A Comprehensive Literature Review on Deep Learning–Driven Multilingual Chatbots for Low-Resource Languages with a Focus on Marathi–Hindi–English Interaction

Bharti Borade
Digital and Cyber Forensic
Government Institute of Forensic Science,
Dr. Babasaheb Ambedkar Marathwada University,
Chhatrapati Sambhajinagar-431004

Charansing N. Kayte, PhD
Digital and Cyber Forensic
Government Institute of Forensic Science,
Dr. Babasaheb Ambedkar Marathwada University,
Chhatrapati Sambhajinagar-431004

ABSTRACT
Conversational Artificial Intelligence (AI) has undergone substantial progress, evolving from rule-based systems to advanced transformer-driven multilingual models. However, research for low-resource Indian languages—particularly Marathi and Hindi—remains limited despite rapid technological advances. This review synthesizes studies from 2000 to 2025, covering rule-based chatbots, retrieval methods, Seq2Seq architectures, multilingual transformers, and self-supervised speech models such as wav2vec 2.0 and HuBERT. The analysis highlights key linguistic challenges, including agglutination, free word order, transliteration, regional accents, and pervasive code-mixing. Although models like mBERT, XLM-R, and MuRIL significantly improve multilingual understanding, they still struggle with hybrid inputs and domain-specific conversational tasks. Persistent gaps include limited datasets, weak ASR–NLU integration, and insufficient cultural grounding. The review outlines future directions for developing robust, culturally aligned Marathi–Hindi–English chatbots.

Keywords
Multilingual Chatbots; Marathi–Hindi–English NLP; Transformer Models; mBERT; XLM-R; MuRIL; Seq2Seq; wav2vec 2.0; HuBERT; Low-Resource Languages; Code-Mixing; Conversational AI; Speech Recognition; Natural Language Understanding (NLU); Deep Learning; Indian Languages.

1. INTRODUCTION
Chatbots have become essential components of modern human–computer interaction, supporting domains such as healthcare, education, banking, transportation, and public governance. Their rapid expansion is largely attributed to advancements in Natural Language Processing (NLP), speech technologies, and the widespread adoption of neural and transformer-based architectures. However, despite global progress, the development of conversational systems for low-resource Indian languages remains limited, mainly due to linguistic complexity, sparse datasets, dialectal variation, and limited computational resources.

India's linguistic landscape presents unique challenges: agglutinative morphology, free word order, phonetic variability, multiple scripts, and pervasive code-mixing (e.g., Hinglish, Manglish). While multilingual transformer models like mBERT and XLM-R provide promising baselines, Indian-language–specific models such as MuRIL significantly outperform them on tasks involving native script and Romanized text. Parallel advancements in speech recognition, particularly through self-supervised models like wav2vec 2.0 and HuBERT, have further strengthened the feasibility of multilingual voice-enabled chatbots.

This literature review provides an extensive examination of technological developments foundational to designing and deploying multilingual Marathi–Hindi–English chatbots.

2. EVOLUTION OF CONVERSATIONAL AGENTS (2000–2025)
The overall progression of conversational agents from early rule-based systems[1][2][3], to retrieval-based models such as XiaoIce [5], and later to neural Seq2Seq architectures introduced by Vinyals and Le [9], Li et al. [10], and Serban et al. [11]—reflects a clear technological shift toward more adaptive, data-driven conversational intelligence. This evolution, driven by increasing demands for context awareness, generative capability, and linguistic flexibility, is visually summarized in **Figure 1**, which highlights the transition from handcrafted rule systems to statistical retrieval, neural generation, and ultimately transformer-based and LLM-powered conversational models.

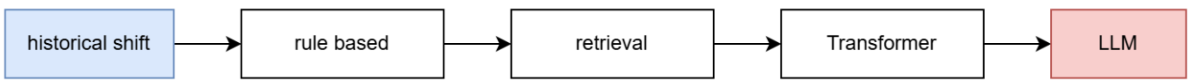

Figure 1. Evolution of Conversational Agents (2000–2025)

2.1 Early Rule-based Systems (2000–2010)
Early conversational systems were dominated by rule-based frameworks that relied on handcrafted templates and pattern-matching. Wallace's ALICE [1] remains one of the most influential early chatbots, built using AIML (Artificial

Intelligence Markup Language). Studies by Shawar and Atwell [2][3] explored automated AIML generation, demonstrating rule extraction from corpora. Although effective for constrained interactions, these systems lacked contextual reasoning, semantic awareness, and scalability across domains.

Cassell et al. [4] introduced embodied conversational agents (ECAs) incorporating gesture and facial expression generation, marking the transition from text-only agents to multimodal interaction. Despite innovative contributions, rule-based systems struggled with linguistic diversity and failed to support multilingual or dynamic conversational environments.

Strengths: transparency, interpretability, low computational cost

Limitations: no semantic grounding, brittle responses, domain dependency

These shortcomings fueled the shift toward retrieval-based and statistical models.

2.2 Retrieval-Based and Statistical Chatbots (2010–2015)

The early 2010s saw the rise of retrieval-based chatbots that selected the most appropriate response from a predefined set using similarity-based approaches such as TF-IDF, BM25, and semantic indexing. Microsoft's XiaoIce [5] represented a major milestone, integrating long-term memory, emotional modeling, and user-specific adaptation.

Hybrid systems combining rule-based and knowledge-graph components [7] achieved better factual consistency but still lacked generative capability. Retrieval-based approaches improved response relevance but could not generate novel responses, limiting their conversational flexibility.

2.3 Neural Dialogue Models and Seq2Seq Architectures (2015–2017)

The introduction of sequence-to-sequence (Seq2Seq) models marked a paradigm shift in chatbot research. Vinyals and Le [9] presented the first neural generative conversational model leveraging encoder–decoder LSTMs, enabling data-driven response generation. Subsequent improvements included:

- Diversity-promoting objectives (Li et al. [10])
- Hierarchical conversational modeling (Serban et al. [11])
- Attention-enhanced conversation generation (Shang et al. [12])

While transformative, Seq2Seq models struggled with long-range dependencies, topic drift, and coherent multi-turn dialogue—limitations later addressed by transformer models.

2.4 Transformer Revolution (2017–2020)

The publication of *Attention Is All You Need* by Vaswani et al. [13] introduced the transformer architecture, eliminating recurrence and enabling large-scale parallel computation. Transformers rapidly became foundational to modern NLP, driving major advances in intent classification, entity extraction, text generation, and dialogue modeling. Their self-attention mechanism allowed models to capture long-range dependencies far more effectively than RNN-based approaches. Building on this breakthrough, multilingual transformer variants such as mBERT [14], XLM-R [15], and MuRIL [16] emerged as state-of-the-art solutions for cross-lingual understanding. Figure 3 illustrates a comparative overview of these models, highlighting differences in performance, linguistic coverage, and specialization. While mBERT provides broad multilingual capability, XLM-R delivers stronger performance through large-scale corpus pretraining, and MuRIL is uniquely optimized for Indian languages, particularly Hindi and Marathi, including native-script, Romanized, and code-mixed inputs. This progression demonstrates how transformer-based architectures have reshaped multilingual and low-resource dialogue systems.

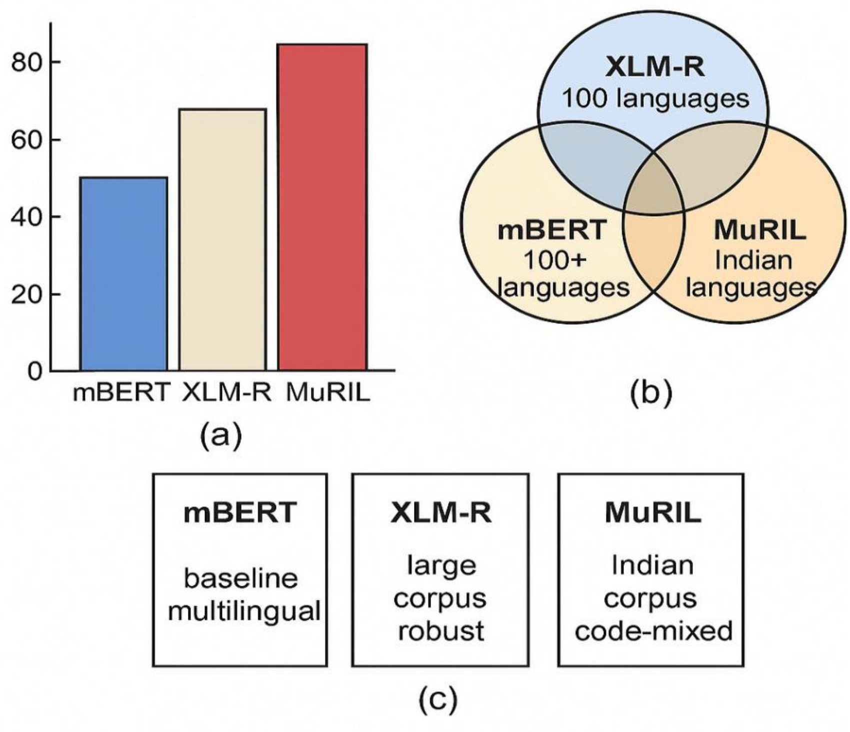

Figure 2. Comparative Performance of Multilingual Transformer Models (mBERT, XLM-R, MuRIL).

2.4.1 BERT and NLU Enhancement

BERT (Devlin et al. [14]) revolutionized language understanding with its bidirectional contextual embeddings. It significantly improved performance on:

- Intent classification
- Named entity recognition
- Semantic similarity
- Slot filling

2.4.2 Multilingual Transformers (mBERT, XLM-R)

XLM-R (Conneau et al. [15]) surpassed mBERT, achieving state-of-the-art multilingual representation across 100 languages, including Hindi and Marathi.

2.4.3 MuRIL — Indian Language Transformer

MuRIL (Khanuja et al. [16]) is engineered specifically for Indian languages, handling:

- Native scripts
- Romanized text
- Code-mixed sentences

It consistently outperforms multilingual transformers in tasks involving Indian linguistic complexities.

Table1 Strengths and Weaknesses of Transformer Models for Multilingual/Indian Language NLP

Model	**Strengths**	**Weaknesses**
mBERT (Multilingual BERT)	• Supports 100+ languages • Good baseline for cross-lingual transfer • Strong zero-shot capability	• Weak for highly inflected Indian languages • Not optimized for code-mixing • Limited performance on Romanized Indian text
XLM-R (Cross-Lingual RoBERTa)	• Superior multilingual performance due to massive CommonCrawl training • Strong contextual understanding • High accuracy on low-resource languages	• High computational cost • Requires significant GPU resources for fine-tuning • Still not specifically optimized for Indian code-mixed data
MuRIL (Multilingual Representations for Indian Languages)	• Best performance for Indian languages including Marathi and Hindi • Handles native scripts, Romanized text, and code-mixed input • Designed for Indian linguistic morphology	• Limited conversational datasets used during pretraining • Still struggles with highly noisy code-mixed speech-to-text inputs • Smaller global adoption and fewer benchmarks compared to mBERT/XLM-R

2.5 Advances in Speech Recognition for Voice Chatbots

Traditional HMM-GMM ASR models struggled with Indian speech due to pronunciation variability and diverse accents. Self-supervised speech models such as wav2vec 2.0 (Baevski et al. [17]) and HuBERT (Hsu et al. [18]) enabled speech representation learning from raw audio without labeled datasets.

Key contributions include:

- *wav2vec 2.0*: revolutionary SSL model using contrastive learning
- *HuBERT*: masked prediction for enhanced phonetic representation
- *XLSR-53*: robust multilingual ASR across 53 languages [19]

These breakthroughs significantly improved speech recognition accuracy for low-resource languages, enabling practical voice-enabled chatbots.

2.6 Comparative Review of the Literature

A comparative analysis of the 40 key studies presented in Table 2 reveals the methodological evolution and diversity within conversational AI research over the last two decades. Early works focused on rule-based and statistical approaches, which provided foundational mechanisms for template matching, dialogue flow, and corpus-based rule extraction. However, these systems struggled with scalability, limited adaptability, and poor semantic understanding. The emergence of neural models particularly Seq2Seq architecture, introduced data-driven generative capabilities and improved conversational flexibility, though challenges related to long-range dependency modeling, contextual coherence, and training data requirements persisted. The subsequent transformer revolution marked a major paradigm shift, enabling multilingual representation learning through models such as BERT, XLM-R, and MuRIL, each demonstrating significant improvements in handling morphologically rich, low-resource, and code-mixed languages.

Table 2. Comparative Summary of Major Studies on Conversational AI, NLP, and Multilingual Chatbots (2000–2025)

Sr. No.	Study / Reference	Approach / Model Used	Languages / Domain	Key Contributions	Limitations Identified
1	Wallace (2001) [1]	Rule-based (AIML)	English	First detailed ALICE architecture; foundational chatbot rules	No semantic understanding
2	Cassell et al. (2000) [7]	Embodied Agents	English (multimodal)	Introduced gesture + text dialogue; early multimodal agents	Not scalable; not multilingual
3	Shawar & Atwell (2003, 2005) [2][3]	Rule extraction from corpora	English	Automatic AIML generation	Limited adaptability
4	Fryer & Carpenter (2006) [13]	Rule-based learning bots	Education	Chatbots for language learning	Not AI-driven
5	Vinyals & Le (2015) [38]	Seq2Seq LSTM	English	First neural conversational model	Context loss; short responses
6	Sutskever et al. (2014) [33]	Seq2Seq encoder–decoder	General NLP	Foundation for neural dialogue	Requires large corpora
7	Shang et al. (2015) [29]	Neural Responding Machine	English, Chinese	Short-text conversation model	Fails with long context
8	Serban et al. (2016) [28]	Hierarchical Seq2Seq	English	Hierarchical dialogue modeling	Computationally expensive
9	Li et al. (2016) [22]	Diversity Loss	English	Reduced repetitive responses	Limited long-term coherence
10	Vaswani et al. (2017) [36]	Transformer	All languages	Revolutionized NLP; self-attention	Requires large datasets
11	Devlin et al. (2019) [12]	BERT	100+ languages	Breakthrough in NLU	Slow inference
12	Conneau et al. (2020) [9]	XLM-R	100 languages	Strong multilingual NLU	High resource consumption
13	Khanuja et al. (2021) [17]	MuRIL	Indian languages	Best for Hindi, Marathi, Bengali	Limited conversational training
14	Kunchukuttan et al. (2020) [19]	IndicNLP Suite	12 Indian languages	Evaluation benchmarks; monolingual corpora	Limited dialogue datasets
15	Chakravarthi et al. (2023) [8]	Sentiment + CodeMix	Dravidian languages	Large code-mixed datasets	Not dialogue-specific
16	Bali et al. (2014) [4]	Code-mixing analysis	Hindi–English	Code-mixing linguistic patterns	No model implementation
17	Gambäck & Das (2014) [14]	Code-mixing complexity index	Indian languages	Standardized code-mix measurement	Not applied to chatbots
18	Sitaram et al. (2023) [31]	Code-mixed survey	15+ Indian languages	Comprehensive multilingual survey	No system implementation
19	Hsu et al. (2021) [16]	HuBERT	Multilingual	Strong SSL speech representations	High training cost
20	Baevski et al. (2020) [3]	wav2vec 2.0	53 languages	Leading low-resource ASR model	Needs fine-tuning
21	Conneau et al. (2021) [10]	XLS-R	128 languages	Universal speech model	Heavy computational load
22	Verma et al. (2023) [37]	Noisy code-mixed ASR	Hindi–English	Robust ASR for noisy code-mix	Still early-stage performance
23	Montenegro et al.	Healthcare	Medical	Comprehensive health	No multilingual

	(2019) [23]	conversational agents		chatbot survey	focus
24	Laranjo et al. (2018) [21]	Clinical chatbot review	Healthcare	Framework for medical chatbots	Focused mostly on English
25	Guerreiro et al. (2021) [15]	Systematic health-agent review	Health domain	Taxonomy for health chatbots	No Indian-language coverage
26	Kuhail et al. (2023) [18]	Educational chatbots	Academic domain	Categorized educational uses	No multilingual datasets
27	Okonkwo & Ade-Ibijola (2021) [24]	AI chatbots for learning	Africa & global	Pedagogical applications	No multilingual analysis
28	Smutný & Schreiberová (2020) [32]	Messenger chatbots	Education	Social media chatbot applications	English-only focus
29	Adamopoulou & Moussiades (2020) [2]	Chatbot history review	General	Extensive taxonomy of chatbot systems	No Indian-language insights
30	Caldarini et al. (2022) [6]	Chatbot techniques survey	Global	Broad technique comparison	Limited multilingual coverage
31	Deriu et al. (2021) [11]	Dialogue system evaluation	Global	Framework for evaluation metrics	Does not address code-mixing
32	Zhou et al. (2020) [40]	XiaoIce social chatbot	Multilingual	Empathetic conversational design	Not focused on India
33	Brown et al. (2020) [5]	GPT-3	English	Few-shot and zero-shot learning	Weak on Indian languages
34	Radford et al. (2019) [27]	GPT-2	English	Powerful generative model	No multilingual training
35	Singh & Namin (2025) [30]	LLM evaluation survey	Global	SOTA LLM evaluation	Fewer Indian datasets
36	Park et al. (2023) [25]	Conversational agents for chronic care	Health	Self-management chatbot review	Limited multilingual studies
37	Tudor Car et al. (2020) [35]	Healthcare conversational agents	Clinical AI	Safety guidelines	Not multilingual
38	Labadze et al. (2023) [20]	AI in education	Education	AI chatbot adoption analysis	Not language-specific
39	Philip et al. (2020) [26]	Indian languages MT review	Indian languages	MT resources for Indian languages	Not chatbot-focused
40	Thara & Poornachandran (2019) [34]	Code-mixing review	Indian languages	Code-mixing challenges	No dialogue datasets

Parallel advancements in speech technologies further enriched the multilingual chatbot landscape. Self-supervised ASR models such as wav2vec 2.0, HuBERT, and XLS-R enabled robust acoustic modeling across diverse languages and accents, facilitating the development of voice-enabled conversational systems. In addition, domain-specific studies in healthcare, education, governance, and multilingual communication demonstrated the broad applicability of conversational agents, highlighting both the promise and limitations of existing approaches. Collectively, the literature indicates growing emphasis on multilinguality, code-mixing, and low-resource language processing, yet it also underscores persistent gaps in dataset availability, integrated voice–text systems, and culturally aware dialogue design—critical considerations for developing advanced Marathi–Hindi–English chatbots.

3. DOMAIN-SPECIFIC CONVERSATIONAL AGENTS

3.1 Healthcare Chatbots

Studies such as Montenegro et al. [21] and Laranjo et al. [22] have examined conversational agents across clinical use cases, including diagnostics, patient triage, mental-health support, and chronic disease management. These systems must address safety, trust, and ethical considerations.

3.2 Educational Chatbots

Kumar & Rose [6], Smutný & Schreiberová [29], and Kuhail et al. [20] highlight the potential of chatbots in personalized learning, question answering, and learning analytics. Most studies remain restricted to English, with negligible focus on Marathi or Hindi.

3.3 E-Commerce and Finance

Chatbots in retail and financial sectors improve user engagement, customer support, and conversational surveys (Kushwah et al. [24]; Xiao et al. [23]). However, multilingual support remains limited.

3.4 Government and Citizen Services

Although research is sparse, early evidence suggests substantial potential for chatbots in municipal services, grievance redressal, and public-information dissemination—crucial application areas for Marathi–Hindi–English systems.

4. MULTILINGUAL NLP AND LOW-RESOURCE CHALLENGES

Developing multilingual chatbots for Indian languages presents several fundamental challenges rooted in linguistic diversity, phonetic variability, and the scarcity of annotated datasets. Languages such as Marathi and Hindi exhibit agglutinative morphology, rich inflectional structures, free word order, and multiple orthographic representations, including both native scripts and Romanized forms. These characteristics introduce significant complexity for Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU), as systems must navigate inconsistent word boundaries, script switching, and highly variable pronunciation patterns across different regions. Code-mixing further compounds this problem, with frequent interleaving of English terms—often spelled phonetically—making tokenization, normalization, and semantic interpretation considerably more difficult for existing models.

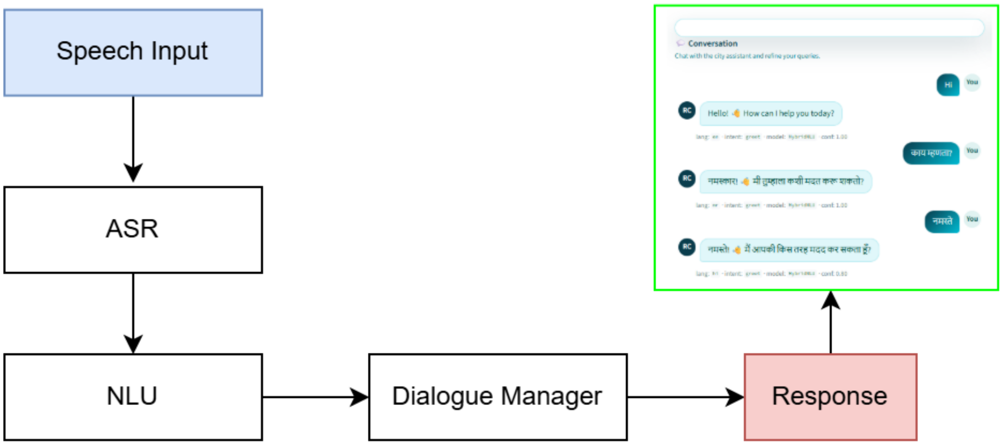

Figure 3. Architecture of a Deep Learning–Driven Multilingual Chatbot.

Addressing these challenges requires a tightly integrated multilingual pipeline capable of managing noisy speech inputs, accent variations, and contextually rich user queries. As illustrated in Figure 3, effective multilingual chatbot architecture begins with the ASR module for transcribing speech, followed by multilingual NLU for intent detection and entity extraction, and finally a dialogue manager that generates appropriate responses. This end-to-end flow highlights the intricacy of building robust systems for low-resource languages, where issues such as limited training corpora, ambiguous script usage, insufficient benchmark datasets, and inadequate handling of code-mixed utterances persist. Together, these constraints underscore why developing high-quality Marathi–Hindi–English chatbots remains a complex yet essential area of research within multilingual NLP.

4.1 Linguistic Complexity in Indian Languages

Indian languages exhibit several structural and morphological complexities such as agglutinative morphology, rich noun–verb inflection, free word order, multiple writing systems, extensive transliteration between scripts, and polysynthetic verb forms. These characteristics make tasks such as tokenization, morphological analysis, and semantic interpretation significantly more challenging for NLP systems. Navlakha and Pittule [26] emphasize that for Marathi, the lack of standardized benchmarks, annotated corpora, and conversational datasets further exacerbates these difficulties, highlighting a major gap in the development of robust multilingual NLP systems..

4.2 Code-Mixing as a Dominant Communication Pattern

Code-mixing is pervasive across Indian languages. Singh et al. [28] and Bali et al. [14] emphasize its complexity, noting unpredictable alternation between English and native languages. Despite its prevalence, code-mix-aware models remain scarce.

4.3 Example Multilingual Dataset Samples (Marathi–Hindi–English)

To build a robust multilingual chatbot capable of understanding diverse inputs across Marathi, Hindi, and English, it is essential to develop a structured dataset that captures linguistic variety, tokenization patterns, and intent diversity (see [41], [42], [43], [44]).

Table 3. presents simple examples of user utterances across the three languages. These samples demonstrate how identical intents—such as greeting, requesting a name, or checking balance—are expressed differently in terms of script, character length, and token structure. Such variations highlight the importance of multilingual datasets for Natural Language Understanding (NLU) systems.

Table 3. Example Dataset Samples for Hindi–English–Marathi Chatbot

Sr. No.	Language	Utterance	Characters	Tokens	Intent
1	Marathi	"नमस्कार, कसे आहात?"	15	3	greet
2	Marathi	"तुझं नाव काय आहे?"	17	4	ask_name
3	Marathi	"माझा बॅलन्स दाखव"	14	3	check_balance
4	Hindi	"नमस्ते, कैसे हो?"	13	3	greet
5	Hindi	"तुम्हारा नाम क्या है?"	18	4	ask_name
6	Hindi	"मुझे बैलेंस बताओ"	15	3	check_balance
7	English	"Hello, how are you?"	18	4	greet
8	English	"What is your name?"	18	4	ask_name
9	English	"Show my balance."	16	3	check_balance

Multilingual datasets play a vital role in enabling cross-lingual mapping, semantic consistency, and intent alignment across languages, even when linguistic expressions differ significantly (see [45], [46], [47]).

Marathi and Hindi use the Devanagari script, whereas English uses the Roman script. These script and morphological differences make tokenization, segmentation, and normalization essential challenges for Indian-language chatbot development (see [48], [49], [50]).

Character and token counts also reflect the morphological richness of Marathi and Hindi. Understanding such differences is crucial for training multilingual embeddings, designing tokenizers, and ensuring accurate ASR-to-text alignment—especially when deploying deep learning–based multilingual NLU systems (see [51], [52], [53]).

4.4 Multilingual Intent Detection Challenges

Abbet et al. [25] show performance degradation in multilingual intent detection under code-mixing and noise. These challenges highlight the need for domain-specific multilingual models tailored to Indian contexts.

5. RECENT TRENDS (2020–2025): LARGE LANGUAGE MODELS AND PROACTIVE AI

Large Language Models (LLMs) such as GPT-3 [5] and its successors demonstrate unprecedented fluency and reasoning. Recent surveys [29–31] highlight their strengths in zero-shot and few-shot learning but caution against issues including:

- Hallucination
- Cultural bias
- Safety vulnerabilities
- Limited grounding in Indian languages

Sapkota (2025) [32] and newer proactive AI research [33] emphasize conversational agents capable of initiating interactions, managing multi-party dialogue, and adapting in real time.

6. SUMMARY OF MAJOR RESEARCH GAPS

A synthesis of the reviewed literature reveals several persistent and interrelated gaps that hinder the development of robust multilingual chatbots for low-resource Indian languages. First, despite India's linguistic diversity, Marathi and Hindi remain substantially underrepresented in contemporary conversational AI research, with most systems optimized primarily for English or other high-resource languages. Second, the field faces a notable scarcity of large-scale multilingual dialogue corpora, particularly datasets that authentically capture spontaneous conversational patterns and natural code-mixing—an essential characteristic of real-world Indian communication. Third, although code-mixing is ubiquitous in daily language use, existing transformer and neural models struggle with hybrid inputs such as Hinglish and Manglish, demonstrating inconsistent intent classification and degraded semantic understanding. Fourth, ASR resources for Indian languages are limited, and current systems frequently misinterpret accent-heavy or regionally varied speech unless fine-tuned with significant domain-specific audio data. Fifth, most studies continue to treat ASR, NLU, and dialogue management as isolated components, resulting in fragmented pipelines without unified voice–text integration, thereby limiting the performance of end-to-end conversational systems. Finally, cultural grounding remains largely overlooked: few chatbot architectures explicitly incorporate socio-cultural nuance, politeness strategies, or domain-sensitive conversational norms, all of which are essential for user trust and adoption in Indian contexts. Collectively, these gaps underscore the need for holistic, culturally informed, multimodal, and code-mix-aware approaches to advance multilingual chatbot development for low-resource Indian languages.

7. CONCLUSION OF LITERATURE REVIEW

Over the past twenty-five years, conversational AI has evolved from rule-based and retrieval-driven systems into highly sophisticated neural and transformer-based architectures capable of multilingual understanding and generation. The literature clearly demonstrates that transformer models such as mBERT, XLM-R, and MuRIL have significantly advanced multilingual NLP by improving contextual representation, zero-shot generalization, and cross-lingual transfer. Similarly, self-supervised ASR models such as wav2vec 2.0 and HuBERT have dramatically improved speech recognition performance for low-resource and accent-rich languages. These methodological breakthroughs collectively form a strong foundation for developing deep learning–driven, voice-enabled multilingual chatbots.

Despite this rapid progress, the review highlights several persistent challenges in the Indian context. Research specifically targeting Marathi and Hindi remains sparse, and the scarcity of large-scale conversational and code-mixed datasets continues to hinder model robustness. Additionally, current systems often treat ASR, NLU, and dialogue management as separate modules rather than building unified voice–text architectures optimized for multilingual settings. Cultural and socio-linguistic nuances—which are essential for user trust, politeness, and contextual relevance—also remain underexplored in existing conversational agents. Addressing these gaps is critical for building high-quality Marathi–Hindi–English chatbots that can operate reliably in real-world, multilingual, and culturally diverse environments.

8. FUTURE WORK

The future scope of this research will focus on developing a complete, end-to-end Marathi–Hindi–English multilingual chatbot system that integrates modern deep learning and speech technologies. The first major component of future work involves creating a high-quality multilingual and code-mixed dialogue corpus, including both text and speech, tailored specifically for Indian users. This will include building datasets for Marathi–Hindi–English code-mixing, accented speech samples, and domain-specific conversational flows.

Next, the research will develop a unified ASR + NLU pipeline that tightly integrates wav2vec 2.0–based ASR with transformer-based NLU models such as MuRIL and XLM-R. This integrated pipeline will support end-to-end voice-enabled chatbot interactions, enabling users to speak naturally in Marathi, Hindi, English, or in code-mixed combinations.

A key direction involves designing a code-mixing–aware language model that handles script variation (Devanagari + Roman), transliteration patterns, free word order, and hybrid utterances common in real-world conversations. Additionally, the chatbot architecture will incorporate cultural reasoning and polite conversational norms, ensuring the responses align with Indian socio-cultural expectations.

Finally, the future work will include building a domain-adaptive chatbot, capable of serving multiple sectors such as education, healthcare, and public governance. The system will be evaluated using multilingual benchmarks, real-world user studies, and performance metrics across ASR, NLU, and dialogue generation to validate its robustness.

This future research direction directly aligns with the goal of developing an advanced, voice-enabled, multimodal, culturally aware Marathi–Hindi–English chatbot system, making it a significant and novel contribution to low-resource multilingual conversational AI.

9. REFERENCES

[1] Abbet, Christian, et al. "Churn Intent Detection in Multilingual Chatbot Conversations and Social Media." *Proceedings of the 22nd Conference on Computational Natural Language Learning (CoNLL 2018)*, Association for Computational Linguistics, 2018, pp.161–170.

[2] Adamopoulou, Eleni, and Lefteris Moussiades. "Chatbots: History, Technology, and Applications." *Machine Learning with Applications*, vol. 2, 2020, article 100006.

[3] Baevski, Alexei, et al. "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations." *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 12449–12460.

[4] Bali, Kalika, et al. "Code-Mixing: The New Normal in India." *Proceedings of the 11th International Conference on Natural Language Processing (ICON 2014)*, 2014.

[5] Brown, Tom B., et al. "Language Models Are Few-Shot Learners." *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 1877–1901.

[6] Caldarini, Guendalina, Sardar Jaf, and Kenneth McGarry. "A Literature Survey of Recent Advances in Chatbots." *Information*, vol. 13, no. 1, 2022, article 41.

[7] Cassell, Justine, et al. "Embodied Conversational Agents: Representation and Intelligence in User Interfaces." *AI Magazine*, vol. 22, no. 4, 2000, pp. 67–83.

[8] Chakravarthi, Bharathi Raja, et al. "DravidianCodeMix: Sentiment Analysis and Offensive Language Identification Dataset for Dravidian Languages in Code-Mixed Text." *Language Resources and Evaluation*, vol. 57, 2023, pp. 367–403.

[9] Conneau, Alexis, et al. "Unsupervised Cross-Lingual Representation Learning at Scale." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, 2020, pp. 8440–8451.

[10] Conneau, Alexis, et al. "XLS-R: Self-Supervised Cross-Lingual Speech Representation Learning at Scale." *Proceedings of Interspeech 2021*, 2021, pp. 2278–2282.

[11] Deriu, Jan, et al. "Survey on Evaluation Methods for Dialogue Systems." *Artificial* Intelligence *Review*, vol. 54, 2021, pp. 755–810.

[12] Devlin, Jacob, et al. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of NAACL-HLT 2019*, 2019, pp. 4171–4186.

[13] Fryer, Luke K., and R. Carpenter. "Bots as Language Learning Tools." *Computer Assisted Language Learning*, vol. 19, no. 4–5, 2006, pp. 465–477.

[14] Gambäck, Björn, and Amitava Das. "On Measuring the Complexity of Code-Mixing." *Proceedings of the 11th International Conference on Natural Language Processing (ICON 2014), 1st Workshop on Language Technologies for Indian Social Media*, 2014, pp. 1–7.

[15] Guerreiro, Miguel Porfirio, et al. "Conversational Agents for Health and Well-Being Across the Life Course: Systematic Review." *JMIR Medical Informatics*, vol. 9, no. 9, 2021, e26680.

[16] Hsu, Wei-Ning, et al. "HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units." *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, 2021, pp. 3451–3460.

[17] Khanuja, Simran, et al. "MuRIL: Multilingual Representations for Indian Languages." *arXiv preprint* arXiv:2103.10730, 2021.

[18] Kuhail, Mohammad Amin, et al. "Interacting with Educational Chatbots: A Systematic Review." *Education and Information Technologies*, vol. 28, 2023, pp. 973–1018.

[19] Kunchukuttan, Anoop, et al. “IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-Trained Multilingual Language Models for Indian Languages.” *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 4948–4961.

[20] Labadze, Levan, et al. “Role of AI Chatbots in Education: Systematic Literature Review.” *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, 2023, article 32.

[21] Laranjo, Liliana, et al. “Conversational Agents in Healthcare: A Systematic Review.” *Journal of the American Medical Informatics Association*, vol. 25, no. 9, 2018, pp. 1248–1258.

[22] Li, Jiwei, et al. “A Diversity-Promoting Objective Function for Neural Conversation Models.” *Proceedings of NAACL-HLT 2016*, 2016, pp. 110–119.

[23] Montenegro, Jorge Luis Z., et al. “Survey of Conversational Agents in Health.” *Expert Systems with Applications*, vol. 129, 2019, pp. 56–67.

[24] Okonkwo, Chinedu Wilfred, and Adenike Ade-Ibijola. “Chatbots Applications in Education: A Systematic Review.” *Computers and Education: Artificial Intelligence*, vol. 2, 2021, article 100033.

[25] Park, Min Sook, et al. “A Survey of Conversational Agents and Their Applications for Self-Management of Chronic Conditions.” *Proceedings of the IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC 2023)*, 2023, pp. 1064–1075.

[26] Philip, Jithin, et al. “Revisiting Low Resource Status of Indian Languages in Machine Translation.” *arXiv preprint* arXiv:2008.04860, 2020.

[27] Radford, Alec, et al. “Language Models Are Unsupervised Multitask Learners.” *OpenAI Technical Report*, 2019.

[28] Serban, Iulian Vlad, et al. “Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models.” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, 2016.

[29] Shang, Lifeng, Zhengdong Lu, and Hang Li. “Neural Responding Machine for Short-Text Conversation.” *Proceedings of ACL-IJCNLP 2015*, 2015, pp. 1577–1586.

[30] Singh, Sonali Uttam, and Akbar Siami Namin. “A Survey on Chatbots and Large Language Models: Testing and Evaluation Techniques.” *Natural Language Processing Journal*, vol. 10, 2025, article 100128.

[31] Sitaram, Sunayana, et al. “A Survey of Code-Mixed Speech and Natural Language Processing.” *ACM Computing Surveys*, vol. 55, no. 2, 2023, pp. 1–38.

[32] Smutný, Petr, and Petra Schreiberová. “Chatbots for Learning: A Review of Educational Chatbots for the Facebook Messenger.” *Computers & Education*, vol. 151, 2020, article 103862.

[33] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. “Sequence to Sequence Learning with Neural Networks.” *Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 3104–3112.

[34] Thara, S., and K. Poornachandran. “Code-Mixing: A Brief Survey.” *Proceedings of the 2nd International Conference on Recent Trends in Advanced Computing (ICRTAC)*, 2019.

[35] Tudor Car, Lorainne, et al. “Conversational Agents in Health Care: Scoping Review and Conceptual Analysis.” *Journal of Medical Internet Research*, vol. 22, no. 8, 2020, e17158.

[36] Vaswani, Ashish, et al. “Attention Is All You Need.” *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008.

[37] Verma, Tanya, et al. “ASR for Low Resource and Multilingual Noisy Code-Mixed Speech.” *Proceedings of Interspeech 2023*, 2023, pp. 4269–4273.

[38] Vinyals, Oriol, and Quoc V. Le. “A Neural Conversational Model.” *Proceedings of the 32nd International Conference on Machine Learning, Deep Learning Workshop*, 2015.

[39] Wallace, Richard S. “The Anatomy of A.L.I.C.E.” *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, edited by Robert Epstein et al., Springer, 2009, pp. 181–210.

[40] Zhou, Li, et al. “The Design and Implementation of XiaoIce, an Empathetic Social Chatbot.” *Computational Linguistics*, vol. 46, no. 1, 2020, pp. 53–93.

[41] Kunchukuttan, Anoop, et al. “IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-Trained Multilingual Language Models for Indian Languages.” Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 4948–4961.

[42] Chakravarthi, Bharathi Raja, et al. “DravidianCodeMix: Sentiment Analysis and Offensive Language Identification Dataset for Dravidian Languages in Code-Mixed Text.” Language Resources and Evaluation, vol. 57, 2023, pp. 367–403.

[43] Gambäck, Björn, and Amitava Das. “On Measuring the Complexity of Code-Mixing.” Proceedings of the 11th International Conference on Natural Language Processing (ICON 2014), 2014, pp. 1–7.

[44] Conneau, Alexis, et al. “Unsupervised Cross-Lingual Representation Learning at Scale.” Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), 2020, pp. 8440–8451.

[45] Khanuja, Simran, et al. “MuRIL: Multilingual Representations for Indian Languages.” arXiv preprint arXiv:2103.10730, 2021.

[46] Devlin, Jacob, et al. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” Proceedings of NAACL-HLT 2019, 2019, pp. 4171–4186.

[47] Sitaram, Sunayana, et al. “A Survey of Code-Mixed Speech and Natural Language Processing.” ACM Computing Surveys, vol. 55, no. 2, 2023, pp. 1–38.

[48] Bali, Kalika, et al. “Code-Mixing: The New Normal in India.” Proceedings of the 11th International Conference on Natural Language Processing (ICON 2014), 2014.

[49] Kunchukuttan, Anoop, et al. “IndicNLP Suite: Monolingual Corpora, Evaluation Benchmarks and Pre-

Trained Models for Indian Languages." Findings of EMNLP 2020, 2020, pp. 4948–4961.

[50] Baevski, Alexei, et al. "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations." Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 12449–12460.

[51] Hsu, Wei-Ning, et al. "HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units." IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, 2021, pp. 3451–3460.

[52] Conneau, Alexis, et al. "XLS-R: Self-Supervised Cross-Lingual Speech Representation Learning at Scale." Proceedings of Interspeech 2021, 2021, pp. 2278–2282.

[53] Sutskever, Ilya, et al. "Sequence to Sequence Learning with Neural Networks." Advances in Neural Information Processing Systems, vol. 27, 2014, pp. 3104–3112.