How are LLMs being employed by primary school teachers in the UK and what are the barriers to their adoption in key areas of teaching and learning?

Introduction

Large Language Models (LLMs) - linguistic, data-trained machine learning models (Raff, Farris & Biderman, 2025) - have received a significant increase in publicity since ChatGPT-3.5 was made available to the wider public in November 2022 (Jauhiainen & Guerra, 2023). A subset of generative AI (GAI), LLMs process language to generate content (Toner, 2023) and are proving to be a highly accessible, powerful and flexible technology that has been rapidly adopted throughout the education sector (Furze et al., 2024).

This literature review aims to give an overview of how this technology has been utilised within education, beginning with how digital learning development has paved the way for its rapid adoption. The strengths and weaknesses of LLMs as educational tools will be outlined in relation to the ways in which it is reportedly being employed in classrooms around the world: as research assistants, plagiarism detectors, assessors, role-playing partners, and learning resource generators (Jeon & Lee, 2023; Gates, 2023; Pesovski et al., 2024; Zou et al., 2025; Jauhiainen & Guerra, 2023; Ng, 2024). These developments will then be viewed within the context of the UK Government's plans for the use of LLMs in education, alongside a brief look at pre-trained, LLM-based software currently being marketed to teachers. Finally, this paper will outline how the current study will contribute to current knowledge regarding the impact of LLMs on present and future teaching and learning practices.

The development of Digital Learning

Digital learning has seen a progression from the 1960s rule-based systems that were programmed on predetermined rules, to those of the 1990s and 2000s that utilised machine learning to respond to performance data in order to track performance over time (He, 2024; Alsafari et al, 2024). Deep learning systems of the 2010s then allowed for language generation, responsive educational applications with learning tools and automatically generated content that mirrored human tutoring (He, 2024). These however were still constrained by predetermined programming while development was resource-intensive (Alsafari et al, 2024). Modern LLM-based GAI systems have taken tailored learning further still by providing rapid, contextual responses that are able to tutor students, assess work, generate content, converse with and assist teachers, amongst its other uses (Jeon & Lee, 2023; Gates, 2023; Pesovski et al., 2024; Zou et al., 2025; Jauhiainen & Guerra, 2023; Ng, 2024). Additionally, the knowledge base of LLMs is far more up to date than older, non-AI systems, with vastly more scope for further expansion (Alsafari et al., 2024). Future developments in LLMs

for education appear to be heading towards much more emotionally aware systems that can be integrated into android-like hardware to provide more human-like assistance (Ng, 2024; Gates, 2023; Fragakis, Trichopoulos & Caridakis, 2025). However, as impressive as they seem, some view LLMs as advanced pattern-matchers that still require human intervention to craft purposeful educational experiences (McCaleb et al., 2025). Indeed, LLMs' output can be far from perfect, with issues such as erroneous, nonsense answers, bias, inaccuracies, and a lack of emotional intelligence limiting their effectiveness as educational tools (Bai et al., 2024; Ng, 2024).

LLMS as research assistants

LLMs can be useful research aids to both teachers and students, clarifying ideas and information (Jeon & Lee, 2023; Ng, 2024). When using ChatGPT for research, undergraduate students have been observed engaging in purposeful conversation in which the students were able to obtain answers to complex questions before prompting further detailed explanations from the LLM (Durgungoz & Kharrufa, 2025). Although viewed as a significantly useful tool, it has been recognised that there is a high need for further student training on the use of effective prompts to ensure high quality LLM content generation (Durgungoz & Kharrufa, 2025). Such interactions between student and LLM research assistant have been seen to increase the self-efficacy of students, with many continuing their use of the tool over extended periods of time (Durgungoz & Kharrufa, 2025). Free LLM models such as ChatGPT-3.5, however, have been found to be less accurate in answering questions and explaining concepts than more advanced, paid-for models such as ChatGPT-4 (Bai et al., 2024; Borges et al., 2024). Zhang et al. (2025) similarly found that pre-trained models enhanced its understanding of educational content considerably; such models offered more engaging, human-like conversation, offering pedagogically sound, context-rich tutoring capable of offering effective problem-solving guidance (Zhang et al., 2025). Additional uses of LLMs as a research aid include content summarising and literature searches, however erroneous outputs are again a commonly reported limitation in this context (Sohail & Zhang, 2025; Durgungoz & Kharrufa, 2025). While one study found that students became better at spotting the LLM's erroneous answers over time, they tended to rely on intuition rather than a set strategy employed to verify output (Durgungoz & Kharrufa, 2025). Although human training has been suggested as a way to address this limitation (Durgungoz & Kharrufa, 2025), pretrained LLMs also have the capability of identifying such content (Wang et al., 2024).

LLMS for plagiarism detection

In another instance of AI being employed to address the limitations of AI, paid-for, pre-trained LLMs are available to help educators identify signs of AI-generated content within assignments (Qorich & Ouazzani, 2025); GPT Zero, Content at Scale and

Turnitin use such systems within their software (Protheroe, 2025). Although some warn against the growing use of LLMs within assignments and view its use as a threat to academic integrity (Bektik et al., 2024; Borges et al., 2024; Sohail & Zhang, 2025), Furze et al. (2024) believe it should be accepted as a new and innovative learning tool, suggesting a five-point scale on which the levels of AI use can be detailed, clearly outlining the degree of AI assistance that is deemed to be acceptable within an assessed task. When trialled, this assessment system resulted in a significant reduction in AI-related academic misconduct, while improving student engagement and developing more advanced teaching practices (Furze et al., 2024). On this subject, Jeon & Lee (2023) similarly advocate for the development of students as active investigators who can utilise LLMs and formulate impactful questions so that LLMs become research partners rather than mere knowledge generators.

LLMS as assessors

LLMs such as ChatGPT have been shown to capably assess work - even in advanced writing assignments - against complex criteria (Zou et al., 2025). Accurate output however seems to be more consistent when assessing simpler answer forms, and LLMs have been seen to particularly struggle with more complex mathematics reasoning problems (Lee et al., 2024). Furthermore, although LLMs have been found to provide valuable feedback on essay-style assignments, repetitive language, inaccuracies and a lack of emotional awareness have been shown to limit their impact on subsequent student learning (De Wet, Da Silva & Bohnsack, 2025). Human input on assessment feedback is still sought by students, with one study reporting that students preferred a hybrid feedback option of human and LLM input as opposed to solely human or solely LLM feedback (De Wet, Da Silva & Bohnsack, 2025). Such collaboration with LLMs have been seen to significantly improve educational instructors' ability to provide more timely, personalised feedback (Xavier et al., 2025). LLMs have even been employed to create assessments, however they often lack originality and can be overly similar to existing material found online (Zou et al., 2025).

LLMS as role-playing partners

The use of LLMs as an educational role-playing partner within hypothetical scenarios has been used in modelling language interaction to pupils (Jeon & Lee, 2023). LLMs playing the role of a dialogue partner have been shown to significantly improve second language acquisition, providing increased opportunities for students to have one-on-one conversations in their developing language (Shan et al., 2025; Zheng et al., 2025). This has been shown to significantly increase oral proficiency, self-efficacy and motivation in foreign language communication, while reducing foreign language anxiety (Zheng et al., 2025). Despite these benefits, studies have reported limitations in non-English output due to the majority of LLMs having been trained on English training data (Bai et al., 2024; Lee, 2024).

LLMs as learning resource generators

The Department for Education (DfE) has recently invested in AI learning tool development that has led to the development of Aila – a free-too-use beta AI-powered lesson assistant that aids teachers in planning lesson content (DfE, 2025). LLMs have also been shown to be effective at generating student learning materials, with research reporting positive outcomes of reported enjoyment and engagement with learning materials provided by ChatGPT-3.5 (Jauhiainen & Guerra, 2023). A further study found that personalised learning materials, particularly those varied in content, resulted in increased study time among undergraduate students (Pesovski et al., 2024). Such learning materials were generated in the nuanced styles of characters from popular culture along with a traditional professor – a method that seemed to improve enjoyment and engagement (Pesovski et al., 2024). Indeed, personalised content generation is seen to be one of the areas where AI will have the biggest impact in education (Gates, 2023), with future incarnations of LLMs predicted to be able to find ways to offer deeper personalisation and nuance (Ng, 2024; Fragakis, Trichopoulos & Caridakis, 2025).

LLMs in UK schools

In July 2025, the UK Government signed a strategic partnership with OpenAI to advance progress towards embedding the use of ChatGPT in public sector services such as education (Department for Science, Innovation and Technology (DfSIT) & Kyle, 2025). The DfE (2025) acknowledges the importance of LLMs pre-trained on curriculum content in order to maximise output quality, however, despite laying out further intention to invest in developing AI-education tools, the DfE acknowledges its limitations and the need for schools to ensure that any use of AI adheres to guidelines around data protection, child safeguarding, and intellectual property (DfE, 2025). Indeed, Tasdelen & Bodemer (2025) emphasise the need for greater consistency in data processing transparency for LLMs used in the classroom. The relatively unproven impact of AI within education is also emphasised by the DfE, suggesting that much more development is needed to maximise the usefulness of this technology within education (DfE, 2025).

A growing number of paid-for LLM-based education software is currently being marketed to educators as time-saving devices. MagicSchool.ai and Learnt.ai for example can be used for the creation of multiple-choice questions, assessment questions, and feedback generation, while software such as Lesson Robot focusses specifically on lesson plan generation (Protheroe, 2025). More general LLM-based applications such as Teachmate and Diffit are capable of all aforementioned functions

as well as being able to differentiate content and generate presentations and personalised reports (TeachMateAI Ltd, 2025; Diffit, N.D.).

Conclusion

Reported use of LLMs in education include its employment as a research assistant, plagiarism detector, assessor, role-playing partner, and learning resource generator (Jeon & Lee, 2023; Gates, 2023; Pesovski et al., 2024; Zou et al., 2025; Jauhiainen & Guerra, 2023; Ng, 2024). Limitations of current LLMs, and possible barriers to widespread adoption of LLMs within the aforementioned roles, include nonsensical replies; inaccurate output; bias; a lack of emotional awareness in output; limited ability to process non-English input; use of repetitive language; unclear data processing techniques; output that is similar to pre-existing material; and a threat to academic integrity without increased assessment flexibility (Jeon & Lee, 2023; Furze et al., 2024; Gates, 2023; Pesovski et al., 2024; Zou et al., 2025; Jauhiainen & Guerra, 2023; Ng, 2024). While such limitations exist, versatility, availability, speed of response and scalability are key strengths, while LLMs themselves can be employed to combat a number of its limitations, such as plagiarism detectors and those trained to identify erroneous data output (Qorich & Ouazzani, 2025; Wang et al., 2024). Quality prompts from trained educators, combined with carefully pre-trained models can also significantly improve output quality (Tasdelen & Bodemer, 2025; Jeon & Lee, 2023). LLMs are often cited as being a supplementary tool, best used alongside human intervention and oversight to ensure quality control and much-needed emotional nuance within student interaction (De Wet, Da Silva & Bohnsack, 2025; McCaleb et al., 2025; Tasdelen & Bodemer, 2025; Jeon & Lee, 2023; Qorich & Ouazzani, 2025).

This study will seek to uncover patterns of LLM usage in the relatively neglected area of primary education (Lang et al., 2025). Within this study, an analysis of LLM usage will help to provide an overview of barriers to LLM adoption in key areas of primary teaching and learning, answering questions such as: ‘Are LLMs not used by some teachers due to the cost of the more accurate, pre-trained models?’; ‘Do erroneous or inaccurate LLM outputs affect teachers’ use of LLMs as research assistants or live teaching assistants?’; ‘Are free LLMs such as ChatGPT-3.5 trusted less than paid-for options?’; ‘Are LLM data processes trusted enough for the teacher to enable direct student-LLM interaction?’ The answers to such questions should go some way to informing much-needed teacher and student training in LLM-usage (Jeon & Lee, 2023; Durgungoz & Kharrufa, 2025) while shedding light on future areas of development for educational LLMs in the UK’s primary education setting.

References

Alsafari, B., Atwell, E., Walker, A., & Callaghan, M. (2024) ‘Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants’, *Natural Language Processing Journal*, 8, pp. 100101. Available at: https://doi.org/10.1016/j.nlp.2024.100101 (Accessed 25 August 2025).

Bai, Y., Li, J., Shen, J., & Zhao, L. (2024) ‘Investigating the Efficacy of ChatGPT-3.5 for Tutoring in Chinese Elementary Education Settings’, *IEEE Transactions on Learning Technologies*, 17, pp.2102–2117. Available at: https://doi.org/10.1109/TLT.2024.3464560 (Accessed 23 August 2025).

Bektik, D., Ullmann, T. D., Edwards, C., Herodotou, C., & Whitelock, D. (2024) ‘AI-Powered Curricula: Unpacking the Potential and Progress of Generative Technologies in Education’, *Ubiquity Proceedings*, 4(1), pp.38. Available at: https://doi.org/10.5334/uproc.160 (Accessed 21 August 2025).

Borges, B., Foroutan, N., Bayazit, D., Sotnikova, A., Montariol, S., Nazaretsky, T., Banaei, M., Sakhaeirad, A., Servant, P., Neshaei, S. P., Frej, J., Romanou, A., Weiss, G., Mamooler, S., Chen, Z., Fan, S., Gao, S., Ismayilzada, M., Paul, D., … Bosselut, A. (2024) ‘Could ChatGPT get an engineering degree? Evaluating higher education vulnerability to AI assistants’, *Proceedings of the National Academy of Sciences - PNAS*, *121*(49), pp.1-41. Available at: https://doi.org/10.1073/pnas.2414955121 (Accessed 21 August 2025).

Department for Education (DfE). (2025) *Policy paper Generative artificial intelligence (AI) in education*. Available at: https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education (Accessed 02 September 2025).

Department for Science, Innovation and Technology (DfSIT) & Kyle, P. (2025) *OpenAI to expand UK office and work with government departments to turbocharge the UK’s AI infrastructure and transform public services*. Available at:

https://www.gov.uk/government/news/openai-to-expand-uk-office-and-work-with-government-departments-to-turbocharge-the-uks-ai-infrastructure-and-transform-public-services#:~:text=the%20partnership%20arrangement.-,OpenAI%20and%20the%20UK%20government%20have%20today%20signed%20a%20new,use%20of%20the%20latest%20tech (Accessed 30 August 2025).

De Wet, M., Da Silva, M.O. and Bohnsack, R. (2025) ‘Can AI give good feedback on essay-type assignments? An explorative case study of LLMs in higher education’, *Innovations in Education and Teaching International*, pp.1-16. Available at: https://doi.org/10.1080/14703297.2025.2536609 (Accessed 02 September 2025).

Diffit. (N.D.) *Learning resources for all*. Available at: https://web.diffit.me/ (Accessed 04 September 2025).

Durgungoz, A., & Kharrufa, A. (2025) ‘”ChatGPT is like a study buddy, a teacher and sometimes just a friend”: a longitudinal exploration of students’ interactions, perception and acceptance’, *Interactive Learning Environments*, pp.1–20. Available at: https://doi.org/10.1080/10494820.2025.2509276 (Accessed 01 September 2025).

Fragakis, N., Trichopoulos, G., & Caridakis, G. (2025). Empowering Education with Intelligent Systems: Exploring Large Language Models and the NAO Robot for Information Retrieval. Electronics (Basel), 14(6), 1210. https://doi.org/10.3390/electronics14061210

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). ‘The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment’, *Australasian Journal of Educational Technology*, 40(4), pp.38–55. Available at: https://doi.org/10.14742/ajet.9434 (Accessed 02 September 2025).

Gates, B. (2023) *AI Is about to Completely Change How You Use Computers*. Available at: https://www.gatesnotes.com/AI-agents (Accessed: 14 August 2025).

He, Z. (2024) ‘The Journey of LLMs in Education’. *XRDS: Crossroads, The ACM Magazine for Students*, *31*(1), pp.61. Available at: https://doi.org/10.1145/3688092 (Accessed 03 August 2025).

Henkel, O., Hills, L., Roberts, B., & McGrane, J. (2025) ‘Can LLMs grade open response reading comprehension questions? An empirical study using the ROARs dataset’, *International journal of artificial intelligence in education*, *35*(2), pp.651-676. Available at: http://arxiv.org/abs/2310.18373 (Accessed 15 August 2025).

Jauhiainen, J. S., & Guerra, A. G. (2023) ‘Generative AI and ChatGPT in School Children’s Education: Evidence from a School Lesson’, *Sustainability*, *15*(18), pp.14025. Available at: https://doi.org/10.3390/su151814025 (Accessed 15 August 2025).

Jeon, J. and Lee, S. (2023) ‘Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT’, *Education and Information Technologies*, 28, pp.15873–15892. Available at: doi:10.1007/s10639-023-11834-1 (Accessed 15 August 2025).

Lang, Q., Wang, M., Yin, M., Liang, S., & Song, W. (2025) ‘Transforming Education With Generative AI (GAI): Key Insights and Future Prospects’, *IEEE Transactions on Learning Technologies*, 18, pp.230–242. Available at: https://doi.org/10.1109/TLT.2025.3537618 (Accessed 16 August 2025).

Lee, Y. (2024) ‘Development of Open Large Language Models for Artificial Intelligence Digital Textbooks’, *TEM Journal*, 13(4), pp.2773–2783. Available at: https://doi.org/10.18421/TEM134-14 (Accessed 02 September 2025).

Lee, U., Kim, Y., Lee, S., Park, J., Mun, J., Lee, E., Kim, H., Lim, C. and Yoo, Y.J. (2024) ‘Can we use GPT-4 as a mathematics evaluator in education?: Exploring the efficacy and limitation of LLM-based automatic assessment system for open-ended mathematics question’, *International Journal of Artificial Intelligence in Education*, pp.1-37. Available at: https://doi.org/10.1007/s40593-024-00448-4 (Accessed 01 September 2025).

McCaleb, L., Henriksen, D., Mishra, P., & Oster, N. (2025). ‘Between Imperative and Panic: Navigating AI’s Educational Moment with Leon Furze’, *TechTrends*, 69(4), pp.700–706. Available at: https://doi.org/10.1007/s11528-025-01116-7 (Accessed 29 August 2025).

Ng, L. (2024) ‘Llms in education’, *XRDS: Crossroads, The ACM Magazine for Students*, *31*(1), pp.66-70. Available at: https://doi.org/10.1145/3688094 (Accessed 20 August 2025).

Pesovski, I.; Santos, R.; Henriques, R.; Trajkovik, V. (2024) ‘Generative AI for Customizable Learning Experiences’. *Sustainability, 16*(7), pp.3034. Available at: https://doi.org/10.3390/su16073034 (Accessed 25 August 2025).

Protheroe, R. (2024) *12+ Incredible AI Tools for Educators*. Available at: https://pandaeducation.co.uk/blog/incredible-ai-tools-for-trainers/ (Accessed 10 September 2025).

Qorich, M., & El Ouazzani, R. (2025) ‘Detection of artificial intelligence-generated essays for academic assessment integrity using large language models’, *Expert Systems with Applications*, *291*, pp.128405. Available at: https://doi.org/10.1016/j.eswa.2025.128405 (Accessed 22 August 2025).

Raff, E., Farris, D., & Biderman, S. (2025) *How Large Language Models Work.* 1st edn. New York: Manning Publications Co.

Shan, Z., Song, Z., Jiang, X., Chen, W., & Chen, L. (2025) ‘Complementing but Not Replacing: Comparing the Impacts of GPT-4 and Native-Speaker Interaction on Chinese L2 Writing Outcomes’, *Behavioral Sciences*, *15*(4), pp.540. Available at: https://doi.org/10.3390/bs15040540 (Accessed 19 August 2025).

Sohail, A., & Zhang, L. (2025) ‘Using large language models to facilitate academic work in the psychological sciences’, *Current Psychology*, *44*(9), pp.7910–7918. Available at: https://doi.org/10.1007/s12144-025-07438-2 (Accessed 01 September 2025).

Tasdelen, O., & Bodemer, D. (2025) ‘Generative AI in the Classroom: Effects of Context-Personalized Learning Material and Tasks on Motivation and Performance’, *International Journal of Artificial Intelligence in Education*, pp.1-25. Available at: https://doi.org/10.1007/s40593-025-00491-9 (Accessed 19 August 2025).

TeachMateAI Ltd. (2025) *Instant support for teachers*. *Heavy lifting by AI*. Available at: https://teachmateai.com/ (Accessed 05 September 2025).

Toner, H. (2023) *What Are Generative AI, Large Language Models, and Foundation Models?* Available at: https://cset.georgetown.edu/article/what-are-generative-ai-

large-language-models-and-foundation-models/#:~:text=The%20Surprising%20Power%20of%20Next,that%20predict%20the%20next%20word (Accessed 02 September 2025).

Wang, J., Zhu, Z., Liu, C., Li, R., & Wu, X. (2024) 'LLM-Enhanced multimodal detection of fake news', *PloS One*, *19*(10), pp.1-21. Available at: https://doi.org/10.1371/journal.pone.0312240 (Accessed 10 September 2025).

Xavier, C., Rodrigues, L., Costa, N., Neto, R., Alves, G., Falcao, T. P., Gasevic, D., & Mello, R. F. (2025) 'Empowering Instructors With AI: Evaluating the Impact of an AI-Driven Feedback Tool in Learning Analytics', *IEEE Transactions on Learning Technologies*, *18*, pp.498–512. Available at: https://doi.org/10.1109/TLT.2025.3562379 (Accessed 21 August 2025).

Xie, T., Zhou, Y., & Yu, J. (2025) 'Ce-LLMs: Status and trends of education-specific large language models developed in China', *Future in Educational Research*, pp.505-525. Available at: https://doi.org/10.1002/fer3.70008 (Accessed 21 August 2025).

Zhang, F., Li, C., Henkel, O., Xing, W., Baral, S., Heffernan, N., & Li, H. (2025) 'Math-LLMs: AI Cyberinfrastructure with Pre-trained Transformers for Math Education', *International Journal of Artificial Intelligence in Education*, 35(2), pp.509–532. Available at: https://doi.org/10.1007/s40593-024-00416-y (Accessed 21 August 2025).

Zheng, Y.-B., Zhou, Y.-X., Chen, X.-D., & Ye, X.-D. (2025) 'The influence of large language models as collaborative dialogue partners on EFL English oral proficiency and foreign language anxiety', *Computer Assisted Language Learning*, pp.1–27. Available at: https://doi.org/10.1080/09588221.2025.2453191 (Accessed 23 August 2025).

Zou, W., Goh, T.-T., Zhu, H., Liu, M., & Yang, B. (2025) 'Algorithmic Learning: Assessing the Potential of Large Language Models (LLMs) for Automated Exercise Generation and Grading in Educational Settings', *International Journal of Human-Computer Interaction*, pp.1–18. Available at: https://doi.org/10.1080/10447318.2025.2520931 (Accessed 25 August 2025).