186/s41239-023-00436-z>
- Campbell, M., Hoane, A. J., Jr., & Hsu, F.-H. (2002). Deep Blue. *Artificial Intelligence*, 134(1–2), 57–83. [https://doi.org/10.1016/S0004-3702\(01\)00129-1](https://doi.org/10.1016/S0004-3702(01)00129-1)
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>

- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, Article 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Deng, L., & Yu, D. (2014). Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3–4), 197–387. <https://doi.org/10.1561/20000000039>
- Diputera, A. M., Zulpan, & Eza, G. N. (2024). Memahami konsep pendekatan deep learning dalam pembelajaran anak usia dini yang meaningful, mindful dan joyful: Kajian melalui filsafat pendidikan [Understanding the deep learning approach in early childhood learning that is meaningful, mindful and joyful: A study through the philosophy of education]. *Jurnal Bunga Rampai Usia Emas*, 10(2), 108–120. <https://doi.org/10.24114/jbrue.v10i2.65978>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Fukuyama, M. (2018). Society 5.0: Aiming for a new human-centered society. *Japan SPOTLIGHT*, July/August, 47–50.
- Fullan, M., & Langworthy, M. (2014). *A rich seam: How new pedagogies find deep learning*. Pearson.
- Harsanti, H. R., Sudibjo, N., Riady, S., & Yu, P. (2025). Exploring Indonesian teachers' intention to use artificial intelligence in schools: The impact of digital leadership, digital readiness, and perceived usefulness. *Cogent Social Sciences*, 11(1), Article 2593596. <https://doi.org/10.1080/23311886.2025.2593596>
- Hitachi-UTokyo Laboratory. (2020). *Society 5.0: A people-centric super-smart society*. Springer. <https://doi.org/10.1007/978-981-15-2989-4>
- Ji, Z., Lee, N., Friesler, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 248. <https://doi.org/10.1145/3571730>
- Kasneçi, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneçi, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kestin, G., Miller, K., Kiales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, 15(1), Article 17458. <https://doi.org/10.1038/s41598-025-97652-6>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology*, 46(1), 4–11. <https://doi.org/10.1111/j.2044-8279.1976.tb02980.x>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- OECD.AI. (n.d.-a). *National strategy for artificial intelligence (Stranas KA) Indonesia 2020–2045* [Policy initiative record]. OECD.AI Policy Navigator. Retrieved October 2, 2025, from <https://oecd.ai/en/dashboards/policy-initiatives/national-ai-strategy-4627>
- OECD.AI. (n.d.-b). *White paper on the National Artificial Intelligence Roadmap* [Policy initiative record]. OECD.AI Policy Navigator. Retrieved October 2, 2025, from <https://oecd.ai/en/dashboards/policy-initiatives/national-artificial-intelligence-roadmap>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744. <https://doi.org/10.52202/068431-2011>
- Potočan, V., Mulej, M., & Nedelko, Z. (2021). Society 5.0: Balancing of Industry 4.0, economic advancement and social problems. *Kybernetes*, 50(3), 794–811. <https://doi.org/10.1108/K-12-2019-0858>
- Pusat Penelitian Badan Keahlian DPR RI. (2024, November). Pendekatan deep learning [The deep learning approach] [Isu Sepekan]. https://berkas.dpr.go.id/pusaka/files/isu_sepekan/Isu%20Sepekan---II-PUSLIT-November-2024-201.pdf

- Republika. (2025). Deep learning akan diterapkan, Mendikdasmen: Memperbaiki pembelajaran di Indonesia [Deep learning will be implemented, Minister of Primary and Secondary Education: Improving learning in Indonesia]. <https://news.republika.co.id/berita/svj8ta451/deep-learning-akan-diterapkan-mendikdasmen-memperbaiki-pembelajaran-di-indonesia>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419), 1140–1144. <https://doi.org/10.1126/science.aar6404>
- Subandowo, M. (2022). Teknologi pendidikan di era Society 5.0 [Educational technology in the Society 5.0 era]. *Jurnal Sagacious*, 9(1), 24–35. <https://rumahjurnal.net/sagacious/article/view/1139>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- World Economic Forum. (2015). *New vision for education: Unlocking the potential of technology*. https://www3.weforum.org/docs/WEFUSA_NewVisionforEducation_Report2015.pdf
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>