Journal of Multidisciplinary and Translational Research (JMTR)

journal homepage: https://journals.kln.ac.lk/jmtr/

Data scarcity in adaptive learning: Challenges and opportunities with transfer learning and knowledge tracing

Mithara Fernandopulle[1], Wijendra Gunathilake[1]*, Malshani Hettiarachchi[2] and Jayani Rajakaruna[1]

[1]*University of Sri Jayawardenapura*
[2]*University of Moratuwa*

Abstract

Adaptive learning systems personalize instruction by tracking what each student knows and adjusting content to match their current knowledge state. Knowledge Tracing (KT) is the main technique used for this purpose. It has evolved from probabilistic models to deep learning architectures that use attention mechanisms to interpret a student's answer history and predict future performance. These models perform well only after training on large volumes of interaction data. A platform built for a new course, subject, student group, or national curriculum usually starts with little or no recorded history of student attempts. This is called the cold start problem. This review brings together recent work on knowledge tracing, transfer learning, and synthetic data generation to examine how these methods address data scarcity in adaptive learning. The review followed a systematic search and screening process across five academic databases, applied explicit inclusion and exclusion criteria, and used a structured quality appraisal completed independently by three reviewers. Seventy-nine records were identified; twenty-six were retained for the final synthesis (N = 26). The findings show that transfer learning techniques, particularly domain adaptation and cross-disciplinary knowledge transfer, can reduce the local data needed to reach acceptable prediction accuracy. However, none of the reviewed methods eliminate the cold start penalty; they only reduce it. The review also finds that most published work comes from a small number of research groups in the United States and China. Of the 26 included studies, 19 (73%) drew benchmark data or model architecture from these countries, leaving a gap for curricula, subjects, and languages elsewhere. Synthetic and simulated interaction data offer a complementary route: a new platform can begin with a working model before real student data accumulates. This research paper closes with guidance for researchers and developers building adaptive learning tools in data-scarce settings and directions for future study.

Keywords: Adaptive learning, Cold-start problem, Data scarcity, Knowledge tracing, Transfer learning

Article info

Article history:
Received 30th May 2026
Received in revised form 12th June 2026
Accepted 25th July 2026
Available online 25th August 2026

ISSN (E-Copy): ISSN 3051-5262
ISSN (Hard copy): ISSN 3051-5602
Doi:
ORCID iD: https://orcid.org/0000-0003-1762-930X
*Corresponding author:
E-mail address: wijendrapg@sjp.ac.lk
(Wijendra Gunathilake)

Introduction

Personalized education has been a continued goal of computer-assisted learning. If a system can infer what a student already understands and where they still struggle, it can offer the right material at the right time rather than moving the whole class through one fixed sequence. This can speed up learning and avoid wasted time on concepts already known. Adaptive learning platforms put this idea into practice, and the most commonly used technique for estimating understanding is knowledge tracing (KT), which predicts whether a student will answer the next question correctly based on past responses, treating that prediction as a working estimate of mastery of the underlying skill.

Knowledge tracing began with Bayesian Knowledge Tracing (BKT), a probabilistic model that updates a single mastery estimate for each skill after every attempt (Corbett & Anderson, 1994). This approach is simple and easy to interpret, but it treats each skill in isolation and needs a reasonable number of attempts before its estimates stabilize. Deep Knowledge Tracing (DKT) replaced this with a recurrent neural network that reads a student's full sequence of attempts and learns patterns across skills rather than treating them separately (Piech et al., 2015). Attention-based models followed, including the Self-Attentive model for Knowledge Tracing (SAKT) and Context-Aware Attentive Knowledge Tracing (AKT), which compute relevance weights between past and current questions rather than relying on a single recurrent state (Ghosh et al., 2020; Pandey & Karypis, 2019). Each step also demanded more training data for its larger parameter set, a pattern documented across many model variants in a recent survey (Shen et al., 2024).

This creates a concern at the center of adaptive learning. The models that perform best need the most data. Yet the platforms that most need personalization are often the ones with the least data available to train on. A new course, school, subject area, or national curriculum typically begins with no recorded student interactions. Even an established platform faces a smaller version of the same problem whenever it adds a new question or enrolls a new student, since no history yet exists for that question or student. This is widely known as the cold start problem. Recent experimental work confirms that every major model architecture shows a measurable drop in accuracy when tested on genuinely new students, compared with held-out interactions from students the model has already partly seen (Bhattacharjee & Wayllace, 2025).

Data scarcity is not unique to education. It is one of the most studied problems in deep learning, and the wider machine learning community has built a substantial set of responses. These include data augmentation, synthetic sample generation, and transfer learning from a related source task or domain (Alzubaidi et al., 2023; Bansal et al., 2022). Transfer learning asks whether a model trained on one task, where data is plentiful, can be adapted to a related task where data is limited, instead of training a new model from scratch (Farahani et al., 2021; Zhuang et al., 2021;). Applied to knowledge tracing, this raises a practical question. Can a model trained on a large public dataset from one subject, country, or platform be adapted to work reasonably well where almost no local data exists?

A small but growing body of research answers this question directly. These studies propose methods that move knowledge tracing models across subjects, courses, and student populations using only limited target data (Cheng et al., 2020; Cheng et al., 2022; Deng et al., 2025). Alongside

this, a separate line of work generates synthetic learner interactions, either through simulation or through generative models such as Generative Adversarial Networks (GANs) and large language models (LLMs), to give a new platform a working dataset before real students arrive (Zhang et al., 2024, Zhang et al., 2025). These two strands are often discussed separately, even though they address the same problem and can reasonably combine into a single pipeline.

This review brings the two strands together and asks two research questions. RQ1: What challenges does data scarcity create for knowledge tracing models, and how have researchers characterized these challenges across model families? RQ2: What opportunities do transfer learning and synthetic data generation offer, and what gaps remain? The review pays particular attention to settings where data scarcity is not a temporary hurdle that disappears once enough students have used a system, but a lasting feature of the environment. National curricula outside the well-resourced settings that produced most existing benchmark datasets are one example. Building a working personalization engine for these settings depends on exactly the techniques surveyed here, since waiting years to accumulate a large local dataset is rarely realistic for a small institution or a new public education initiative.

Methodology

This review follows a hybrid narrative-structured approach. A structured search defined the scope and selection criteria for the literature, and a narrative synthesis then interpreted and connected the selected studies, since the topic spans several distinct subfields, including educational data mining, transfer learning, and data augmentation, rarely reviewed together in a single systematic protocol. The identification and screening stages followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guideline, which sets a standard structure for recording how many records are identified, screened, excluded, and retained, so another researcher could repeat the search and reach a comparable set of studies (Page et al., 2021). PRISMA does not require a full meta-analysis; here, it keeps the search and screening stages transparent.

The search was carried out between 15 January 2026 and 10 February 2026 across five academic databases and search engines: Google Scholar, Semantic Scholar, the ACM Digital Library, arXiv, and SpringerLink. Search terms combined a core concept from knowledge tracing with a term describing the data problem or solution. The full Boolean search string, adapted slightly for each database's syntax, was:

("knowledge tracing" OR "student modeling" OR "student modelling") AND ("cold-start" OR "data scarcity" OR "transfer learning" OR "domain adaptation" OR "synthetic data")

Searches were not restricted to a fixed publication window, except that a study had to postdate the original Bayesian Knowledge Tracing paper, the start of knowledge tracing as a distinct research area. Screening was conducted by all four authors, with disagreements resolved by discussion. Table 1 sets out the inclusion and exclusion criteria applied during screening.

Table 1: Inclusion and exclusion criteria

Criterion	Inclusion	Exclusion
Topic	Knowledge tracing, transfer learning, domain adaptation, or data augmentation applied to student or learner performance data	General artificial intelligence in education studies with no technical treatment of learner modeling or data scarcity
Application context	Intelligent tutoring systems, adaptive quiz platforms, Massive Open Online Courses (MOOCs), or any technology-assisted learning environment	Studies unrelated to technology-assisted or adaptive learning
Date	1994 to 2026	Before 1994
Source type	Peer-reviewed journal articles, peer-reviewed conference proceedings, and clearly identified preprints from a recognized repository such as arXiv	Blog posts, marketing material, and opinion pieces without a technical method or result
Language	English	Other languages

Titles and abstracts of candidate records were screened against these criteria, with full text checked when the abstract lacked enough detail to decide. No formal quality-appraisal instrument (e.g., GRADE, Newcastle–Ottawa) was applied, consistent with this review's narrative rather than systematic-review design.

Peer-reviewed sources were included on the basis of publication venue; preprints were included only where the methodology and results were reported with sufficient technical detail to be independently assessed by the review team, and where the work was subsequently cited in, or closely aligned with, peer-reviewed literature in the same area. Studies were retained if they reported a knowledge tracing model; a transfer learning or domain adaptation method applied to student data; a data augmentation or synthetic data generation method aimed at learner performance data; or a survey-level treatment of data scarcity informing the review's general framing.

Eighty-two candidate records were identified across the five databases. After removing 12 duplicates, 70 records were screened by title and abstract. 42 records were excluded at this stage (mostly opinion pieces, marketing material, and papers mentioning knowledge tracing only in passing), leaving 28 records for full-text assessment. 2 were further excluded at full-text review for falling outside the criteria in Table 1, and the review retained twenty-six sources for direct citation (N = 26). Figure 1 presents the PRISMA 2020 flow diagram for the identification, screening, eligibility assessment, and inclusion of studies.

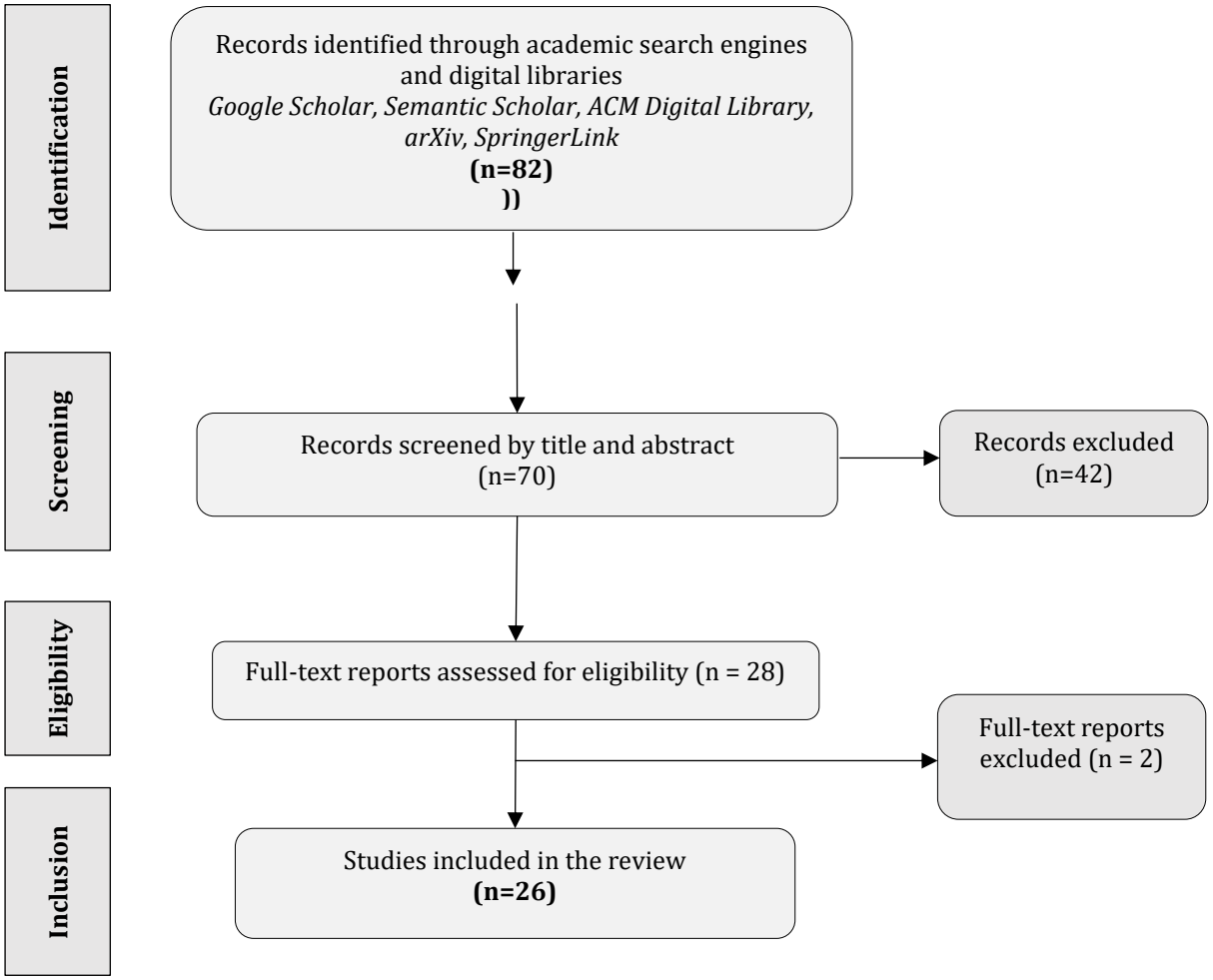

Figure 1. PRISMA 2020 flow diagram showing the identification, screening, eligibility assessment, and inclusion of studies in the review

The narrative synthesis that follows organizes these sources thematically rather than chronologically, since the review aims to connect a technical problem, data scarcity, with the range of proposed solutions, rather than trace a strict publication timeline. Included studies were grouped thematically according to the mechanism they used to address data scarcity: cold-start modelling, transfer learning, synthetic data generation, and low-resource/geographic context, with each theme synthesized narratively by identifying recurring findings, points of disagreement, and gaps across the studies within it.

Results

The results are organized around the two questions set out in the introduction. The first two themes answer the first research question, on the challenges data scarcity creates: the data scarcity problem in knowledge tracing, and the evolution of knowledge tracing models alongside their data requirements. The next two themes answer the second research question, on opportunities and remaining gaps: transfer learning and data augmentation with synthetic data generation. A fifth theme, data scarcity in low-resource educational contexts, draws on both questions, since it shows where the challenge is most severe and the opportunities least tested.

The data scarcity problem in knowledge tracing

Responding to the first research question, this theme sets out what data scarcity looks like inside a knowledge tracing system. Knowledge tracing systems face data scarcity in multiple forms, and the literature distinguishes among them, even though "cold-start" is often used loosely to cover them all (Liu et al., 2021). The first and most familiar form is a student cold-start, where a new student has answered too few questions for the model to say much about their knowledge state. The second is the item or question cold-start, where not enough students have attempted a new question for the model to learn how it relates to existing skills. The third, less discussed but increasingly relevant, is platform or domain cold-start, where an entire course, subject, or institution is new with no historical data to train on (Abdelrahman et al., 2023).

A controlled experimental study compared DKT, Dynamic Key-Value Memory Networks (DKVMN), and SAKT under a genuine cold start protocol. The study trained each model only on past students and tested it only on entirely new students. All three model families showed a clear early accuracy drop, with performance recovering gradually as a new student answered more questions (Bhattacharjee & Wayllace, 2025). The attention-based model reached higher initial accuracy than the recurrent and memory-based alternatives. Even so, it still fell well short of its later performance once the student had built a longer history. Newer architectures reduce the cold start penalty somewhat, but none of the published model families remove it.

Platform-level data scarcity is harder to study directly, since it requires comparing a new deployment against an established one. The broader deep learning literature offers a useful frame for this problem. Alzubaidi et al. (2023) and Bansal et al. (2022) describe data scarcity as a structural property of an application area rather than a temporary inconvenience. It arises when labelled examples are costly to collect, constrained by privacy rules, or simply not yet generated because the system is new. Education fits this pattern closely. Collecting a useful volume of labelled student data takes a term or a year of real use, time that a new platform has not yet had.

Evolution of knowledge tracing models and their data requirements

The history of knowledge tracing models shows a steady trade-off between expressive power and data appetite, summarised in Table 2. BKT represents a student's mastery of each skill as a single hidden binary state, updated using four parameters per skill: an initial mastery probability, a learning rate, a slip probability, and a guess probability (Corbett & Anderson, 1994). Because each skill is modelled independently with only a handful of parameters, this approach produces reasonable estimates from a modest number of attempts. It cannot share information across related skills, though, and it struggles with more complex patterns.

The DKT model moved away from this skill-by-skill structure. It processes a student's entire interaction sequence through a recurrent neural network and learns relationships between skills directly from data, rather than assuming skills are independent (Piech et al., 2015). This flexibility came at a cost. A recurrent network of this kind has far more parameters than a Bayesian model, shared across all students and skills, so it needs a larger and more varied training set before its internal representations become reliable. A direct empirical comparison across nine real-world datasets supports this trade-off: logistic regression and BKT stayed competitive on small and

moderate-sized datasets, while DKT only pulled ahead once a dataset grew large enough to make use of its extra parameters (Gervet et al., 2020).

Attention-based models kept the idea of learning shared representations from data, but changed how the model decides which past interactions matter for a current prediction. The SAKT model computes attention weights between a student's past questions and the next question, letting the model focus on the most relevant history instead of treating recent and distant attempts equally (Pandey & Karypis, 2019). AKT adds a monotonic decay function and a Rasch model-based embedding, so the model can account for question difficulty and forgetting (Ghosh et al., 2020). This added structure improved accuracy over simpler attention models without a large rise in parameter count. Careful architectural choices can partly offset the trend toward larger data requirements in deeper models, though they do not eliminate it.

Table 2 sets out these four model families side by side, along with their core mechanisms and how each tends to behave with limited data, drawing on the sources discussed above.

Table 2: Knowledge tracing model families and their data requirements

Model family	Core mechanism	Typical data demand	Note on data scarcity
Bayesian Knowledge Tracing (Corbett & Anderson, 1994; Shen et al., 2024)	Per skill probabilistic update with four parameters per skill	Low to moderate	Settles quickly per skill, but cannot share information across related skills
Deep Knowledge Tracing (Piech et al., 2015; Gervet et al., 2020)	A recurrent neural network reads the full interaction sequence	High	Learns cross-skill patterns but needs a large and varied training set
Self-Attentive Knowledge Tracing (Pandey & Karypis, 2019; Shen et al., 2024)	Self-attention over past questions and responses	Moderate to high	Performs reasonably on shorter histories but still needs a broad training set
Context-Aware Attentive Knowledge Tracing (Ghosh et al., 2020; Shen et al., 2024)	Monotonic attention with Rasch model-based embeddings	Moderate to high	Added structure improves accuracy without a large rise in parameter count

Transfer learning approaches to data scarcity in knowledge tracing

Turning to the second research question, this theme examines the first of two major opportunities. Transfer learning offers a direct response to the trend described above. Rather than training a large model entirely from local data, it starts from a model already trained on a related source task, where data is abundant, and adapts only part of that model to a target task where data is scarce (Zhuang et al., 2021). Farahani et al. (2021) distinguish domain adaptation, where the source and target tasks are the same, but the data distributions differ, from broader

transfer learning, where the tasks themselves may differ, noting that most practical applications combine elements of both.

Within knowledge tracing specifically, an early formulation of this problem as a domain adaptation task set out two linked goals: building an accurate model for each domain, and transferring a well-performing model between domains when one lacks sufficient data (Cheng, et al., 2020). The same group later proposed AdaptKT, a three-stage method that selects training instances from the source domain resembling the target domain, reduces the statistical gap between the domains' feature distributions, and fine-tunes only the output layer on the limited target data, leaving lower layers unchanged (Cheng et al., 2022), reflecting a common deep learning finding: lower layers learn general-purpose representations while upper layers specialize, so freezing the lower layers during transfer preserves useful general knowledge while still letting the model adjust to a new domain.

More recent work has tackled harder versions of this problem. Deng et al. (2025) address cross-disciplinary cold-start, where source and target domains are entirely different subjects with little or no overlap in skills or vocabulary, by clustering student knowledge states from a pre-trained source model and routing target students through a mixture of expert networks guided by those clusters, with an adversarial component pulling together students with similar learning patterns while pushing apart dissimilar ones. Bai et al. (2025) address cold-start questions with a kernel bias and cone attention mechanism that gives the model a sensible default representation for a question lacking response data of its own, while a related route uses the structure of a concept graph instead, since a new question's position relative to known questions can supply a useful prior before any student has attempted it (Nakagawa et al., 2019). Earlier work by Zhao et al. (2020) tackled the same problem with an attentive neural Turing machine, combining external memory with attention so the model draws on patterns from other students even when a student's own history is short, matching what the literature calls few-shot learning, where a model performs well after seeing only a handful of examples for a new case (Song et al., 2023).

Taken together, these methods show that transfer learning for knowledge tracing has moved from a general statement of the problem toward increasingly specific architectures targeted at particular cold-start scenarios, whether involving a new student, a new question, or an entirely new subject area. What they share is the assumption that a model need not learn everything from the target domain's own limited data, since much of what makes a good knowledge tracing model, such as recognizing general patterns of practice, forgetting, and partial understanding, can be learned once from a data-rich source and reused.

Data augmentation and synthetic data generation as a complementary opportunity

Still under the second research question, this theme covers the opportunity of manufacturing more data rather than transferring a model. While transfer learning reduces how much target-domain data a model needs, a second, complementary line of work increases the effective amount of target-domain data available, either by simulating plausible student behavior or by generating synthetic interactions, an approach broader surveys treat as a standard companion to transfer learning rather than a substitute for it (Fayaz et al., 2024). This idea is not new in knowledge tracing: the original Deep Knowledge Tracing paper used a synthetic dataset from item response

theory to test the model before applying it to real student logs, simulating thousands of virtual students answering a fixed set of questions (Piech et al., 2015).

More recent approaches use generative models rather than a fixed statistical formula to produce synthetic data. Zhang et al. (2024, 2025) represent sparse learner performance data as a three-dimensional tensor across learners, questions, and attempts, then use tensor factorization with generative adversarial networks and large language model-based generation to fill in missing values and produce additional plausible interactions tailored to clusters of similar learning patterns. Their comparisons, repeated across two related studies, found the generative adversarial network produced more reliable synthetic data than the language model-based approach, a reminder that newer, general-purpose generative tools do not automatically outperform older, specialized ones for a structured task like this.

The broader literature on data scarcity in deep learning supports the same principle outside education. Bansal et al. (2022) and Alzubaidi et al. (2023) treat data augmentation and synthetic data generation as standard tools alongside transfer learning, noting that the choice between them depends on how much structure is already known about the target task. Where a model of how students learn already exists, such as an item response theory model or a curriculum dependency map, simulation can cheaply generate synthetic interaction data with full control over its properties; where less is known in advance, generative approaches trained on whatever limited real data exists may capture patterns a hand-built simulation would miss, at the cost of being harder to validate.

A notable limitation across the synthetic-data literature is that generated records are rarely validated against real student behavior beyond aggregate accuracy metrics, which risks models learning statistical artifacts of the generator rather than genuine learning patterns: a concern that current studies in this review do not adequately address.

Data scarcity in low-resource educational contexts

This final theme draws on both research questions, showing where the challenge is most acute and the opportunities least tested. Most of the knowledge tracing and transfer learning literature reviewed above is built and tested on a small number of public datasets, most prominently ASSISTments, a platform developed in the United States and used as the benchmark source domain in several transfer learning studies discussed earlier (Feng et al., 2009). This pattern is consistent across the review: the majority of the 26 studies cited here originate from research groups in the United States or China, and the benchmark datasets they rely on (e.g., ASSISTments) were likewise built in those settings.

A recent review of artificial intelligence in education across low and middle-income countries found that generative and adaptive tools in these settings are constrained not only by infrastructure such as electricity and internet access, but also by a more fundamental data gap, since the corpora and benchmark datasets underpinning most AIED research are drawn predominantly from a small number of countries and languages (Landa-Blanco, 2026). For knowledge tracing specifically, a platform built around a national mathematics curriculum outside that group cannot simply download a pre-trained model and expect it to transfer cleanly,

since the source domain may differ from the target in language, curriculum sequencing, and skill structure (National Institute of Education Sri Lanka, 2023).

Existing e-learning tools built for Sri Lankan school students illustrate this gap directly. A mobile application designed to motivate and assist primary school students tracks correctness and engagement and adjusts question difficulty based on a student's score, but does not estimate a hidden knowledge state for each concept the way knowledge tracing does (Silva et al., 2021), suggesting the techniques described here have not yet reached many locally built platforms. This is precisely the situation in which combining transfer learning and synthetic data generation, rather than either technique alone, is most useful. A pre-trained source model supplies general patterns of learning and forgetting that do not depend heavily on subject, while a curriculum-specific synthetic dataset, built by mapping the target curriculum's concepts to the source dataset's knowledge components, supplies the local structure a generic transfer step would miss. None of the studies reviewed here test this combination within an under-resourced deployment, pointing to a clear opportunity for future work rather than a settled finding.

In summary, the reviewed literature shows a consistent pattern: model accuracy and data appetite rise together, every architecture suffers a measurable cold-start penalty, and transfer learning together with synthetic data generation are the two most developed responses — though both remain tested almost exclusively on well-resourced, Western-built platforms.

Table 3: Transfer learning and data augmentation approaches addressing data scarcity in knowledge tracing

Approach	Method	Type of scarcity addressed
Domain adaptation for knowledge tracing (Cheng, et al., 2020)	Formal definition of cross-domain transfer for knowledge tracing	Domain-level scarcity, new subject or course
AdaptKT (Cheng et al., 2022)	Instance selection, distribution discrepancy minimization, output layer fine-tuning	Domain-level scarcity with some source-target overlap
Cross-disciplinary cold-start framework (Deng et al., 2025)	Mixture of experts guided by clustered knowledge states, adversarial feature alignment	Domain-level scarcity with little or no source-target overlap
Attentive neural Turing machine (Zhao et al., 2020; Song et al., 2023)	External memory with attention over prior learners, related to few-shot learning more broadly	Student-level cold-start
Kernel bias, cone attention, and graph-based representations (Bai et al., 2025; Nakagawa et al., 2019)	Default representation for new questions using attention bias or concept-graph position	Item or question-level cold-start

Discussion

Taken together, the findings tell a consistent story across two bodies of research usually discussed apart. Knowledge tracing work shows newer, more accurate models also need more data, and every architecture tested still loses accuracy on a genuine cold start. Transfer learning and data augmentation work show this loss can shrink, though never disappear, when a model reuses patterns from a richer source, when synthetic interactions fill in for missing real ones, or when both combine. For practice, the base model matters less than how it gets adapted; an attention-based model such as SAKT balances accuracy and data needs well but still benefits from a transfer step. The transfer technique should match the scarcity at hand, with instance selection and distribution alignment suiting some source-target overlap and mixture of expert or adversarial methods suiting almost none, while synthetic data works best as a bridge through the leanest early period.

These approaches also differ in cost and fragility. Instance-selection and distribution-alignment methods (e.g., AdaptKT) are comparatively lightweight and interpretable but assume meaningful overlap between source and target domains, an assumption that breaks down for cross-disciplinary transfer. Mixture-of-experts and adversarial approaches (e.g., Deng et al., 2025) handle domain mismatch better but need substantially more source data and computation to train reliably, and their added complexity makes them harder to diagnose when transfer fails. No single method dominates; the right choice depends on how much source-target overlap actually exists, which is rarely known in advance for a genuinely new deployment.

Two gaps stand out. First, almost no published work tests these methods across curricula and populations beyond a handful of well-documented, mostly Western platforms; since transfer learning relies on similarity between source and target domains, it remains an open question how these methods hold up once the target departs more sharply, for instance, through a different language of instruction or topic sequencing. Second, most studies report prediction accuracy against held out portions of an existing dataset rather than asking whether better prediction leads to better content sequencing or learning outcomes, the actual point of these systems. Work connecting model level metrics to usability and learning outcomes through testing with real students and teachers would strengthen the case for any particular method.

Conclusions

This review examined how data scarcity affects knowledge tracing, the core technique behind adaptive learning systems, and how transfer learning and synthetic data generation respond to that challenge. Knowledge tracing models have grown more accurate and more data hungry together, from simple per skill probability models through recurrent networks to attention based architectures, and every model family shows a measurable cold-start penalty with new students, new questions, or new subjects. Transfer learning methods built for knowledge tracing ease this penalty by reusing patterns from a related, data rich source, while synthetic data generation offers a complementary route to a working dataset before real interactions accumulate.

The review also found a clear gap. Almost three-quarters of the methods and benchmark datasets discussed here come from a small number of well-resourced platforms in the United States and

China, leaving open how well they perform across curricula, languages, and populations outside that group. In practice, this means institutions in under-resourced settings should not assume a published model will transfer out of the box; a practical starting point is pairing a pre-trained source model with a small, curriculum-specific synthetic dataset rather than either collecting years of local data or adopting a source model unmodified. This matters for educational equity, since institutions most likely to benefit from a working adaptive learning tool are often least represented in the data these methods are tested on. Future research should prioritize testing these methods on underrepresented curricula and languages, combined approaches pairing a pre-trained source model with curriculum-specific synthetic data, and evaluation frameworks linking prediction accuracy to usability and learning outcomes for the students and teachers who would use these systems.

Acknowledgements

We would like to express our sincere appreciation to the academic staff of the Faculty of Computing, University of Sri Jayewardenepura, for the opportunity and guidance throughout this study.

Declaration of Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest statement

The author declares no conflict of interest.

Data availability statement

This is a literature review and no new data were generated or analyzed for this study. All sources discussed are publicly available and are listed in the references.

Author Contribution

All authors accepted responsibility for the content of this paper, reviewed the findings, and approved the final version for submission. Conceptualization and study design: M.F.; literature review and original draft preparation: M.F.; methodology development: M.F., M.H.; analysis and interpretation of findings: M.F., J.R.; writing – review, editing and supervision: M.H., J.R., W.G.; project oversight and final approval: W.G.

References

Abdelrahman, G., Wang, Q., & Nunes, B. (2023). Knowledge tracing: A survey. *ACM Computing Surveys, 55*(11), Article 224, 1–37. https://doi.org/10.1145/3569576

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., & Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications. *Journal of Big Data, 10*, Article 46. https://doi.org/10.1186/s40537-023-00727-2

Bai, Y., Li, X., Liu, Z., Huang, Y., Guo, T., Hou, M., Xia, F., & Luo, W. (2025). csKT: Addressing cold-start problem in knowledge tracing via kernel bias and cone attention. *Expert Systems with Applications, 266*, 125988. https://doi.org/10.1016/j.eswa.2024.125988

Bansal, A., Sharma, R., & Kathuria, M. (2022). A systematic review on data scarcity problem in deep learning: Solution and applications. *ACM Computing Surveys, 54*(10s), 1–29. https://doi.org/10.1145/3502287

Bhattacharjee, I., & Wayllace, C. (2025). *Cold start problem: An experimental study of knowledge tracing models with new students*. arXiv. https://doi.org/10.48550/arXiv.2505.21517

Cheng, S., Liu, Q., & Chen, E. (2020). *Domain adaption for knowledge tracing*. arXiv. https://doi.org/10.48550/arXiv.2001.04841

Cheng, S., Liu, Q., Chen, E., Zhang, K., Huang, Z., Yin, Y., Huang, X., & Su, Y. (2022). AdaptKT: A domain adaptable method for knowledge tracing. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 123–131. https://doi.org/10.1145/3488560.3498379

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*(4), 253–278. https://doi.org/10.1007/BF01099821

Deng, Y., Guan, Z., He, M., Wang, X., Liu, J., & Li, Z. (2025). *Adaptive knowledge transfer for cross-disciplinary cold-start knowledge tracing*. arXiv. https://doi.org/10.48550/arXiv.2511.20009

Farahani, A., Voghoei, S., Rasheed, K., & Arabnia, H. R. (2021). A brief review of domain adaptation. In R. Stahlbock, G. M. Weiss, M. Abou-Nasr, C.-Y. Yang, H. R. Arabnia, & L. Deligiannidis (Eds.), *Advances in data science and information engineering* (pp. 877–894). Springer. https://doi.org/10.1007/978-3-030-71704-9_65

Fayaz, S., Shah, S. Z. A., Mohi ud din, N., Gul, N., & Assad, A. (2024). Advancements in data augmentation and transfer learning: A comprehensive survey to address data scarcity challenges. *Recent Advances in Computer Science and Communications, 17*(8), 14–35. https://doi.org/10.2174/0126662558286875231215054324

Feng, M., Heffernan, N., & Koedinger, K. R. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction, 19*(3), 243–266. https://doi.org/10.1007/s11257-009-9063-7

Gervet, T., Koedinger, K., Schneider, J., & Mitchell, T. (2020). When is deep learning the best approach to knowledge tracing? *Journal of Educational Data Mining, 12*(3), 31–54. https://doi.org/10.5281/zenodo.4143614

Ghosh, A., Heffernan, N., & Lan, A. S. (2020). Context-aware attentive knowledge tracing. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2330–2339). Association for Computing Machinery. https://doi.org/10.1145/3394486.3403282

Landa-Blanco, M. (2026). Artificial intelligence in education: Applications and limitations for teachers in low- and middle-income countries. *Frontiers in Education, 10*, Article 1681836. https://doi.org/10.3389/feduc.2025.1681836

Liu, Q., Shen, S., Huang, Z., Chen, E., & Zheng, Y. (2021). *A survey of knowledge tracing*. arXiv. https://doi.org/10.48550/arXiv.2105.15106

Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019). Graph-based knowledge tracing: Modeling student proficiency using graph neural network. In *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 156–163). Association for Computing Machinery. https://doi.org/10.1145/3350546.3352513

National Institute of Education Sri Lanka. (2023). Primary school curriculum framework, Grades 3–5. NIE Publications.

Pandey, S., & Karypis, G. (2019). A self-attentive model for knowledge tracing. In C. F. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), *Proceedings of the 12th International Conference on Educational Data Mining* (pp. 384–389). International Educational Data Mining Society. https://doi.org/10.48550/arXiv.1907.06837

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., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in neural information processing systems* (Vol. 28, pp. 505–513).

Shen, S., Liu, Q., Huang, Z., Zheng, Y., Yin, M., Wang, M., & Chen, E. (2024). A survey of knowledge tracing: Models, variants, and applications. *IEEE Transactions on Learning Technologies, 17*, 1898–1879. https://doi.org/10.1109/TLT.2024.3383325

Silva, P. H. D. D., Sudasinghe, S. A. V. D., Hansika, P. D. U., Gamage, M. P., & Gamage, M. P. A. W. (2021). AI base e-learning solution to motivate and assist primary school students. In *2021 3rd International Conference on Advancements in Computing (ICAC)* (pp. 294–299). IEEE. https://doi.org/10.1109/ICAC54203.2021.9671209

Song, Y., Wang, T., Cai, P., Mondal, S. K., & Sahoo, J. P. (2023). A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys, 55*(13s), 1–40. https://doi.org/10.1145/3582688

Zhang, L., Lin, J., Borchers, C., Cao, M., & Hu, X. (2024a). *3DG: A framework for using generative AI for handling sparse learner performance data from intelligent tutoring systems*. arXiv. https://doi.org/10.48550/arXiv.2402.01746

Zhang, L., Lin, J., Sabatini, J., Borchers, C., Weitekamp, D., Cao, M., Hollander, J., Hu, X., & Graesser, A. C. (2025). Data augmentation for sparse multidimensional learning performance data using generative AI. *IEEE Transactions on Learning Technologies, 18*, 145–164. https://doi.org/10.1109/TLT.2025.3526582

Zhao, J., Bhatt, S. P., Thille, C., Gattani, N., & Zimmaro, D. (2020). Cold start knowledge tracing with attentive neural Turing machine. In *Proceedings of the Seventh ACM Conference on Learning @ Scale* (pp. 333–336). Association for Computing Machinery. https://doi.org/10.1145/3386527.3406741

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE, 109*(1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555