Automatic question generation (AQG): A systematic review of methods, applications, challenges and implications in educational technology

Supriyadi Supriyadi
Bandung Institute of Technology, Indonesia
https://orcid.org/0000-0003-1960-1327

Fetty Fitriyanti Lubis
Bandung Institute of Technology, Indonesia
https://orcid.org/0000-0002-0397-9782

Yusep Rosmansyah
Bandung Institute of Technology, Indonesia
https://orcid.org/0000-0003-2314-6734

The rapid transition to digital education has intensified the demand for personalised evaluation, yet the manual creation of high-quality assessment items remains critically resource-intensive. To address this challenge, this systematic review examined automatic question generation (AQG) methods, applications, challenges and implications in educational technology. A total of 127 peer-reviewed studies published between 2018 and 2025 were analysed. AQG approaches are classified into template-based, rule-based, machine learning and hybrid methods. While template and rule-based systems offer stability and interpretability, they face scalability and rigidity constraints. Conversely, machine learning-based techniques, particularly transformer architectures, demonstrate superior adaptability, though performance remains hindered by data scarcity. Emerging hybrid models show promise in balancing precision with flexibility. Key challenges identified include multilingual corpus availability, the generation of complex cognitive questions and ethical concerns regarding algorithmic transparency. We recommend embedding AQG within adaptive platforms to enhance instructional efficiency. Future research should prioritise expansive multilingual data sets, automated rule formulation and robust ethical frameworks to guide responsible deployment.

Implications for practice or policy:

- Educational technology leaders should embed AQG into adaptive systems to enable real-time feedback
- Teacher training must include AQG competencies, focusing on question validation.
- Higher education institutions should pilot AQG within learning management systems to automate assessments.
- Policymakers must establish regulatory frameworks for data privacy and algorithmic transparency.
- Researchers should conduct longitudinal studies to provide empirical evidence on learning outcomes.

Keywords: automatic question generation, educational technology, learning outcomes, personalised assessment, Bloom's taxonomy, systematic literature review

Introduction

Question generation (QG) refers to generating questions from a given context, optionally based on specific answers (Akyön et al., 2022). The processes involved in automatic question generation (AQG) have garnered significant research interest in the field of natural language processing (NLP; Xie, Fang et

al., 2022). AQG is defined as the automated process of generating questions, which involves creating natural language questions from various input sources such as images, natural language texts, structured web tables and knowledge base sub-graphs (Bao et al., 2018). AQG is essentially the reverse concept of automatic question answering (AQA). While both share the common goal of employing natural language approaches for text processing, their objectives differ: AQA focuses on generating answers to questions, whereas AQG aims to produce questions from answers or input texts (Sun et al., 2020).

Although AQG is often perceived as primarily applicable to education, it can be implemented in various fields. In education, question generation is an integral part of the assessment process, closely tied to achieving the learning objectives. These objectives are typically articulated in measurable cognitive levels, such as those defined by Bloom's taxonomy (Krathwohl, 2002). AQG proves particularly beneficial in creating various types of questions, including multiple-choice tests (Chomphooyod et al., 2023; Kumar et al., 2024; Liu et al., 2018), quizzes for instructional materials (Murakhovs'ka et al., 2022), as well as human-machine dialogue systems, and interactive question-answering systems (Bousmaha et al., 2020).

The complexity of AQG is heavily influenced by the intricacies of natural language understanding. Consequently, AQG research often intersects with multiple disciplines, including education, particularly assessment, and computational fields. Recent advancements in NLP techniques and machine learning (ML) algorithms have significantly increased interest in AQG. This emerging research domain delves into the automation of question generation processes, to produce coherent, relevant and domain-specific questions. As the demand for scalable and adaptive learning technologies rises, AQG's role in educational technologies, assessment systems and information retrieval applications has become increasingly prominent.

To provide a structured analysis, this systematic literature review adopted the sample, phenomenon of interest, design, evaluation, research type framework (Cooke et al., 2012). It sought to address the following research questions (RQs) within the context of tertiary educational technology:

1. RQ1: How do different AQG methods compare in terms of technical effectiveness and computational efficiency?
2. RQ2: What are the primary applications of AQG research, and how do they influence the types of questions generated?
3. RQ3: What are the key challenges and limitations faced by current question generation techniques?
4. RQ4: What are the broader implications and benefits of AQG for educational technology stakeholders, specifically in enhancing assessment quality and personalised learning?

RQ1 evaluated the effectiveness (linguistic and pedagogical quality) and efficiency (scalability and workload reduction) of rule-based, template-based and data-driven AQG. It synthesised how advancements in NLP, ML and deep learning have reshaped these methodologies to ensure reliable educational outcomes.

RQ2 explored AQG applications across various question types and cognitive levels. It examined how formats from multiple-choice to open-ended are aligned with frameworks like Bloom's taxonomy (Murakhovs'ka et al., 2022) to meet specific contextual and measurable requirements.

RQ3 identified technical challenges in semantic understanding, linguistic variatio, and algorithm selection, alongside social factors like computational trust. Pinpointing these obstacles provides a roadmap for innovative solutions and future research directions.

RQ4 examined the broader ramifications of AQG on pedagogical practices and assessment systems. By evaluating both beneficial and adverse consequences across diverse contexts, this study aimed to optimised outcomes and proactively manage emerging technological challenge

Methodology

Search strategy and selection criteria

Following the search strategy outlined in Table 1, this review adhered to the sample, phenomenon of interest, design, evaluation, research framework (Cooke et al., 2012) to maintain methodological rigour. The literature search focused on AQG methods, applications, challenges and educational technology implications within the 2018–2025 timeframe.

Table 1
Sample, phenomenon of interest, design, evaluation, research framework for study selection

Element	Description
Sample	Tertiary education students, instructors and higher education institutions
Phenomenon of Interest	AQG systems and their technical and pedagogical integration
Design	Systematic literature review
Evaluation	Technical effectiveness, pedagogical efficiency, challenges and instructional implications
Research type	Peer-reviewed journal articles

Electronic databases including Google Scholar, Scopus, IEEE Xplore, ACM Digital Library, Web of Science and ERIC were systematically searched. Keywords included "question generation", ”automatic question generation”, ”question generation AND methods”, ”question generation AND applications”, ”question generation AND challenges”, ”question generation AND implications”, and ”question generation AND educational technology” to ensure high-impact coverage across both engineering and pedagogical literature.

Screening process

This review employed the PRISMA 2020 framework (Page et al., 2021), building upon the foundational reporting items first proposed by Moher et al. (2009). As illustrated in Figure 1, the identification, screening and inclusion stages were meticulously followed to ensure transparency. During the identification phase, a total of 706 records were retrieved from the selected databases. After an automated duplication check, 372 records were removed, leaving 334 articles for initial screening.

During retrieval, 217 articles were sought for full-text assessment. Technical factors and restricted access excluded 78 articles, while 12 systematic literature reviews were removed to prioritise primary studies. Ultimately, 127 articles were included for thematic analysis. To minimise bias, we independently conducted the screening and resolved any discrepancies via consensus.

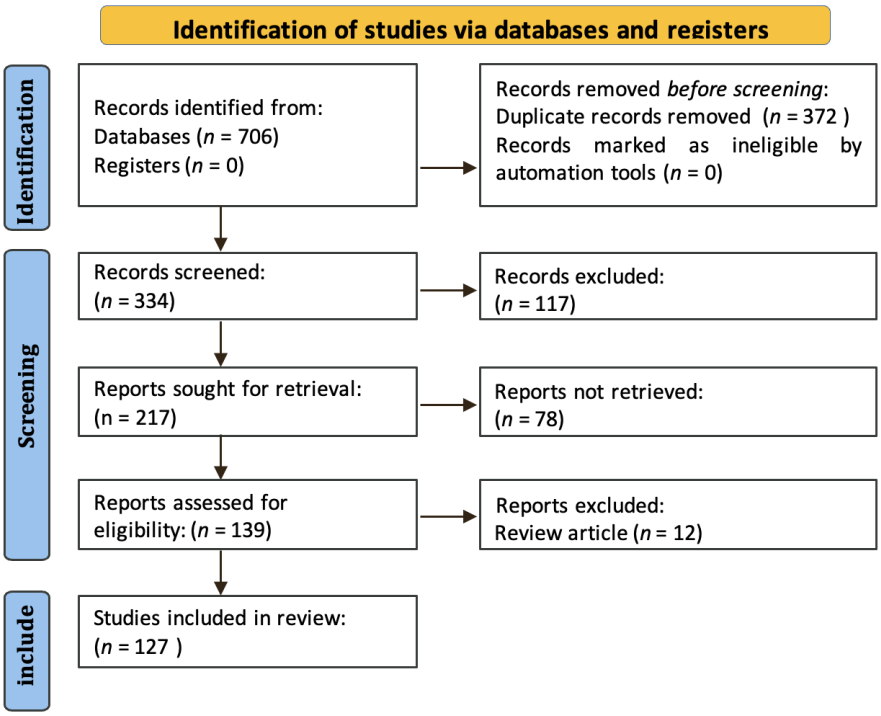


Figure 1. Phases in the systematic literature review process

Figure 2 shows a significant upward trajectory in AQG research, peaking in 2024. This trend reflects momentum driven by neural language model advancements in educational technology. Additionally, 12 existing reviews (Table 2) were used as secondary references to ensure the systematic literature review provided unique contributions without duplicating prior synthesis.

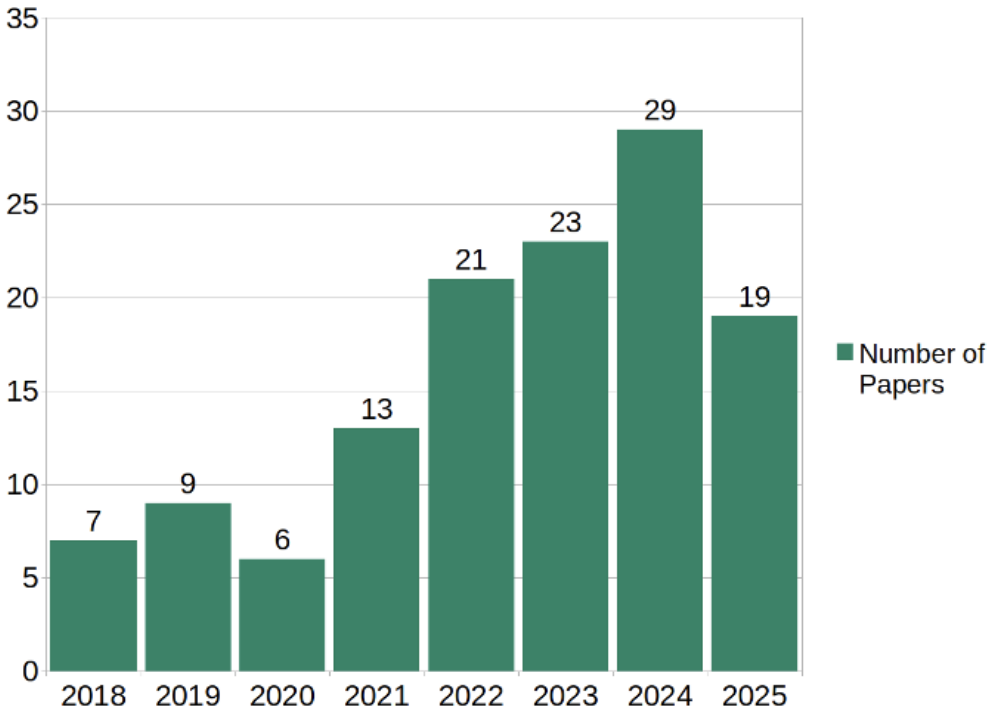


Figure 2. Temporal distribution of included articles (2018–2025)

Table 2

Systematic literature review articles on AQG from 2018 to 2025

Papers	Summary	Year	Review range	Papers analysed
(Kurdi et al., 2020)	AQG trends: Difficulty control gaps & template prevalence	2020	2015–2019	93
(Ch & Saha, 2020)	Six-phase multiple-choice question (MCQ) workflow: Pre-processing to distractors	2020	1999–2019	86
(Das et al., 2021)	Self-directed AQG: Sentence extraction & evaluation	2021	2001–2018	78
(Alshboul & Baksa-Varga, 2022)	Programming AQG: Pre- vs post-COVID developments	2022	2017–2022	18
(Zhang et al., 2022)	Rule-based vs neural-based QG for augmentation	2022	2017–2021	252
(Mulla & Gharpure, 2023a)	Taxonomy: Standalone, visual & conversational QG	2023	2008–2018	151
(Al Faraby et al., 2023)	Neural QG: Higher-order & multi-paragraph context	2023	2016–2022	224
(Yan et al., 2024)	Large language models: Readiness, transparency, ethics & privacy	2024	2017–2023	118
(Kıyak & Emekli, 2024)	ChatGPT: Prompt engineering for medical MCQs	2024	Until 2024	23
(Prieto-Andreu & Labisa-Palmeira, 2024)	GPT-3: Classification trends & gaps in large language model research	2024	2020–2023	34
(Tao et al., 2025)	GenAI review: 12 topics via topic modelling	2025	2010–2024	1158
(Dong et al., 2025)	QG in education review: Data sets, objectives & evaluation	2025	2015–2025	48

Synthesis and analysis

Data from the 127 articles were synthesised using comparative analysis for RQ1 and thematic analysis (Braun & Clarke, 2006) for RQ2–RQ4. This six-phase process transformed diverse findings into a cohesive synthesis, identifying patterns in domains, stakeholder benefits and ethical challenges. To ensure validity and minimise bias, themes were validated through cross-case synthesis. Finally, an adapted PRISMA checklist (Page et al., 2021) was applied to ensure the synthesised evidence remains robust and academically rigorous.

Results and discussion

RQ1: Comparative analysis of QG methods

The fundamental goal of AQG is to automate question creation across diverse domains by determining the correct answer basis from given inputs. This process generally mirrors manual crafting through three core stages: reading input data, identifying key ideas and transforming those ideas into formal questions (Divate & Salgaonkar, 2017), as illustrated in Figure 3.

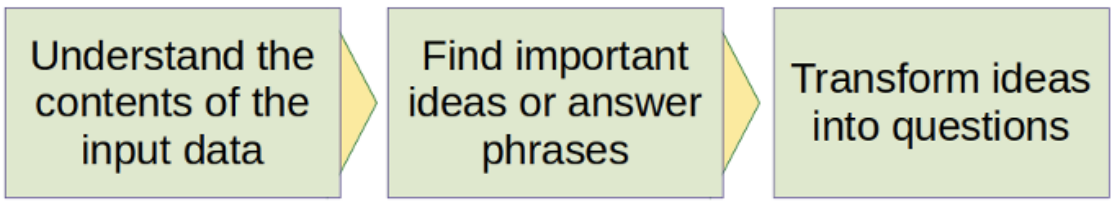


Figure 3. The process of manual question generation

AQG research (Figure 4) has expanded into education, assessment, information retrieval and dialogue systems. These systems generate diverse outputs, from factual close-ended formats to complex open-ended and analytical questions. While text-based systems dominate, a shift towards image, knowledge and hybrid inputs is increasing context awareness (Alshboul & Baksa-Varga, 2024).

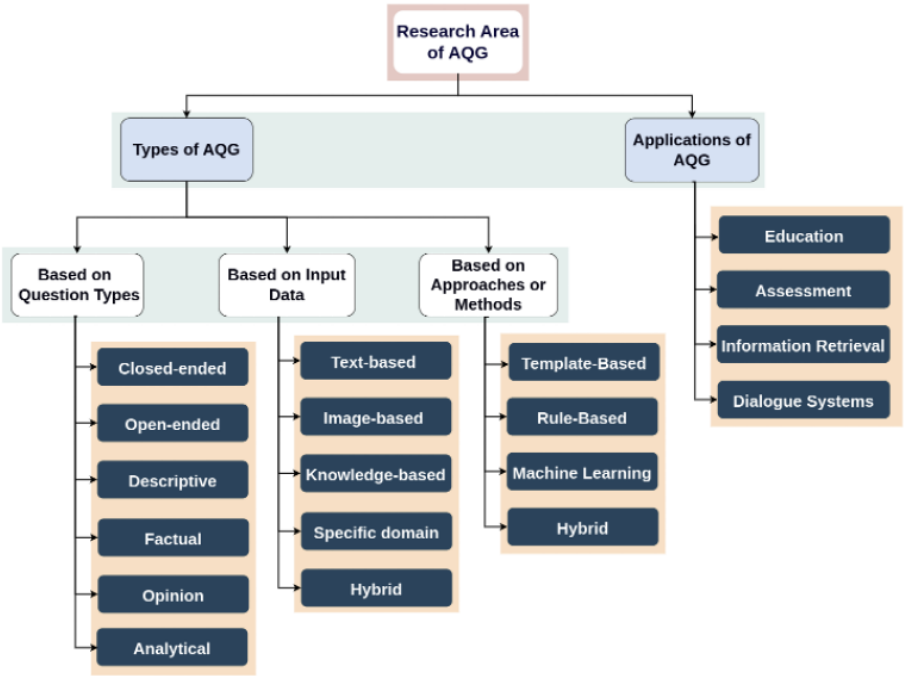


Figure 4. Research areas of AQG

Our comparative analysis identified four primary methodological clusters, each offering distinct trade-offs in pedagogical utility and technical complexity, as detailed in Table 3. Initially, template-based methods provided high efficiency by utilising pre-defined frameworks, including text-based templates (Brito et al., 2019; Willert & Thiemann, 2024) and query-based templates (Kusuma et al., 2022). While these systems ensure predictable output, they are often constrained by linguistic rigidity (Sewunetie & Kovács, 2022). To enhance flexibility, rule-based methods emerged by employing semantic rules (Alshboul & Baksa-Varga, 2024; Vasić et al., 2021) and ontology-based frameworks (Demaidi et al., 2018; Kumar et al., 2024). However, the manual effort required to craft these rules limits their scalability across different subjects (Keklik et al., 2019).

Table 3
Classification of AQG development methods

Category	Specific techniques	Count	Representative papers
Template-based	Text-based templates	5	(Brito et al., 2019; Das et al., 2022; Espinosa-López et al., 2020; Izquierdo-Amo et al., 2025; Willert & Thiemann, 2024)
Template-based	Query-based templates	2	(Atchariyachanvanich et al., 2019; Kusuma et al., 2022)
Rule-based	Semantic rules	17	(Alshboul & Baksa-Varga, 2024; Bousmaha et al., 2020; Deena & Raja, 2019, 2022b, 2022a; Dhomse & Sharma, 2024; Faza et al., 2024; Grubišić et al., 2023; Hsiao & Chung, 2022; Keklik et al., 2019; Khandait et al., 2022; Muñoz et al., 2024; Patil et al., 2024; Sewunetie & Kovacs,

Category	Specific techniques	Count	Representative papers
			2024; Tiwari et al., 2019; Vasić et al., 2021; Vyas et al., 2022)
Rule-based	Ontology-based rules	3	(Demaidi et al., 2018; Kumar et al., 2024; Leo et al., 2019)
ML-based	Traditional ML: Support vector machine, linear regression, K-nearest neighbours, computerised adaptive test, GloVe	2	(Liu et al., 2018; Susanti et al., 2020)
ML-based	Neural network architectures: neural network, recurrent neural network, long short-term memory, sequence-to-sequence, hierarchical encoding	23	(Bi et al., 2025; Chu et al., 2018; Dong et al., 2024; Ji et al., 2021; Kachare et al., 2023; Madri & Meruva, 2025; Murakhovs'ka et al., 2022; Nio & Murakami, 2018; Patro et al., 2020; Sarrouti et al., 2021; Shen et al., 2024; Steuer et al., 2022; Sun et al., 2020; Sun et al., 2021; Terao et al., 2020; Tomikawa et al., 2024; Vincentio & Suhartono, 2022; Wang et al., 2020, 2021; Xie et al., 2024; Xie, Chen et al., 2023; Xie, Fang et al., 2023; Zhu et al., 2023)
ML-based	Transformer-based models: Bidirectional encoder representations from transformers, text-to-text transfer transformer, bidirectional and auto-regressive transformers, generative pre-trained transformer, retrieval-augmented generation	40	(Agrawal & Shukla, 2023; Akyön et al., 2022; Alajmi et al., 2025; Atmajaya et al., 2025; Babakhani et al., 2024; Barlybayev & Matkarimov, 2024; Bashmal et al., 2023; Bitew et al., 2024; Chen & Shiu, 2025; Chomphooyod et al., 2023; Fahad et al., 2024; Gourari et al., 2025; Grévisse et al., 2024; Huang et al., 2024; Hwang et al., 2023; Jabr & Azmi, 2025; Ji et al., 2022; Kim et al., 2025; Lafkiar & En Nahnahi, 2024; Lee et al., 2021; Li et al., 2024; Liu & Zhang, 2022; Lohr et al., 2025; Lopez et al., 2024; Matsumori et al., 2023; Moreno-Cediel et al., 2024; Mulla & Gharpure, 2023b; Nikolovski et al., 2025; Pandraju & Mahalingam, 2021; Patil et al., 2022; Rao et al., 2021; Reid et al., 2025; Ruma et al., 2023; Schroeder et al., 2022; Suhartono et al., 2024, 2024; Wang et al., 2025; Weinzierl & Harabagiu, 2023; Wen & Chu, 2025; Xie, Peng et al., 2022; Zhuge et al., 2025)
ML-based	Advanced architectures: Generative adversarial networks, graph convolutional networks, reinforcement learning, attention networks	11	(Bao et al., 2018; Chen, Shen et al., 2024; Chung et al., 2024; Guo et al., 2021; Shoaib et al., 2025; Uehara & Harada, 2024; Xie et al., 2025; Xu et al., 2021; Zeng et al., 2021, 2024; Zhao et al., 2025)
Hybrid methods	Integration of ML with rule-based, template-based or ML-based constraints	24	(Andersson & Picazo-Sanchez, 2023; Bachiri & Mouncif, 2023; Basuki & Kusuma, 2018; Bernasconi et al., 2025; Chen, Wu et al., 2024; Chen, Xiao et al., 2024; Contreras et al., 2019; De Kuthy et al., 2022; Fung et al., 2023; Guan et al., 2023; Kalpakchi & Boye, 2024; Lu et al., 2021; Maheen et al., 2022; Murad et al., 2023; Qin et al., 2024; Riza et al., 2019; Roy et al., 2023;

Category	Specific techniques	Count	Representative papers
			Sewunetie & Kovács, 2022; Shan et al., 2022; Shao et al., 2022; Shwe et al., 2025; Xie, Fang et al., 2022; Yu et al., 2021; Zheng et al., 2018)

Recent years have shown a transition towards ML and hybrid architectures. Figure 5 reveals a paradigm shift dominated by transformer-based models (Agrawal & Shukla, 2023; Mulla & Gharpure, 2023b), alongside neural network, reinforcement learning and attention networks (Bi et al., 2025; Guo et al., 2021). Hybrid methods, peaking in 2022–2023, integrate ML with rule-based constraints to balance linguistic fluency and pedagogical accuracys (Lu et al., 2021; Murad et al., 2023).

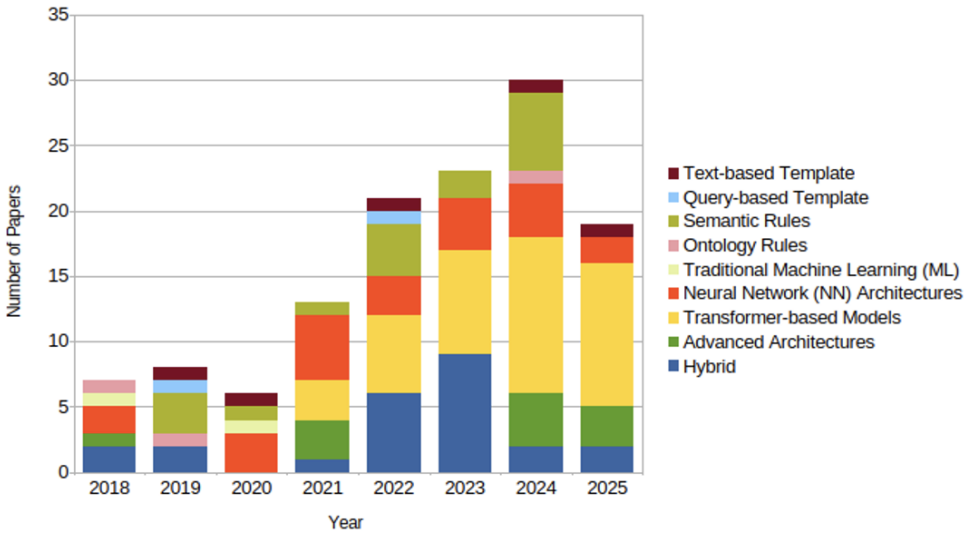


Figure 5. Evolutionary trends of AQG development methodologies 2018–2025

This choice dictates output granularity. Figure 6 shows a balanced focus between close-ended (47 papers) and open-ended formats (44 papers). Modern architectures have expanded beyond MCQs towards descriptive and analytical outputs (Bachiri & Mouncif, 2023; Shao et al., 2022). While text remains the primary input (86%), Figure 7 highlights a rise in multimodal architectures involving image, audio and video (Alshboul & Baksa-Varga, 2024). This evolution prioritises both technical efficiency and pedagogical versatility.

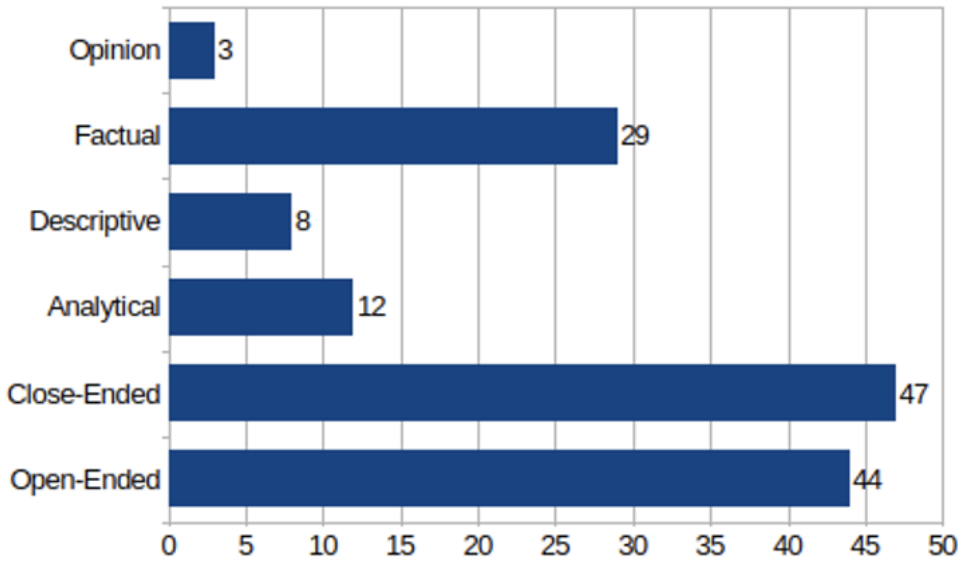


Figure 6. Types of questions generated by AQG

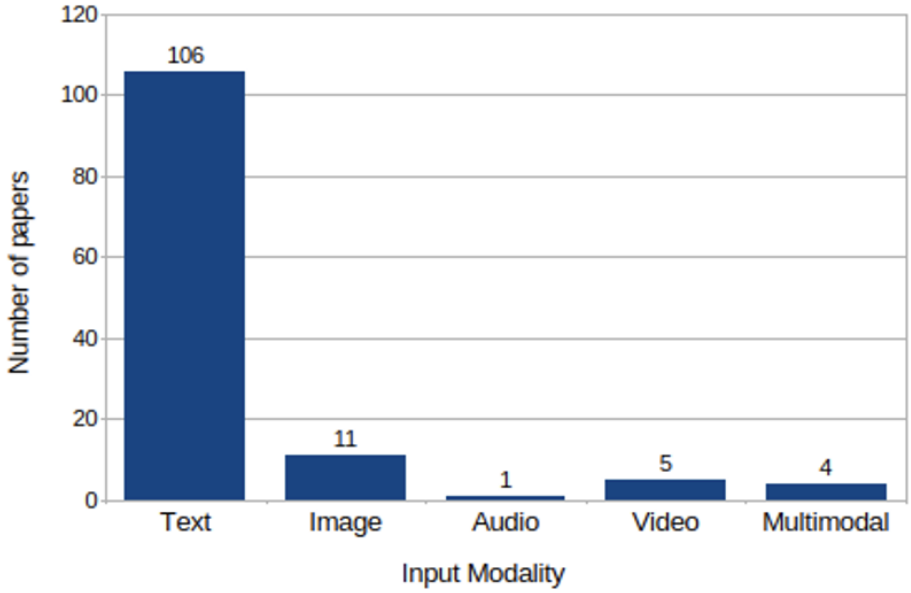


Figure 7. Types of input data in AQG

RQ2: Primary applications of AQG research

To identify the thematic landscape and domain coverage of these advancements, a bibliometric analysis based on abstract data was conducted using VOSviewer. As illustrated in Figure 8, the map is heavily dominated by technical nodes such as "question", "model" and "dataset," reflecting a primary focus on establishing robust benchmarks (Andersson & Picazo-Sanchez, 2023; Bitew et al., 2024). However, the presence of smaller nodes like "learner", "student" and "assessment" indicates an emerging shift towards applied educational utility (Bachiri & Mouncif, 2023). Synthesis of the 127 selected studies reveals that the education sector is the primary domain at 82.11%, followed by healthcare (11.58%), e-commerce (3.16%) and general domains (13.68%).

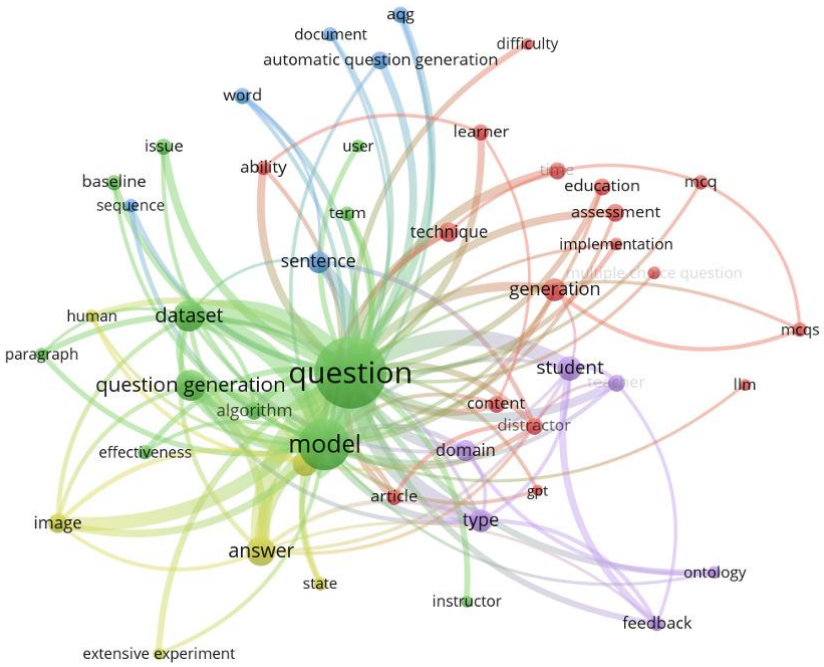


Figure 8. Network visualisation of term co-occurrence derived from article abstracts

In the education sector, development focuses on linguistic diversity and academic specificity. Research remains localised for languages including Arabic (Bousmaha et al., 2020), Chinese (Liu & Zhang, 2022), Indonesian (Basuki & Kusuma, 2018) and English (Matsumori et al., 2023), with specialised applications in Bengali, Croatian, German, Tibetan and Turkish. Specialised academic AQGs, such as multi version question generation for mathematics (Brito et al., 2019) and tools for programming (Hsiao & Chung, 2022), engineering (Kumar et al., 2024), and standardised testing like the Test of English as a Foreign Language (Riza et al., 2019), demonstrate the method's ability to handle diverse cognitive outcomes.

As a pedagogical tool, AQG facilitates personalised learning by integrating frameworks like Bloom's taxonomy to tailor content to specific cognitive levels (Kumar et al., 2024; Murad et al., 2023). These systems dynamically adjust to a learner's proficiency, fostering engagement through targeted goals (Al Faraby et al., 2024). Beyond classrooms, AQG scales assessment in massive open online courses (Bachiri & Mouncif, 2023), enhances corporate training and interview skills (Qin et al., 2024) and powers interactive chatbots (Shao et al., 2022). By reducing manual costs and enhancing communicative competence, AQG is becoming a cornerstone of intelligent tutoring and data-driven pedagogy.

RQ3: Challenges and limitations in question generation

AQG development faces diverse challenges across different methodological paradigms. In this this study, we defined question types as a dual construct involving structural formats (e.g., close-ended, open-ended and factual) and cognitive levels. Rule-based and template-based systems are primarily constrained by grammatical rigidity and a lack of scalability. Creating comprehensive rules is labour-intensive and often results in question redundancy, where generated outputs follow repetitive patterns that fail to capture linguistic variety (Bashmal et al., 2023; Matsumori et al., 2023). In contrast, ML-based architectures suffer from data scarcity and domain-specific degradation. Models trained on general data sets often struggle with high-quality, domain-specific knowledge extraction, making them prone to hallucinations and semantic inconsistencies in complex scenarios (Agrawal & Shukla, 2023).

To mitigate these issues, hybrid systems attempt to reconcile ML fluency with rule-based control; however, they face significant architectural complexity. Balancing the "black-box" nature of neural networks with formal pedagogical constraints remains a technical hurdle that often leads to evaluation difficulties (Sewunetie & Kovács, 2022). Pedagogically, aligning generated questions with specific cognitive levels remains a major barrier across all methods. Questions that lack instructional calibration or contain morphological errors can hinder learner progress. Consequently, overcoming these limitations requires rigorous empirical validation to ensure outputs meet both technical and pedagogical standards (Kachare et al., 2023; Qin et al., 2024).

RQ4: Pedagogical implications of AQG in education

AQG enhances assessment quality and fosters personalised learning by impacting three primary entities: teachers, students and educational institutions (Figure 9). While these systems automate resource creation, human validation remains indispensable to ensure accountability (Lu et al., 2021). The systematic mapping of these implications is summarised in Table 4.

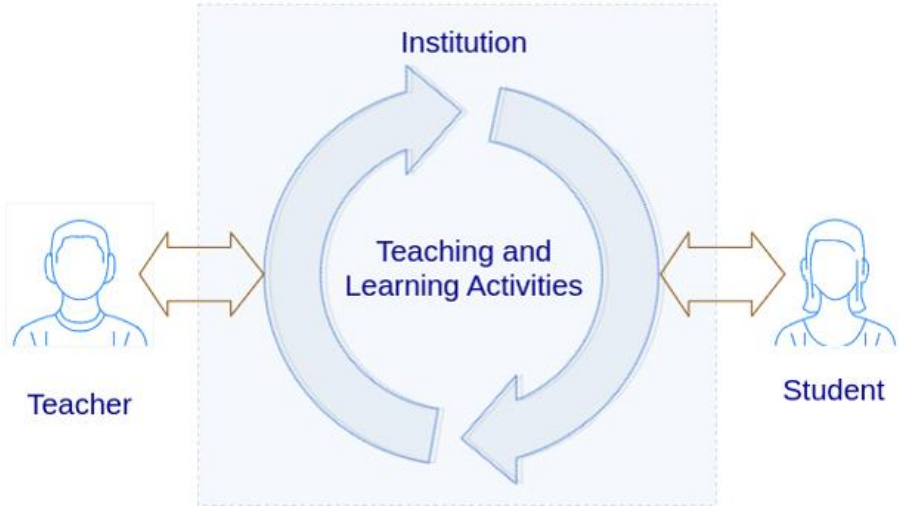


Figure 9. Interaction of teacher, student and educational institution in the learning process

Table 4
Application areas of AQG in education

Application area	Research focus	Papers
Assessment & evaluation	Formative and summative tools, alignment with learning outcomes, automated question banks and complexity	(Agrawal & Shukla, 2023; Al Faraby et al., 2023; Alshboul & Baksa-Varga, 2022; Atchariyachanvanich et al., 2019; Bachiri & Mouncif, 2023; Ch & Saha, 2020; Chen, Wu et al., 2024; Das et al., 2021, 2022; De Kuthy et al., 2022; Deena & Raja, 2022b; Guo et al., 2021; Liu et al., 2018; Lu et al., 2021; Mulla & Gharpure, 2023b; Murad et al., 2023; Sewunetie & Kovacs, 2024)
Adaptive learning	Scaffolded practice, real-time adaptation and student engagement	(Hsiao & Chung, 2022; Kiran et al., 2023)
Teacher support	Resource efficiency, enhanced variety and focused remediation	(Akyön et al., 2022; Bao et al., 2018; Bousmaha et al., 2020; Chomphooyod et al., 2023; Deena & Raja, 2019; Guan et al., 2023; Ji et al., 2021, 2022; Kachare et al., 2023; Kumar et al., 2023; Kusuma et al., 2022; Li et al., 2023; Murakhovs'ka et al., 2022; Pandraju & Mahalingam, 2021; Patro et al., 2020; Peng et al., 2024; Qin et al., 2024; Rao et al., 2021; Riza et al., 2019; Sarrouti et al., 2021; Sewunetie & Kovács, 2022; Steuer et al., 2022; Suwarningsih et al., 2015; Vasić et al., 2021; Xie, Chen et al., 2023; Xie, Fang et al., 2022; Xie, Peng et al., 2022; Xu et al., 2021; Yan et al., 2024; Zheng et al., 2018; Zhu et al., 2023)
Gamified platforms	Instant QG for games, multiplayer interactions and dynamic challenges	(Bashmal et al., 2023; Lopez-Martinez et al., 2020; Nyameino et al., 2019; Vincentio & Suhartono, 2022; Zhang et al., 2022)
Language & literacy	Vocabulary, grammar, reading comprehension and dialogue practice	(Fung et al., 2023; Khandait et al., 2022; Lee et al., 2021; Liu & Zhang, 2022; Riza et al., 2023; Yu et al., 2021; Zeng et al., 2021)
Science, technology, engineering and	Equation-based QG, logical analysis and analytical reasoning	(Brito et al., 2019; Dodiya et al., 2021; Zhou et al., 2023)

Application area	Research focus	Papers
mathematics education		
Self-directed study	Self-quizzing, exam preparation and automated feedback mechanisms	(Contreras et al., 2019; Ji et al., 2022; Keklik et al., 2019; Matsumori et al., 2023; Roy et al., 2023; Shao et al., 2022; Susanti et al., 2020)
Intelligent systems	Educational chatbots and quality assurance systems	(Hwang et al., 2023; Nio & Murakami, 2018; Paramasivam & Nirmala, 2022; Ruma et al., 2023; Sun et al., 2020, 2021a; Wang et al., 2021; Weinzierl & Harabagiu, 2023)

In assessment and evaluation, AQG aligns performance with cognitive levels by improving MCQ distractor quality and generating high-quality open-ended questions (Guo et al., 2021; Murad et al., 2023; Xie, Peng et al., 2022). It further enables personalised learning by tailoring content to proficiency and identifying knowledge gaps (Hsiao & Chung, 2022; Hwang et al., 2023). Integrating frameworks like Bloom's taxonomy allows AQG to support scaffolded practice and real-time adaptation (Kumar et al., 2024).

Regarding resource creation, automation reduces manual preparation effort, letting educators focus on remediation while acting as expert validators (Riza et al., 2019; Sun et al., 2021; Vincentio & Suhartono, 2022). AQG also enhances gamified platforms and chatbots via instant feedback and personalised recommendations (Bashmal et al., 2023). In science, technology, engineering and mathematics, it effectively manages complex equations and logical reasoning (Brito et al., 2019; Kumar et al., 2024).

Future work

Findings from this review revealed critical research trajectories. First, based on the technological shifts identified in RQ1, future research should prioritise scalable, plug-and-play frameworks that adjust to diverse topics without extensive retraining (Kumar et al., 2024). Second, to address the application gaps noted in RQ2, studies must expand beyond simple MCQs towards complex reasoning and scenario based assessments, hybridising rule-based precision with the contextual depth in large language models (Mulla & Gharpure, 2023b). Third, following the challenges highlighted in RQ3, new pedagogical metrics must be developed to replace insufficient automated scores like the bilingual evaluation understudy. Finally, responding to the broader impacts discussed in RQ4, the next frontier lies in mapping questions directly to outcome-based education frameworks while addressing ethical imperatives such as algorithmic transparency and privacy (Murad et al., 2023).

Conclusion

This review mapped the AQG landscape from 2018 to 2025, confirming a technical transition from rule-based to deep learning and hybrid models. Despite the education sector's 82.11% dominance in implementation, a gap remains in pedagogical alignment, as most systems prioritise linguistic fluency over educational utility.

The study's primary contribution is identifying a necessary shift from model-centric to student-centric development. While AQG reduces teacher workload and enables personalised practice, its potential depends on human-in-the-loop validation and metrics that measure cognitive alignment beyond syntax. Integrating frameworks like Bloom's taxonomy is vital for ensuring assessment quality (Contreras et al., 2019; Murad et al., 2023).

Ultimately, AQG is a transformative tool if technical efficiency is reconciled with pedagogical effectiveness. By grounding developments in stakeholder needs, AQG can act as a sophisticated assistant that enriches the learner experience and fosters an inclusive educational landscape.

Author contributions

Supriyadi Supriyadi: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing; **Fetty Fitriyanti Lubis:** Formal analysis, Methodology, Supervision, Validation, Writing – review and editing; **Yusep Rosmansyah:** Formal analysis, Methodology, Supervision, Validation, Writing – review and editing.

Acknowledgements

The authors gratefully acknowledge the support and encouragement from colleagues and supervisors at the Winner Laboratory, School of Electrical Engineering and Informatics, Bandung Institute of Technology. This work was partially supported by the Centre for Higher Education Funding and Assessment and the Indonesia Endowment Fund for Education Agency, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia.

References

References with an asterisk indicate the 127 primary studies synthesised for the RQs; those without an asterisk indicate the supplementary references, comprising 12 excluded reviews and theoretical frameworks used for background and discussion.

*Agrawal, A., & Shukla, P. (2023). Context aware automatic subjective and objective question generation using fast text to text transfer learning. I*nternational Journal of Advanced Computer Science and Applications*, *14*(4). https://doi.org/10.14569/IJACSA.2023.0140451

*Akyön, F. Ç., Çavuşoğlu, A. D. E., CengiZ, C., Altinuç, S. O., & TemiZel, A. (2022). Automated question generation and question answering from Turkish texts. *Turkish Journal of Electrical Engineering and Computer Sciences*, *30*(5), 1931–1940. https://doi.org/10.55730/1300-0632.3914

*Alajmi, A., Altabaa, H., Abed, S., & Ahmad, I. (2025). Arabic question generation using transformers. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *24*(3), 1–21. https://doi.org/10.1145/3701559

Al Faraby, S., Adiwijaya, A., & Romadhony, A. (2023). purposes. *International Journal of Artificial Intelligence in Education, 34*(2), 1008–1045. https://doi.org/10.1007/s40593-023-00374-x

*Alshboul, J., & Baksa-Varga, E. (2024). A hybrid approach for automatic question generation from program codes. *International Journal of Advanced Computer Science and Applications*, *15*(1). https://doi.org/10.14569/IJACSA.2024.0150102

Alshboul, J., & Baksa-Varga, E. (2022). A review of automatic question generation in teaching programming. *International Journal of Advanced Computer Science and Applications*, *13*(10). https://doi.org/10.14569/IJACSA.2022.0131006

*Andersson, T., & Picazo-Sanchez, P. (2023). Closing the gap: Automated distractor generation in Japanese Language testing. *Education Sciences*, *13*(12), Article 1203. https://doi.org/10.3390/educsci13121203

*Atchariyachanvanich, K., Nalintippayawong, S., & Julavanich, T. (2019). Reverse SQL question generation algorithm in the DBLearn adaptive e-learning system. *IEEE Access*, *7*, 54993–55004. https://doi.org/10.1109/ACCESS.2019.2912522

*Atmajaya, P. A., Romadhony, A., & Ihsan, A. F. (2025). Multiple-choice question generation in low resource Indonesian language with T5. *International Journal of Intelligent Engineering and Systems*, *18*(5), 363–381. https://doi.org/10.22266/ijies2025.0630.26

*Babakhani, P., Lommatzsch, A., Brodt, T., Sacker, D., Sivrikaya, F., & Albayrak, S. (2024). Opinerium: Subjective question generation using large language models. *IEEE Access*, *12*, 66085–66099. https://doi.org/10.1109/ACCESS.2024.3398553

*Bachiri, Y.-A., & Mouncif, H. (2023). Artificial intelligence system in aid of pedagogical engineering for knowledge assessment on MOOC platforms: Open EdX and Moodle. *International Journal of Emerging Technologies in Learning*, *18*(05), 144–160. https://doi.org/10.3991/ijet.v18i05.36589

*Bao, J., Gong, Y., Duan, N., Zhou, M., & Zhao, T. (2018). Question generation with doubly adversarial nets. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *26*(11), 2230–2239. https://doi.org/10.1109/TASLP.2018.2859777

*Barlybayev, A., & Matkarimov, B. (2024). Development of system for generating questions, answers, distractors using transformers. *International Journal of Electrical and Computer Engineering*, *14*(2), 1851–1863. https://doi.org/10.11591/ijece.v14i2.pp1851-1863

*Bashmal, L., Bazi, Y., Melgani, F., Ricci, R., Al Rahhal, M. M., & Zuair, M. (2023). Visual question generation from remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 3279–3293. https://doi.org/10.1109/JSTARS.2023.3261361

*Basuki, S., & Kusuma, S. F. (2018). Automatic question generation for 5W-1H open domain of Indonesian questions by using syntactical template-based features from academic textbooks. *Journal of Theoretical and Applied Information Technology*, *96*(12), 3908–3923. https://www.jatit.org/volumes/Vol96No12/25Vol96No12.pdf

*Bernasconi, E., Redavid, D., & Ferilli, S. (2025). Enhancing personalised learning with a context-aware intelligent question-answering system and automated frequently asked question generation. *Electronics*, *14*(7), Article 1481. https://doi.org/10.3390/electronics14071481

*Bi, S., Miao, Z., & Min, Q. (2025). LEMON: A knowledge-enhanced, type-constrained, and grammar-guided model for question generation over knowledge graphs. *IEEE Transactions on Learning Technologies*, *18*, 256–272. https://doi.org/10.1109/TLT.2025.3544454

*Bitew, S. K., Hadifar, A., Sterckx, L., Deleu, J., Develder, C., & Demeester, T. (2024). Learning to reuse distractors to support multiple-choice question generation in education. *IEEE Transactions on Learning Technologies*, *17*, 375–390. https://doi.org/10.1109/TLT.2022.3226523

*Bousmaha, K. Z., Chergui, N. H., Mbarek, M. S. A., & Hadrich, L. B. (2020). AQG: Arabic question generator. *Revue d'Intelligence Artificielle*, *34*(6), 721–729. https://doi.org/10.18280/ria.340606

*Brito, I., Almeida, J. J., & Machado, G. J. (2019). Mvgen: Multi version question generation for math courses. *Open Education Studies*, *1*(1), 146–150. https://doi.org/10.1515/edu-2019-0010

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Ch, D. R., & Saha, S. K. (2020). Automatic multiple choice question generation from text: A survey. *IEEE Transactions on Learning Technologies*, *13*(1), 14–25. https://doi.org/10.1109/TLT.2018.2889100

*Chen, C. H., & Shiu, M. F. (2025). KAQG: A knowledge-graph-enhanced rag for difficulty-controlled question generation. *IEEE Access*, *13*, 197234–197244. https://doi.org/10.1109/ACCESS.2025.3633838

*Chen, T., Shen, Y., Chen, X., Zhang, L., & Zhao, S. (2024). TriKF: Triple-perspective knowledge fusion network for empathetic question generation. *IEEE Transactions on Computational Social Systems*, *11*(6), 7186–7199. https://doi.org/10.1109/TCSS.2024.3418820

*Chen, Y., Wu, L., & Zaki, M. J. (2024). Toward subgraph-guided knowledge graph question generation with graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12. https://doi.org/10.1109/TNNLS.2023.3264519

*Chen, Y., Xiao, Y., Li, Z., & Liu, B. (2024). XMQAs: Constructing complex-modified question-answering dataset for robust question understanding. *IEEE Transactions on Knowledge and Data Engineering*, *36*(3), 1371–1384. https://doi.org/10.1109/TKDE.2023.3303916

*Chomphooyod, P., Suchato, A., Tuaycharoen, N., & Punyabukkana, P. (2023). English grammar multiple-choice question generation using text-to-text transfer transformer. *Computers and Education: Artificial Intelligence*, *5*, Article 100158. https://doi.org/10.1016/j.caeai.2023.100158

*Chu, W., Xue, H., Zhao, Z., Cai, D., & Yao, C. (2018). The forgettable-watcher model for video question answering. *Neurocomputing*, *314*, 386–393. https://doi.org/10.1016/j.neucom.2018.06.069

*Chung, H.-L., Chan, Y.-H., & Fan, Y.-C. (2024). Handover QG: Question generation by decoder fusion and reinforcement learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *32*, 3644–3655. https://doi.org/10.1109/TASLP.2024.3426292

*Contreras, J., Hilles, S., & Abubakar, Z. B. (2019). Automated essay scoring using ontology generator and natural language processing with question generator based on blooms taxonomy's cognitive level. *International Journal of Engineering and Advanced Technology*, *9*(1), 2448–2457. https://doi.org/10.35940/ijeat.A9974.109119

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*, *22*(10), 1435–1443. https://doi.org/10.1177/1049732312452938

Das, B., Majumder, M., Phadikar, S., & Sekh, A. A. (2021). Automatic question generation and answer assessment: A survey. *Research and Practice in Technology Enhanced Learning*, *16*, Article 5. https://doi.org/10.1186/s41039-021-00151-1

*Das, B., Majumder, M., Sekh, A. A., & Phadikar, S. (2022). Automatic question generation and answer assessment for subjective examination. *Cognitive Systems Research*, *72*, 14–22. https://doi.org/10.1016/j.cogsys.2021.11.002

*De Kuthy, K., Kannan, M., Santhi Ponnusamy, H., & Meurers, D. (2022). Exploring neural question generation for formal pragmatics: Data set and model evaluation. *Frontiers in Artificial Intelligence*, *5*, Article 966013. https://doi.org/10.3389/frai.2022.966013

*Deena, G., & Raja, K. (2019). Sentence selection using latent semantic analysis for automatic question generation in e-learning system. *International Journal of Innovative Technology and Exploring Engineering*, *8*(9), 86–91. https://doi.org/10.35940/ijitee.I7492.078919

*Deena, G., & Raja, K. (2022a). Keyword extraction using latent semantic analysis for question generation. *Journal of Applied Science and Engineering*, *26*(4), 501–510. https://doi.org/10.6180/jase.202304_26(4).0006

*Deena, G., & Raja, K. (2022b). Objective type question generation using natural language processing. *International Journal of Advanced Computer Science and Applications*, *13*(2), 539–548. https://doi.org/10.14569/IJACSA.2022.0130263

*Demaidi, M. N., Gaber, M. M., & Filer, N. (2018). OntoPeFeGe: Ontology-based personalized feedback generator. *IEEE Access*, *6*, 31644–31664. https://doi.org/10.1109/ACCESS.2018.2846398

*Dhomse, K. B., & Sharma, S. (2024). Dynamic question generation using NER with various feature extraction and NLP techniques. *Advances in Nonlinear Variational Inequalities*, *27*(3), 639–652. https://doi.org/10.52783/anvi.v27.1432

Divate, M., & Salgaonkar, A. (2017). Automatic question generation approaches and evaluation techniques. *Current Science*, *113*(09), 1683–1691. https://doi.org/10.18520/cs/v113/i09/1683-1691

*Dong, C., Shen, Y., Lin, S., Lin, Z., & Deng, Y. (2024). A unified framework for contextual and factoid question generation. *IEEE Transactions on Knowledge and Data Engineering*, *36*(1), 21–34. https://doi.org/10.1109/TKDE.2023.3280182

Dong, X., Zhang, X., Li, Z., Hou, Q., Xue, J., & Li, X. (2025). A literature review of research on question generation in education. *PeerJ Computer Science*, *11*, Article e3203. https://doi.org/10.7717/peerj-cs.3203

*Espinosa-López, A. L., Gómez-Salgado, A. A., & Lorduy-Arellano, D. D. C. (2020). CARINA Based cognitive agent for factoid wh-questions generation in EFL. *International Journal on Advanced Science, Engineering and Information Technology*, *10*(5), 1852–1859. https://doi.org/10.18517/ijaseit.10.5.10138

*Fahad, A. R., Al Nahian, N., Islam, M. A., & Rahman, R. M. (2024). Answer agnostic question generation in Bangla language. *International Journal of Networked and Distributed Computing*, *12*(1), 82–107. https://doi.org/10.1007/s44227-023-00018-5

*Faza, S., Rahmat, R. F., Gunawan, D., Lady, G. D., & Nadi, F. (2024). Automatic generation of multiple-choice questions using template-based semantic web in Indonesian language. *Journal of Theoretical and Applied Information Technology*, *102*(2), 653–663. http://www.jatit.org/volumes/Vol102No2/20Vol102No2.pdf

*Fung, Y.-C., Lee, L.-K., & Chui, K. T. (2023). An automatic question generator for Chinese comprehension. *Inventions*, *8*(1), Article 31. https://doi.org/10.3390/inventions8010031

*Gourari, A. E., Ezzini, M., Raoufi, M., Skouri, M., & Mouachi, R. (2025). Embedded systems and ai-driven education: scalable platform for automated grading, adaptive learning, and remote labs. *International Journal of Fuzzy Logic and Intelligent Systems*, *25*(3), 258–271. https://doi.org/10.5391/IJFIS.2025.25.3.258

*Grévisse, C., Pavlou, M. A. S., & Schneider, J. G. (2024). Docimological quality analysis of LLM-generated multiple choice questions in computer science and medicine. *SN Computer Science*, *5*, Article 636. https://doi.org/10.1007/s42979-024-02963-6

*Grubišić, A., Stankov, S., Žitko, B., Šarić-Grgić, I., Gašpar, A., Brajković, E., & Vasić, D. (2023). The evaluation of a semi-automatic authoring tool for knowledge extraction in the AC&NL Tutor. *JUCS - Journal of Universal Computer Science*, *29*(8), 866–891. https://doi.org/10.3897/jucs.86745

*Guan, M., Mondal, S. K., Dai, H.-N., & Bao, H. (2023). Reinforcement learning-driven deep question generation with rich semantics. *Information Processing & Management*, *60*(2), Article 103232. https://doi.org/10.1016/j.ipm.2022.103232

*Guo, Z., Zhao, Z., Jin, W., Wei, Z., Yang, M., Wang, N., & Yuan, N. J. (2021). Multi-turn video question generation via reinforced multi-choice attention network. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(5), 1697–1710. https://doi.org/10.1109/TCSVT.2020.3014775

*Hsiao, I., & Chung, C.-Y. (2022). AI-infused semantic model to enrich and expand programming question generation. *Journal of Artificial Intelligence and Technology, 2*(2), 47–54. https://doi.org/10.37965/jait.2022.0090

*Huang, P.-C., Chan, Y.-H., Yang, C.-Y., Chen, H.-Y., & Fan, Y.-C. (2024). EQGG: Automatic question group generation. *IEEE Transactions on Learning Technologies*, *17*, 1994–2007. https://doi.org/10.1109/TLT.2024.3430225

*Hwang, M.-H., Shin, J., Seo, H., Im, J.-S., Cho, H., & Lee, C.-K. (2023). Ensemble-NQG-T5: Ensemble neural question generation model based on text-to-text transfer transformer. *Applied Sciences*, *13*(2), Article 903. https://doi.org/10.3390/app13020903

*Izquierdo-Amo, R., Barbero-Aparicio, J. A., Garrido-Labrador, J. L., Olivares-Gil, A., & García-Osorio, C. I. (2025). Automatic generation of moodle cloze questions for the assessment of knowledge about lexical analysis algorithms. *Education Sciences*, *15*(1), Article 75. https://doi.org/10.3390/educsci15010075

*Jabr, R. B., & Azmi, A. M. (2025). Knowledge-aware Arabic question generation: A transformer-based framework. *Mathematics*, *13*(18), Article 2975. https://doi.org/10.3390/math13182975

*Ji, T., Lyu, C., Cao, Z., & Cheng, P. (2021). Multi-hop question generation using hierarchical encoding-decoding and context switch mechanism. *Entropy*, *23*(11), Article 1449. https://doi.org/10.3390/e23111449

*Ji, T., Lyu, C., Jones, G., Zhou, L., & Graham, Y. (2022). QAScore—An unsupervised unreferenced metric for the question generation evaluation. *Entropy*, *24*(11), Article 1514. https://doi.org/10.3390/e24111514

*Kachare, A., Kalla, M., & Gupta, A. (2023). Visual question generation answering (VQG-VQA) using machine learning models. *WSEAS Transactions on Systems*, *22*, 663–670. https://doi.org/10.37394/23202.2023.22.67

*Kalpakchi, D., & Boye, J. (2024). Quinductor: A multilingual data-driven method for generating reading-comprehension questions using universal dependencies. *Natural Language Engineering*, *30*(2), 217–255. https://doi.org/10.1017/S1351324923000037

*Keklik, O., Tuglular, T., & Tekir, S. (2019). Rule-based automatic question generation using semantic role labeling. *IEICE Transactions on Information and Systems*, *E102.D*(7), 1362–1373. https://doi.org/10.1587/transinf.2018EDP7199

*Khandait, K., S.A, B., & S.S, A. (2022). Automatic question generation through word vector synchronization using Lamma. *Indian Journal of Computer Science and Engineering*, *13*(4), 1083–1095. https://doi.org/10.21817/indjcse/2022/v13i4/221304046

*Kim, D., Won Park, S., & Lee, J. (2025). CQELedu: Design and implementation of a LangChain and GPT-4o mini-based web application for custom question generation and error-based learning in education. *IEEE Access*, *13*, 208932–208947. https://doi.org/10.1109/ACCESS.2025.3639087

Kıyak, Y. S., & Emekli, E. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: A literature review. *Postgraduate Medical Journal*, *100*(1189), 858–865. https://doi.org/10.1093/postmj/qgae065

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory Into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

*Kumar, A. P., Nayak, A., K, M. S., Chaitanya, & Ghosh, K. (2024). A novel framework for the generation of multiple choice question stems using semantic and machine-learning techniques. *International Journal of Artificial Intelligence in Education*, *34*(2), 332–375. https://doi.org/10.1007/s40593-023-00333-6

Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, *30*(1), 121–204. https://doi.org/10.1007/s40593-019-00186-y

*Kusuma, S. F., Siahaan, D. O., & Fatichah, C. (2022). Automatic question generation with various difficulty levels based on knowledge ontology using a query template. *Knowledge-Based Systems*, *249*, Article 108906. https://doi.org/10.1016/j.knosys.2022.108906

*Lafkiar, S., & En Nahnahi, N. (2024). An Arabic question generation system based on a shared BERT-base encoder-decoder architecture. *Mathematical Modeling and Computing*, *11*(3), 763–772. https://doi.org/10.23939/mmc2024.03.763

*Lee, H.-G., Jang, Y., & Kim, H. (2021). Machine reading comprehension framework based on self-training for domain adaptation. *IEEE Access*, *9*, 21279–21285. https://doi.org/10.1109/ACCESS.2021.3054912

*Leo, J., Kurdi, G., Matentzoglu, N., Parsia, B., Sattler, U., Forge, S., Donato, G., & Dowling, W. (2019). Ontology-based generation of medical, multi-term MCQs. *International Journal of Artificial Intelligence in Education*, *29*(2), 145–188. https://doi.org/10.1007/s40593-018-00172-w

*Li, L., Zhang, L., Zhu, C., Mao, Z. (2024). QGAE: An end-to-end answer-agnostic question generation model for generating question-answer pairs. *JUSTC*, *54*(1), 0102. https://doi.org/10.52396/JUSTC-2023-0002

*Liu, M., Rus, V., & Liu, L. (2018). Automatic Chinese multiple choice question generation using mixed similarity strategy. *IEEE Transactions on Learning Technologies*, *11*(2), 193–202. https://doi.org/10.1109/TLT.2017.2679009

*Liu, M., & Zhang, J. (2022). Chinese neural question generation: Augmenting knowledge into multiple neural encoders. *Applied Sciences*, *12*(3), Article 1032. https://doi.org/10.3390/app12031032

*Lohr, D., Berges, M., Chugh, A., Kohlhase, M., & Müller, D. (2025). Leveraging large language models to generate course-specific semantically annotated learning objects. *Journal of Computer Assisted Learning*, *41*(1), Article e13101. https://doi.org/10.1111/jcal.13101

*Lopez, C., Morrison, M., & Deacon, M. (2024). Language models for generating programming questions with varying difficulty levels. *European Public & Social Innovation Review*, *9*, 1–19. https://doi.org/10.31637/epsir-2024-760

*Lu, O. H. T., Huang, A. Y. Q., Tsai, D. C. L., & Yang, S. J. H. (2021). Expert-authored and machine-generated short-answer questions for assessing students' learning performance. *Educational Technology & Society*, *24*(3). https://drive.google.com/file/d/1AU9Vfp_Fn2VpW90c2hgn2gjHbGj7V8V6/view

*Madri, V., & Meruva, S. (2025). MCQs generation using ensemble model for student performance assessment. *Advances in Artificial Intelligence and Machine Learning*, *5*(1), 3519–3533. https://doi.org/10.54364/AAIML.2025.51201

*Maheen, F., Asif, M., Ahmad, H., Ahmad, S., Alturise, F., Asiry, O., & Ghadi, Y. Y. (2022). Automatic computer science domain multiple-choice questions generation based on informative sentences. *PeerJ Computer Science*, *8*, Article e1010. https://doi.org/10.7717/peerj-cs.1010

*Matsumori, S., Okuoka, K., Shibata, R., Inoue, M., Fukuchi, Y., & Imai, M. (2023). Mask and cloze: Automatic open cloze question generation using a masked language model. *IEEE Access*, *11*, 9835–9850. https://doi.org/10.1109/ACCESS.2023.3239005

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

*Moreno-Cediel, A., del-Hoyo-Gabaldon, J.-A., Garcia-Lopez, E., Garcia-Cabot, A., & de-Fitero-Dominguez, D. (2024). Evaluating the performance of multilingual models in answer extraction and question generation. *Scientific Reports*, *14*(1), Article 15477. https://doi.org/10.1038/s41598-024-66472-5

*Mulla, N., & Gharpure, P. (2023a). Automatic question generation: A review of methodologies, datasets, evaluation metrics, and applications. *Progress in Artificial Intelligence*, *12*(1), 1–32. https://doi.org/10.1007/s13748-023-00295-9

*Mulla, N., & Gharpure, P. (2023b). Leveraging well-formedness and cognitive level classifiers for automatic question generation on Java technical passages using T5 transformer. *International Journal of Information Technology*, *15*(4), 1961–1973. https://doi.org/10.1007/s41870-023-01262-2

*Muñoz, R. C., Blanco, B. B., Campo-Ávila, J. D., & Rodriguez, J. L. T. (2024). TEaching compilers: Automatic question generation and intelligent assessment of grammars' parsing. *IEEE Transactions on Learning Technologies*, *17*, 1734–1744. https://doi.org/10.1109/TLT.2024.3405565

*Murad, D. F., Wijanarko, B. D., Leandros, R., & Murad, S. A. (2023). Long short-term memory-based encoder-decoder with attention mechanism model for question generation. *ICIC International, 17*(9), 1067–1074. https://doi.org/10.24507/icicel.17.09.1067

*Murakhovs'ka, L., Wu, C.-S., Laban, P., Niu, T., Liu, W., & Xiong, C. (2022). MixQG: Neural question generation with mixed answer types. In *Findings of the 2022 Annual Conference of the North Americal Chapter of the Association for Computational Linguistics* (pp. 1486–1497). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-naacl.111

*Nikolovski, V., Trajanov, D., & Chorbev, I. (2025). Advancing AI in higher education: a comparative study of large language model-based agents for exam question generation, improvement, and evaluation. *Algorithms*, *18*(3), Article 144. https://doi.org/10.3390/a18030144

*Nio, L., & Murakami, K. (2018). Intelligence is asking the right question: a study on Japanese question generation. In Y. Stylianou, V. Karkaletsis, & S. Bangalore (Chairs), *Proceedings of the 2018 IEEE Spoken Language Technology Workshop* (pp. 771–778). IEEE. https://doi.org/10.1109/SLT.2018.8639526

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

*Pandraju, S., & Mahalingam, S. G. (2021). Answer-aware question generation from tabular and textual data using T5. *International Journal of Emerging Technologies in Learning*, *16*(18), 256–267. https://doi.org/10.3991/ijet.v16i18.25121

*Patil, P. M., Bhavsar, R. P., & Pawar, B. V. (2024). Bloom's taxonomy based automatic Marathi question generation. *International Journal of Advanced Technology and Engineering Exploration*, *11*(113), 575–603. https://doi.org/10.19101/IJATEE.2023.10102163

*Patil, S., Chavan, L., Mukane, J., Vora, D., & Chitre, V. (2022). State-of-the-art approach to e-learning with cutting edge NLP transformers: Implementing text summarization, question and distractor generation, question answering. *International Journal of Advanced Computer Science and Applications*, *13*(1), 445–453. https://doi.org/10.14569/IJACSA.2022.0130155

*Patro, B. N., Chauhan, D., Kurmi, V. K., & Namboodiri, V. P. (2020). Revisiting paraphrase question generator using pairwise discriminator. *Neurocomputing, 420*, 149-161. https://doi.org/10.1016/j.neucom.2020.08.022

Prieto-Andreu, J. M., & Labisa-Palmeira, A. (2024). Quick review of pedagogical experiences using GPT-3 in education. *Journal of Technology and Science Education*, *14*(2), 633–647. https://doi.org/10.3926/jotse.2111

*Qin, C., Zhu, H., Shen, D., Sun, Y., Yao, K., Wang, P., & Xiong, H. (2024). Automatic skill-oriented question generation and recommendation for intelligent job interviews. *ACM Transactions on Information Systems*, *42*(1), 1–32. https://doi.org/10.1145/3604552

*Rao, P. S. B., Agnihotri, M., & Babu Jayagopi, D. (2021). Improving asynchronous interview interaction with follow-up question generation. *International Journal of Interactive Multimedia and Artificial Intelligence*, *6*(5), 79–89. https://doi.org/10.9781/ijimai.2021.02.010

*Reid, M., French, M., Andreopoulos, S., Wong, C., & Kee, N. (2025). AI-generated multiple-choice questions in health science education: Stakeholder perspectives and implementation considerations. *Current Research in Physiology*, *8*, Article 100160. https://doi.org/10.1016/j.crphys.2025.100160

*Riza, L. S., Pertiwi, A. D., Rahman, E. F., Munir, M., & Abdullah, C. U. (2019). Question generator system of sentence completion in TOEFL using NLP and K-Nearest Neighbor. *Indonesian Journal of Science and Technology*, *4*(2), 294–311. https://doi.org/10.17509/ijost.v4i2.18202

*Roy, K., Gaur, M., Soltani, M., Rawte, V., Kalyan, A., & Sheth, A. (2023). ProKnow: Process knowledge for safety constrained and explainable question generation for mental health diagnostic assistance. *Frontiers in Big Data*, *5*, Article 1056728. https://doi.org/10.3389/fdata.2022.1056728

*Ruma, J. F., Mayeesha, T. T., & Rahman, R. M. (2023). Transformer based answer-aware Bengali question generation. *International Journal of Cognitive Computing in Engineering*, *4*, 314–326. https://doi.org/10.1016/j.ijcce.2023.09.003

*Sarrouti, M., Ben Abacha, A., & Demner-Fushman, D. (2021). Goal-driven visual question generation from radiology images. *Information*, *12*(8), Article 334. https://doi.org/10.3390/info12080334

*Schroeder, K. T., Hubertz, M., Van Campenhout, R., & Johnson, B. G. (2022). Teaching and learning with AI-generated courseware: Lessons from the classroom. *Online Learning*, *26*(3), 73–87. https://doi.org/10.24059/olj.v26i3.3370

*Sewunetie, W. T., & Kovács, L. (2022). Comparison of template-based and multilayer perceptron-based approach for automatic question generation system. *Indonesian Journal of Electrical Engineering and Computer Science*, *28*(3), 1738–1748. https://doi.org/10.11591/ijeecs.v28.i3.pp1738-1748

*Sewunetie, W. T., & Kovacs, L. (2024). Automatic question generation using extended dependency parsing. *Indonesian Journal of Electrical Engineering and Computer Science*, *33*(2), 1108–1115. https://doi.org/10.11591/ijeecs.v33.i2.pp1108-1115

*Shan, J., Nishihara, Y., Maeda, A., & Yamanishi, R. (2022). Question generation for reading comprehension test complying with types of question. *Journal of Information Science and Engineering*, *38*(3). https://doi.org/10.6688/JISE.202205_38(3).0005

*Shao, T., Cai, F., Chen, W., & Chen, H. (2022). Self-supervised clarification question generation for ambiguous multi-turn conversation. *Information Sciences*, *587*, 626–641. https://doi.org/10.1016/j.ins.2021.12.040

*Shen, K., Wu, L., Tang, S., Xu, F., Long, B., Zhuang, Y., & Pei, J. (2024). Ask questions with double hints: Visual question generation with answer-awareness and region-reference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(12), 9648–9660. https://doi.org/10.1109/TPAMI.2024.3425222

*Shoaib, M., Husnain, G., Sayed, N., Yasin Ghadi, Y., Alajmi, M., & Qahmash, A. (2025). Automated generation of multiple-choice questions for computer science education using conditional generative adversarial networks. *IEEE Access*, *13*, 16697–16715. https://doi.org/10.1109/ACCESS.2025.3530474

*Shwe, L. L., Matayong, S., & Witosurapot, S. (2025). Enabling cognitive and unified similarity-based difficulty ranking mechanisms for AQG on multimedia content. *Expert Systems with Applications*, *277*, Article 127244. https://doi.org/10.1016/j.eswa.2025.127244

*Steuer, T., Filighera, A., & Tregel, T. (2022). Investigating Educational and noneducational answer selection for educational question generation. *IEEE Access*, *10*, 63522–63531. https://doi.org/10.1109/ACCESS.2022.3180838

*Suhartono, D., Majiid, M. R. N., & Fredyan, R. (2024). Towards automatic question generation using pre-trained model in academic field for Bahasa Indonesia. *Education and Information Technologies*, *29*(16), 21295–21330. https://doi.org/10.1007/s10639-024-12717-9

*Sun, Y., Chen, C., Chen, A., & Zhao, X. (2021). Tibetan question generation based on sequence to sequence model. *Computers, Materials & Continua*, *68*(3), 3203–3213. https://doi.org/10.32604/cmc.2021.016517

*Sun, Y., Tang, D., Duan, N., Qin, T., Liu, S., Yan, Z., Zhou, M., Lv, Y., Yin, W., Feng, X., Qin, B., & Liu, T. (2020). Joint learning of question answering and question generation. *IEEE Transactions on Knowledge and Data Engineering*, *32*(5), 971–982. https://doi.org/10.1109/TKDE.2019.2897773

*Susanti, Y., Tokunaga, T., & Nishikawa, H. (2020). Integrating automatic question generation with computerised adaptive test. *Research and Practice in Technology Enhanced Learning*, *15*, Article 9. https://doi.org/10.1186/s41039-020-00132-w

Tao, L., Deng, H., & Song, Y. (2025). Generative artificial intelligence in education: A topic-based bibliometric analysis. *Educational Technology & Society*, *28*(2), 327–347. https://doi.org/10.30191/ETS.202504_28(2).TP05

*Terao, K., Tamaki, T., Raytchev, B., Kaneda, K., & Satoh, S. (2020). Rephrasing visual questions by specifying the entropy of the answer distribution. *IEICE Transactions on Information and Systems*, *E103.D*(11), 2362–2370. https://doi.org/10.1587/transinf.2020EDP7089

*Tiwari, A., Paul, A., & Suganya, P. (2019). A Smart bot as an interactive medical assistant with voice-based system using natural language processing. *International Journal of Engineering and Advanced Technology*, *8*(5s3), 164–166. https://doi.org/10.35940/ijeat.E1039.0785S319

*Tomikawa, Y., Suzuki, A., & Uto, M. (2024). Adaptive question–answer generation with difficulty control using item response theory and pretrained transformer models. *IEEE Transactions on Learning Technologies*, *17*, 2186–2198. https://doi.org/10.1109/TLT.2024.3491801

*Uehara, K., & Harada, T. (2024). Learning by asking questions for knowledge-based novel object recognition. *International Journal of Computer Vision*, *132*(6), 2290–2309. https://doi.org/10.1007/s11263-023-01976-7

*Vasić, D., Žitko, B., & Ljubić, H. (2021). Automatic question generation using semantic role labeling for morphologically rich languages. *Tehnicki Vjesnik - Technical Gazette*, *28*(3). https://doi.org/10.17559/TV-20200402175619

*Vincentio, K., & Suhartono, D. (2022). Automatic question generation monolingual multilingual pre-trained models using RNN and transformer in low resource Indonesian language. *Informatica*, *46*(7). https://doi.org/10.31449/inf.v46i7.4236

*Vyas, N., Kothari, H., Jain, A., & Joshi, A. R. (2022). Automated question and test-paper generation system. *International Journal of Computer Aided Engineering and Technology*, *16*(3), 362–378. https://doi.org/10.1504/ijcaet.2022.10046066

*Wang, L., Song, R., Guo, W., & Yang, H. (2025). Exploring prompt pattern for generative artificial intelligence in automatic question generation. *Interactive Learning Environments*, *33*(3), 2559–2584. https://doi.org/10.1080/10494820.2024.2412082

*Wang, Y., Rong, W., Zhang, J., Zhou, S., & Xiong, Z. (2020). Multi-turn dialogue-oriented pretrained question generation model. *Complex & Intelligent Systems*, *6*(3), 493–505. https://doi.org/10.1007/s40747-020-00147-2

*Wang, Y., Song, K., Bing, L., & Liu, X. (2021). Harvest shopping advice: Neural question generation from multiple information sources in e-commerce. *Neurocomputing*, *433*, 252–262. https://doi.org/10.1016/j.neucom.2020.12.013

*Weinzierl, M. A., & Harabagiu, S. M. (2023). Epidemic question answering: Question generation and entailment for answer nugget discovery. *Journal of the American Medical Informatics Association*, *30*(2), 329–339. https://doi.org/10.1093/jamia/ocac222

*Wen, Z., & Chu, S. K. W. (2025). Using generative AI for reading question creation based on PIRLS 2011 framework. *Cogent Education*, *12*(1), 2458653. https://doi.org/10.1080/2331186X.2025.2458653

*Willert, N., & Thiemann, J. (2024). TEmplate-based generator for single-choice questions. *Technology, Knowledge and Learning*, *29*(1), 355–370. https://doi.org/10.1007/s10758-023-09659-5

*Xie, J., Chen, J., Fang, W., Cai, Y., & Li, Q. (2023). Visual question generation for explicit questioning purposes based on target objects. *Neural Networks*, *167*, 638–647. https://doi.org/10.1016/j.neunet.2023.08.007

*Xie, J., Chen, J., Liu, Z., Cai, Y., Huang, Q., & Li, Q. (2024). VIdeo question generation for dynamic changes. *IEEE Transactions on Circuits and Systems for Video Technology*, *34*(9), 8710–8721. https://doi.org/10.1109/TCSVT.2024.3391415

*Xie, J., Fang, W., Cai, Y., Huang, Q., & Li, Q. (2022). Knowledge-based visual question generation. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(11), 7547–7558. https://doi.org/10.1109/TCSVT.2022.3189242

*Xie, J., Fang, W., Huang, Q., Cai, Y., & Wang, T. (2023). Enhancing paraphrase question generation with prior knowledge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *31*, 1464–1475. https://doi.org/10.1109/TASLP.2022.3221045

*Xie, J., Peng, N., Cai, Y., Wang, T., & Huang, Q. (2022). Diverse distractor generation for constructing high-quality multiple choice questions. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 280–291. https://doi.org/10.1109/TASLP.2021.3138706

*Xie, J., Zheng, J., Fang, W., Cai, Y., & Li, Q. (2025). Explicitly diverse visual question generation. *Neural Networks*, *184*, Article 107002. https://doi.org/10.1016/j.neunet.2024.107002

*Xu, X., Wang, T., Yang, Y., Hanjalic, A., & Shen, H. T. (2021). Radial graph convolutional network for visual question generation. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(4), 1654–1667. https://doi.org/10.1109/TNNLS.2020.2986029

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112. https://doi.org/10.1111/bjet.13370

*Yu, J., Su, Q., Quan, X., & Yin, J. (2021). Multi-hop reasoning question generation and its application. *IEEE Transactions on Knowledge and Data Engineering*, *35*(1), 725–740. https://doi.org/10.1109/TKDE.2021.3073227

Zhang, R., Guo, J., Chen, L., Fan, Y., & Cheng, X. (2022). a. *ACM Transactions on Information Systems*, *40*(1), 1–43. https://doi.org/10.1145/3468889

*Zeng, H., Wei, B., & Liu, J. (2024). RTRL: Relation-aware transformer with reinforcement learning for deep question generation. *Knowledge-Based Systems*, *300*, Article 112120. https://doi.org/10.1016/j.knosys.2024.112120

*Zeng, H., Zhi, Z., Liu, J., & Wei, B. (2021). Improving paragraph-level question generation with extended answer network and uncertainty-aware beam search. *Information Sciences*, *571*, 50–64. https://doi.org/10.1016/j.ins.2021.04.026

*Zhao, R., Tang, J., Zeng, W., Guo, Y., & Zhao, X. (2025). Towards human-like questioning: Knowledge base question generation with bias-corrected reinforcement learning from human feedback. *Information Processing & Management*, *62*(3), Article 104044. https://doi.org/10.1016/j.ipm.2024.104044

*Zheng, H.-T., Han, J., Chen, J., & Sangaiah, A. (2018). A novel framework for automatic Chinese question generation based on multi-feature neural network model. *Computer Science and Information Systems*, *15*(3), 487–499. https://doi.org/10.2298/CSIS171121018Z

*Zhu, H., Togo, R., Ogawa, T., & Haseyama, M. (2023). Diversity learning based on multi-latent space for medical image visual question generation. *Sensors*, *23*(3), Article 1057. https://doi.org/10.3390/s23031057

*Zhuge, Q., Wang, H., & Chen, X. (2025). TwinStar: A novel design for enhanced test question generation using dual-LLM engine. *Applied Sciences*, *15*(6), Article 3055. https://doi.org/10.3390/app15063055

Corresponding author: Supriyadi Supriyadi, 33222001@std.stei.itb.ac.id

Copyright: Articles published in the *Australasian Journal of Educational Technology* (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence (CC BY-NC-ND 4.0). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.

Please cite as: Supriyadi, S., Lubis, F. F., & Rosmansyah, Y. (2026). Automatic question generation (AQG): A systematic review of methods, applications, challenges and implications in educational technology. *Australasian Journal of Educational Technology*, *42*(5), 162–182. https://doi.org/10.14742/ajet.11445