NIGERIAN ONLINE JOURNAL OF
EDUCATIONAL SCIENCES AND TECHNOLOGY
nojest.unilag.edu.ng
nojest@unilag.edu.ng

EXPLORING THE INTEGRATION OF SPEECH RECOGNITION TECHNOLOGIES TO SUPPORT INCLUSIVE E-LEARNING FOR MULTILINGUAL NIGERIAN DISTANCE LEARNERS WITH DISABILITIES

[1]Omoyajowo Bamidele Stephen (PhD), [1]Awosiji Temiloluwa Akinlabi & [2]Olusanjo, Michael Oluwole (PhD)

[1]Department of Science and Technology Education, Faculty of Education,
Lead City University, Toll Gate Ibadan,
Oyo State Nigeria.
Omoyajowo.bamidele@lcu.edu.ng
+2349039287758, +23470344563375
temiloluwaakinlabi6@gmail.com
+2348145189097
[2]Department of Curriculum and Instructions, Osun State College of Education, Ila-Orangun
mooolusanjo@ossceila.edu.ng
08068316276

Abstract

The rapid expansion of digital learning environments has transformed higher education by enhancing instructional delivery, expanding educational access, and promoting learner participation. In Nigeria, the increasing adoption of distance education has created opportunities to extend higher education to geographically dispersed and underserved populations. Despite these advancements, multilingual learners with disabilities continue to encounter significant barriers related to communication, language diversity, accessibility, and equitable participation in online learning environments. Speech recognition technologies have emerged as promising assistive tools capable of supporting inclusive e-learning through speech-to-text transcription, real-time captioning, and voice-enabled interaction.

This paper explores the integration of speech recognition technologies to support inclusive e-learning for multilingual Nigerian distance learners with disabilities. The study adopts an integrative literature review approach, drawing on contemporary scholarly literature, policy documents, and empirical studies on speech recognition technologies, assistive technologies, inclusive education, multilingual learning, and distance education. While speech recognition technologies have considerable potential to improve accessibility, learner engagement, communication, and participation, their successful implementation is influenced by several interrelated factors, including recognition accuracy, multilingual capability, digital infrastructure, institutional readiness, inclusive pedagogical practices, and supportive policy frameworks.

The paper further argues that effective integration requires coordinated efforts among higher education institutions, policymakers, technology developers, and other stakeholders to address linguistic diversity, technological inequalities, and accessibility challenges within the Nigerian context. It concludes that speech recognition technologies can contribute significantly to inclusive e-learning when supported by appropriate infrastructure, inclusive instructional design, and sustained institutional commitment to equitable digital education.

Keywords: Speech recognition technologies; inclusive e-learning; multilingual learners; learners with disabilities; assistive technology; distance education; Nigerian higher education.

Introduction

Digital technologies have transformed higher education by reshaping instructional delivery, expanding educational access, and increasing opportunities for learner participation. Over the past decade, universities and distance education institutions have increasingly adopted e-learning technologies to support flexible learning and widen access to higher education. However, research on accessible online learning shows that digital learning environments do not automatically guarantee inclusion, especially for learners with disabilities, unless accessibility is deliberately built into design, delivery, and institutional support systems. Recent global education reports similarly emphasise that technology can support access, equity, and inclusion, but may also deepen exclusion where infrastructure, accessibility, and learner support are inadequate (UNESCO, 2023; Lomellini et al., 2025).

Speech recognition technologies, commonly referred to as Automatic Speech Recognition (ASR) systems, are artificial intelligence applications that convert spoken language into machine-readable text through acoustic modelling, language modelling, and deep learning algorithms. These technologies have become increasingly important in educational settings because they facilitate real-time transcription, voice-based interaction, automated captioning, and hands-free navigation of digital learning environments (Jurafsky & Martin, 2025; IBM, 2023). Widely used speech recognition applications, including Google Speech-to-Text, Microsoft Azure Speech Services, and Otter.ai, have expanded opportunities for accessible learning by enabling learners to interact with digital platforms through speech rather than conventional text input.

For multilingual Nigerian distance learners with disabilities, speech recognition technologies offer significant potential to reduce communication barriers and enhance participation in online

learning. Learners with physical disabilities that limit keyboard use can employ voice commands for text entry and navigation, while learners with hearing impairments may benefit from real-time speech-to-text captioning during live lectures and recorded instructional sessions. Furthermore, when adapted to multilingual contexts, speech recognition systems can support more inclusive communication by accommodating diverse linguistic backgrounds and improving learner engagement in digitally mediated learning environments (UNESCO, 2023; Burgstahler, 2021). Nevertheless, the educational effectiveness of these technologies depends on their linguistic accuracy, accessibility, and alignment with inclusive pedagogical practices.

LITERATURE REVIEW

Concept of Speech Recognition Technology

Speech recognition technology, commonly referred to as Automatic Speech Recognition (ASR), is an artificial intelligence application that enables computers to convert spoken language into machine-readable text. Modern ASR systems integrate acoustic modelling, language modelling, and deep neural networks to recognise and transcribe human speech with increasing levels of accuracy (Jurafsky & Martin, 2025). Unlike generative artificial intelligence systems such as ChatGPT, which generate human-like text in response to prompts, ASR systems are designed specifically to recognise spoken language and translate it into written text or executable commands. Consequently, speech recognition technologies have become integral components of voice-enabled applications, digital assistants, automated captioning systems, and assistive technologies.

Recent advances in deep learning have significantly improved speech recognition performance across multiple domains, including healthcare, customer service, and education. Contemporary ASR systems such as OpenAI Whisper, Google Speech-to-Text, Microsoft Azure Speech Services, and Amazon Transcribe support multilingual speech recognition, automatic caption generation, and real-time transcription, making them valuable tools for enhancing accessibility in digital learning environments (Radford et al., 2023). Nevertheless, their performance varies considerably depending on accent variation, background noise, language diversity, and the availability of high-quality speech datasets.

Speech Recognition Technologies and Inclusive Education

Inclusive education seeks to ensure equitable participation for all learners by removing barriers that restrict access to educational opportunities. Within digital learning environments, assistive technologies have become essential tools for supporting learners with disabilities through accessible communication, alternative methods of interaction, and personalised learning experiences (Burgstahler, 2021).

Speech recognition technologies contribute to inclusive education by enabling learners to interact with digital systems through spoken language rather than conventional text input. For learners with physical disabilities affecting typing ability, voice-based interaction improves independence and reduces reliance on assistive personnel. Likewise, learners with hearing impairments benefit from real-time speech-to-text captioning, which increases access to lectures, online discussions, and recorded instructional materials. Studies have demonstrated that captioning significantly improves comprehension, information retention, and learner engagement, particularly in online learning environments (Kent et al., 2018).

Beyond disability support, speech recognition technologies also facilitate flexible learning by enabling note-taking, automated transcription, voice search, and interactive learning experiences. Consequently, ASR should be viewed not merely as an assistive technology but as a broader educational technology capable of improving accessibility for diverse learner populations.

Speech Recognition in Multilingual and Low-Resource Language Contexts

Although speech recognition technologies have achieved remarkable improvements in recent years, their effectiveness remains uneven across multilingual environments. Most commercial ASR systems are trained primarily on large English-language datasets originating from North America and Europe. As a result, recognition accuracy often declines when users speak with regional accents, indigenous languages, or mixed-language speech patterns (Koenecke et al., 2020).

This limitation is particularly relevant in multilingual countries such as Nigeria, where English frequently coexists with indigenous languages including Yoruba, Hausa, Igbo, and numerous minority languages. Learners commonly engage in code-switching and speak English with distinct regional accents that differ from those represented in commercial ASR training datasets. Consequently, existing speech recognition technologies may generate higher Word Error Rates (WER), reducing transcription accuracy and limiting their usefulness in educational contexts. Research on African language technologies similarly indicates that low-resource languages remain

significantly underrepresented in contemporary ASR development due to limited annotated speech corpora, inadequate language resources, and insufficient commercial investment (Nekoto et al., 2022). These limitations present important challenges for implementing speech recognition technologies in multilingual Nigerian distance education.

Speech Recognition Technologies in Nigerian Distance Education

Distance education has expanded significantly within Nigeria over the past two decades, particularly through the National Open University of Nigeria (NOUN) and other institutions offering online and blended learning programmes. The increasing adoption of Learning Management Systems and digital instructional platforms has improved access to higher education for geographically dispersed learners. However, persistent infrastructural challenges continue to constrain effective implementation of digital learning.

Studies consistently identify unstable electricity supply, poor internet connectivity, limited access to digital devices, and inadequate institutional support as major barriers to successful online learning within Nigerian higher education (National Universities Commission, 2022; National Open University of Nigeria, 2023). For learners with disabilities, these barriers are further compounded by limited availability of assistive technologies and insufficient accessibility policies. Within this context, speech recognition technologies offer considerable potential to improve accessibility by supporting automated captioning, voice-based interaction, and accessible communication. However, successful implementation requires institutional investment, language-sensitive technological adaptation, staff training, and policies that promote inclusive digital education.

Research Gap

Existing literature demonstrates the growing importance of speech recognition technologies for improving accessibility within digital learning environments. Nevertheless, several important gaps remain. First, most studies have been conducted in technologically advanced countries, limiting their applicability to multilingual developing contexts such as Nigeria. Second, relatively little research has examined how speech recognition technologies can support multilingual learners with disabilities in distance education. Third, current scholarship pays insufficient attention to the interaction between linguistic diversity, accessibility, infrastructure, institutional readiness, and educational policy in shaping the effectiveness of ASR technologies. This review addresses these

gaps by synthesising contemporary evidence on speech recognition technologies and examining their implications for inclusive e-learning within Nigerian higher education.

Methodology

This study adopted an **integrative literature review** design to examine the role of Automatic Speech Recognition (ASR) technologies in promoting inclusive e-learning for multilingual Nigerian distance learners with disabilities. An integrative review was considered appropriate because it allows for the synthesis of empirical studies, theoretical literature, policy documents, and technical reports to provide a comprehensive understanding of emerging educational technologies and their application within specific contexts (Torraco, 2005; Snyder, 2019). Unlike systematic reviews that are primarily designed to answer narrowly defined research questions, integrative reviews facilitate critical analysis and conceptual integration of evidence from diverse sources, making them particularly suitable for educational technology research.

The literature search was conducted using major academic databases, including **Scopus**, **Web of Science**, **ERIC**, **Google Scholar**, and **ScienceDirect**. These databases were selected because they index high-quality peer-reviewed publications in education, educational technology, computer science, and accessibility studies. To ensure that the review reflected recent developments in speech recognition technologies, particularly advances driven by deep learning and neural-network architectures, the search focused primarily on studies published between **2020 and 2025**. Earlier seminal publications were included where necessary to provide conceptual or theoretical foundations.

The search strategy combined keywords and Boolean operators to maximise retrieval of relevant studies. Search terms included *"automatic speech recognition"*, *"speech recognition technology"*, *"speech-to-text"*, *"inclusive e-learning"*, *"assistive technology"*, *"distance education"*, *"learners with disabilities"*, *"multilingual education"*, *"low-resource languages"*, *"African languages"*, and *"Nigeria"*. Additional manual searches of reference lists were undertaken to identify relevant publications that were not retrieved during the initial database search.

To ensure relevance, studies were included if they: (i) focused on speech recognition or speech-to-text technologies; (ii) examined educational applications, accessibility, or assistive technologies; (iii) addressed multilingual learning, disability inclusion, or distance education; or (iv) discussed speech recognition in low-resource or African language contexts. Publications that

focused exclusively on generative artificial intelligence, large language models, or conversational AI without direct relevance to speech recognition were excluded. Non-scholarly publications lacking sufficient methodological or conceptual rigour were also excluded, except where official policy documents from recognised organisations such as UNESCO, the National Universities Commission (NUC), or the National Open University of Nigeria (NOUN) were used to provide contextual information.

Following the screening process, the selected literature was analysed using a **thematic synthesis** approach. The reviewed studies were examined to identify recurring themes relating to the educational benefits of speech recognition technologies, accessibility for learners with disabilities, multilingual and low-resource language challenges, infrastructural and institutional barriers, and strategies for integrating ASR into inclusive e-learning environments. The synthesis focused on identifying patterns, areas of agreement, emerging debates, and gaps within the existing literature rather than merely summarising individual studies. This analytical approach enabled the development of a comprehensive understanding of the opportunities and challenges associated with implementing speech recognition technologies within Nigerian distance education.

Critical Discussion

Educational Value of Automatic Speech Recognition for Inclusive E-learning

The reviewed literature consistently demonstrates that Automatic Speech Recognition (ASR) technologies have the potential to improve accessibility and participation within digital learning environments. Unlike conventional assistive technologies that often address a single accessibility challenge, ASR supports multiple dimensions of learning simultaneously by facilitating speech-to-text transcription, automated captioning, voice-enabled interaction, and improved communication between learners and digital learning platforms. Consequently, ASR should be viewed not merely as a technological innovation but as an accessibility-enabling technology capable of supporting inclusive educational participation.

The educational benefits identified across the literature extend beyond learners with disabilities. While learners with hearing impairments benefit from real-time captioning and learners with physical disabilities gain greater independence through voice-based interaction, multilingual learners, second-language learners, and learners studying in noisy environments also benefit from improved access to instructional content (Kent et al., 2018). This finding suggests that ASR aligns closely with Universal Design for Learning by creating learning environments that are accessible

to diverse learner populations rather than targeting only specific disability groups (CAST, 2024). Accordingly, the educational value of ASR lies in its capacity to enhance flexibility, learner autonomy, and equitable participation across diverse educational contexts.

Linguistic Diversity as the Major Constraint to ASR Adoption

Although contemporary ASR systems have achieved remarkable improvements in recognition accuracy, the literature indicates that their performance remains uneven across multilingual educational environments. This limitation represents one of the most significant challenges for implementing speech recognition technologies within Nigeria. Most commercial ASR systems have been trained predominantly using high-resource English-language datasets originating from North America and Europe. Consequently, speech recognition accuracy declines considerably when users speak with regional accents, indigenous languages, or code-switched speech patterns (Koenecke et al., 2020). These findings suggest that technological advancement alone cannot guarantee educational inclusion. Rather, accessibility depends on the extent to which ASR systems are capable of recognising the linguistic diversity that characterises multilingual educational contexts.

For Nigeria, this issue is particularly significant because learners routinely combine English with indigenous languages during classroom communication. Existing commercial ASR systems frequently demonstrate limited sensitivity to these multilingual speech patterns, thereby increasing transcription errors and reducing accessibility for learners who depend on speech recognition technologies. Consequently, improving multilingual speech datasets should be regarded as an educational priority rather than solely a technological objective.

Institutional Readiness Determines Successful Implementation

The literature further demonstrates that institutional readiness exerts substantial influence on the successful adoption of ASR technologies. Although technological capability has received considerable scholarly attention, comparatively fewer studies examine the organisational conditions required for sustainable implementation. Within Nigerian higher education, institutional readiness encompasses digital infrastructure, internet connectivity, electricity reliability, staff digital competence, accessibility policies, funding, and technical support. National reports consistently identify infrastructural deficiencies as continuing barriers to effective online learning (National Universities Commission, 2022; National Open University of Nigeria, 2023).

These findings indicate that introducing ASR technologies without corresponding institutional investment is unlikely to produce meaningful improvements in educational accessibility.

Equally important is lecturer preparedness. Speech recognition technologies are most effective when integrated within pedagogically appropriate learning activities rather than being introduced as standalone technological tools. Professional development programmes focusing on accessible instructional design, digital assessment, and inclusive pedagogy therefore represent essential components of successful ASR implementation.

Towards a Framework for Integrating ASR into Inclusive E-learning

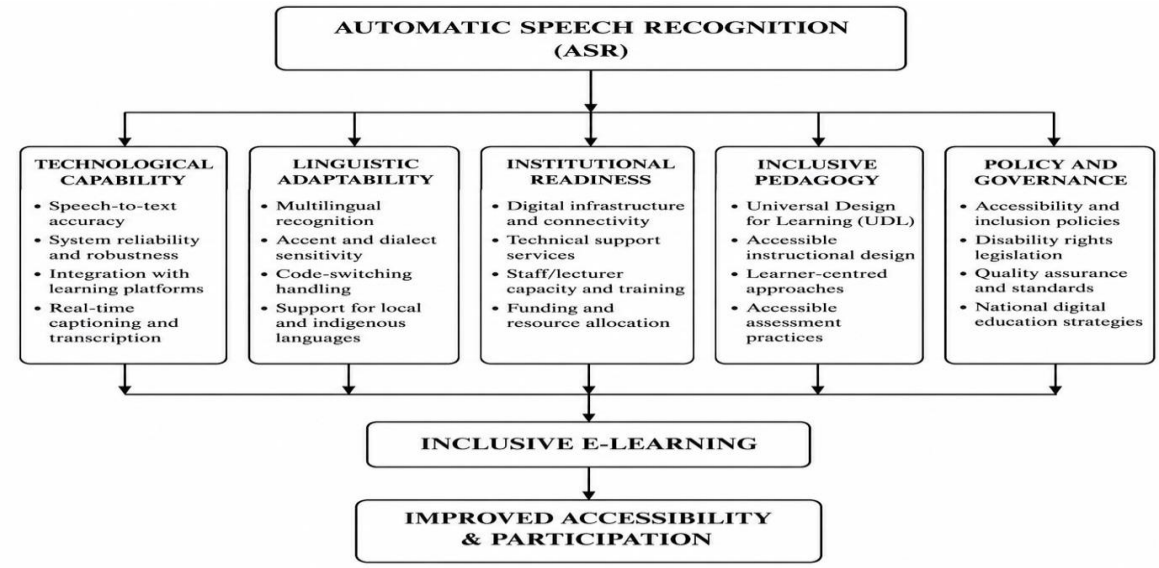
A significant finding emerging from this review is that successful implementation of Automatic Speech Recognition technologies depends upon the interaction of multiple educational, technological, and institutional factors rather than any single variable. Existing studies frequently examine these factors independently, resulting in fragmented understanding of how inclusive digital learning can be achieved within multilingual educational environments.

Based on the reviewed evidence, this paper proposes a conceptual framework comprising five interdependent dimensions represented in Fig.1

1. **Technological Capability**, including transcription accuracy, multilingual functionality, and system reliability;
2. **Linguistic Adaptability**, referring to the capacity of ASR systems to recognise diverse accents, indigenous languages, and code-switched speech;
3. **Institutional Readiness**, encompassing infrastructure, technical support, staff development, and funding;
4. **Inclusive Pedagogy**, involving accessible instructional design, Universal Design for Learning principles, and learner-centred teaching practices; and
5. **Policy and Governance**, including institutional accessibility policies, disability legislation, quality assurance mechanisms, and national digital education strategies.

The framework proposes that meaningful implementation occurs only when these dimensions operate simultaneously. Weakness within any single dimension may substantially reduce the effectiveness of speech recognition technologies regardless of technological sophistication.

Fig 1: Conceptual Framework for Integrating Speech Recognition Technologies into Inclusive E-learning for Multilingual Nigerian Distance Learners with Disabilities.



Source: Authors' conceptualisation based on the synthesis of Burgstahler (2021), Kent et al. (2018), Koenecke et al. (2020), Nekoto et al. (2022), UNESCO (2021, 2023), Williamson and Eynon (2020), and the reviewed literature.

Implications for Policy and Practice

The findings of this review have important implications for policymakers, university administrators, educational technologists, and software developers. Government agencies responsible for higher education should integrate speech recognition technologies within broader national digital inclusion strategies while encouraging institutional compliance with accessibility standards.

Higher education institutions should prioritise investments in accessible learning management systems, multilingual speech recognition technologies, lecturer professional development, and technical support services. Furthermore, technology developers should collaborate with universities, linguists, and disability organisations to develop speech recognition systems capable of recognising Nigerian English varieties and indigenous languages more accurately.

Future research should move beyond conceptual discussions towards empirical investigations examining learner outcomes, recognition accuracy, usability, and long-term educational impacts of ASR implementation within multilingual Nigerian higher education.

Conclusion

The review demonstrates that ASR technologies possess considerable potential to improve accessibility by enabling speech-to-text transcription, real-time captioning, voice-based interaction, and enhanced communication within online learning environments. These technologies can promote learner autonomy, improve engagement, and support inclusive participation when appropriately integrated into teaching and learning processes. Nevertheless, the effectiveness of ASR is influenced by several interrelated factors, including recognition accuracy, multilingual capability, institutional readiness, digital infrastructure, lecturer competence, and supportive accessibility policies.

A major finding of this review is that the successful implementation of ASR technologies extends beyond technological innovation alone. Sustainable integration requires coordinated efforts involving educational institutions, policymakers, technology developers, and disability advocates to create inclusive digital learning ecosystems. In multilingual contexts such as Nigeria, particular attention must be given to developing speech recognition systems capable of recognising Nigerian English varieties, indigenous languages, and code-switched speech to ensure equitable educational access.

The review further proposes a conceptual framework that positions technological capability, linguistic adaptability, institutional readiness, inclusive pedagogy, and policy support as interdependent determinants of successful ASR implementation. This framework contributes to existing scholarship by providing a holistic perspective for understanding the conditions necessary for integrating speech recognition technologies into inclusive e-learning within multilingual higher education contexts.

Although this review provides important conceptual insights, it is limited by the relatively small body of empirical research examining the implementation of ASR technologies within Nigerian higher education. Future research should therefore prioritise empirical investigations that evaluate the educational effectiveness, usability, and accessibility of ASR systems across different disability groups, educational settings, and indigenous language contexts.

Recommendations

Based on the findings of this review, the following recommendations are proposed:

1. **Higher education institutions** should integrate Automatic Speech Recognition technologies into their Learning Management Systems (LMS) to improve accessibility for multilingual learners with disabilities through features such as real-time captioning, speech-to-text transcription, and voice-enabled interaction.

2. **The National Universities Commission (NUC)** and other relevant regulatory agencies should develop comprehensive national guidelines and accessibility standards for the adoption of ASR technologies in higher education to promote inclusive digital learning.

3. **Government and institutional stakeholders** should invest in digital infrastructure, including reliable internet connectivity, stable electricity supply, cloud-based learning technologies, and accessible digital platforms, to support the effective implementation of ASR-enabled learning environments.

4. **Technology developers and Artificial Intelligence researchers** should collaborate with linguists, educational technologists, and disability specialists to develop multilingual ASR systems capable of recognising Nigerian English, indigenous languages, regional accents, and code-switched speech with improved accuracy.

5. **Universities and distance learning institutions** should provide continuous professional development programmes for lecturers, instructional designers, and technical support staff on the pedagogical integration of ASR technologies, inclusive instructional design, and Universal Design for Learning (UDL) principles.

6. **Institutional policies** should ensure that accessibility considerations are incorporated into the procurement, deployment, and evaluation of educational technologies, with particular emphasis on learners with disabilities and other vulnerable learner groups.

References

Ajadi, T. O., Salawu, I. O., & Adeoye, F. A. (2008). E-learning and distance education in Nigeria. Turkish Online Journal of Educational Technology, 7(4), 61–70.

Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2020). Common Voice: A massively-multilingual speech corpus. In Proceedings of the 12th Language Resources and Evaluation Conference (pp. 4218–4222).

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology, 8(1), 19–32. https://doi.org/10.1080/1364557032000119616

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. In Advances in Neural Information Processing Systems, 33, 12449–12460.

Bates, A. W. (2019). Teaching in a digital age: Guidelines for designing teaching and learning (2nd ed.). Tony Bates Associates Ltd.

Burgstahler, S. (2021). Universal design in higher education: From principles to practice (3rd ed.). Harvard Education Press.

CAST. (2024). Universal Design for Learning guidelines version 3.0. https://udlguidelines.cast.org

Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2021). Unsupervised cross-lingual representation learning for speech recognition. In Proceedings of Interspeech 2021 (pp. 2426–2430). https://doi.org/10.21437/Interspeech.2021-329

Edyburn, D. L. (2013). Critical issues in advancing the special education technology evidence base. Exceptional Children, 80(1), 7–24. https://doi.org/10.1177/001440291308000107

Eze, S. C., Chinedu-Eze, V. C., Okike, C. K., & Bello, A. O. (2018). The utilisation of e-learning facilities in the educational delivery system of Nigeria: A study of M-University. International Journal of Educational Technology in Higher Education, 15, Article 34. https://doi.org/10.1186/s41239-018-0116-z

Federal Ministry of Education. (2019). National policy on information and communication technologies in education. Federal Ministry of Education.

Graves, A., Mohamed, A.-r., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 6645–6649). https://doi.org/10.1109/ICASSP.2013.6638947

GSMA. (2023). The state of mobile internet connectivity report 2023. GSMA.

Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., & Pang, R. (2020). Conformer: Convolution-augmented Transformer for speech recognition. In Proceedings of Interspeech 2020 (pp. 5036–5040). https://doi.org/10.21437/Interspeech.2020-3015

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 29(6), 82–97. https://doi.org/10.1109/MSP.2012.2205597

Imam, S. H., Nakatumba-Nabende, J., & Muhammad, S. H. (2025). Automatic speech recognition for African low-resource languages. In Proceedings of the Sixth Workshop on African Natural Language Processing. Association for Computational Linguistics.

International Telecommunication Union. (2023). Measuring digital development: Facts and figures 2023. ITU.

Jurafsky, D., & Martin, J. H. (2025). Speech and language processing (3rd ed., draft). https://web.stanford.edu/~jurafsky/slp3/

Kent, M., Ellis, K., Latter, N., & Peaty, G. (2018). Mainstreaming captions for online lectures in Australian higher education: Alternative approaches to engaging with video content at scale. Australasian Journal of Educational Technology, 34(2), 161–174. https://doi.org/10.14742/ajet.3422

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. Proceedings of the National Academy of Sciences, 117(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117

Moore, M. G., & Kearsley, G. (2012). Distance education: A systems view of online learning (3rd ed.). Wadsworth.

National Open University of Nigeria. (2023). Annual report on open and distance learning in Nigeria. National Open University of Nigeria.

National Universities Commission. (2023). Guidelines for e-learning in Nigerian universities. National Universities Commission.

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Fagbohungbe, T., Akinola, S. O., Muhammad, S., Kabongo, S., Osei, S., Sackey, F., Niyongabo, R. A., Macharm, R., Ogayo, P., Ahia, O., Berhe, M. M., Adeyemi, M., Mokgesi-Selinga, M., Okegbemi, L., Martinus, L., … Bashir, A. (2020). Participatory research for low-resourced machine translation: A case study in African languages. In Findings of the Association for Computational Linguistics: EMNLP 2020 (pp. 2144–2160). https://doi.org/10.18653/v1/2020.findings-emnlp.195

Nigerian Communications Commission. (2024). Subscriber/network data report. Nigerian Communications Commission.

Okebukola, P. (2021). Academic integrity and quality assurance in Nigerian universities. Nigerian Journal of Higher Education, 18(2), 45–61.

Ògúnrèmí, T., Manning, C. D., & Jurafsky, D. (2023). Multilingual self-supervised speech representations improve the speech recognition of low-resource African languages with code-switching. In Proceedings of the 6th Workshop on Computational Approaches to Linguistic Code-Switching (pp. 83–88). Association for Computational Linguistics.

Oye, N. D., Salleh, M., & Iahad, N. A. (2011). Challenges of e-learning in Nigerian university education based on the experience of developed countries. International Journal of Managing Information Technology, 3(2), 39–48. https://doi.org/10.5121/ijmit.2011.3204

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). LibriSpeech: An ASR corpus based on public domain audio books. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 5206–5210). https://doi.org/10.1109/ICASSP.2015.7178964

Pratap, V., Xu, Q., Sriram, A., Synnaeve, G., & Collobert, R. (2020). MLS: A large-scale multilingual dataset for speech research. In Proceedings of Interspeech 2020 (pp. 2757–2761). https://doi.org/10.21437/Interspeech.2020-2826

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.), Proceedings of the 40th International Conference on Machine Learning (Vol. 202, pp. 28492–28518). PMLR.

Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Polity Press.
Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
Tatman, R. (2017). Gender and dialect bias in YouTube's automatic captions. In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing (pp. 53–59). https://doi.org/10.18653/v1/W17-1606
Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), 356–367. https://doi.org/10.1177/1534484305278283
Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews: Checklist and explanation. Annals of Internal Medicine, 169(7), 467–473. https://doi.org/10.7326/M18-0850
UNESCO. (2021). Artificial intelligence and education: Guidance for policy-makers. UNESCO.
UNESCO. (2023). Global education monitoring report 2023: Technology in education: A tool on whose terms? UNESCO.
W3C. (2023). Web Content Accessibility Guidelines (WCAG) 2.2. World Wide Web Consortium. https://www.w3.org/TR/WCAG22/
Watanabe, S., Hori, T., Karita, S., Hayashi, T., Nishitoba, J., Unno, Y., Yalta, N. E., Heymann, J., Wiesner, M., Chen, N., Renduchintala, A., & Ochiai, T. (2018). ESPnet: End-to-end speech processing toolkit. In Proceedings of Interspeech 2018 (pp. 2207–2211). https://doi.org/10.21437/Interspeech.2018-1456
Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. Journal of Advanced Nursing, 52(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x
Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995
World Bank. (2020). Nigeria digital economy diagnostic report. World Bank.
World Bank. (2022). Disability inclusion and education: Global report. World Bank.