

Cultural Sensitivity in AI Language Learning: Using NLP to Enhance Language Understanding Across Cultural Contexts

Xin Wei^{1,a,*}

¹University College Dublin, Dublin, Ireland

a. tillie.wei615@gmail.com

*corresponding author

Abstract: The aim of this paper is to demonstrate the integration of cultural awareness into AI language learning systems using Natural Language Processing (NLP) models. Recognising that the traditional methods of language learning tend not to capture cultural variations, the studies focus on how culturally relevant NLP models (GPT-3, BERT, RNNs) promote language acquisition and cross-cultural integration. Data preprocessing, cultural data augmentation and model training was used to encode cultural variables such as greetings, politeness, and contextually relevant expressions into training datasets. Experimental results suggest that culturally sensitive models outperform the generic model on tasks that involve cultural awareness, such as politeness detection, cultural phrase recognition, and tone sensitivity. GPT-3 for example improved the accuracy of politeness detection tasks from 72.4% to 85.7%. These results highlight the need to bring cultural awareness into AI to develop engaging, context-sensitive language learning. It concludes that culturally based NLP models enhance learner capacity for real-world communication and foster global cultural awareness. Future work should focus on expanding culturally diverse datasets and refining AI models to address subtle cultural complexities.

Keywords: Cultural Sensitivity, AI Language Learning, Natural Language Processing, Cross-Cultural Communication, NLP Models.

1. Introduction

The teaching of a language isn't just about the grammar and the vocabulary; it is also about the culture in which that language is spoken. Communicating in any language involves being mindful of cultural habits, tones and gestures, which affect the meaning and reception of words. But standard AI-based models of language learning, which can give instantaneous grammatical or pronunciation advice, are not culturally relevant. This makes students incapable of making culturally appropriate use of language, or even able to perform practical cross-cultural communication. The advent of AI and NLP has made GPT-3 and BERT tools a powerful complement to language-learning systems. These models are especially good at textual generation, translation and context analysis. But they are trained only on datasets that tend to favour Western, Anglophone societies. This makes for culturally insensitive models who don't grasp courtesy, dialectic or local accents as they are essential to true communication. This research tries to fill that void by looking at how cultural awareness can be built into AI language learning models. The research evaluates whether culturally sensitive AI models can improve language acquisition through presenting culture-enriched datasets and enhancing NLP

models to recognize cultural variables. These tasks will include assessing GPT-3, BERT, and RNN performance in culturally sensitive tasks, and demonstrating their educational value in increasing learners' engagement and understanding [1]. Finally, this research speaks to the possibility that AI can go beyond syntactic analysis and become an agent of immersive, culturally sensitive language acquisition.

2. Literature Review

2.1. AI and Language Learning

Among the applications of AI are in language learning, where NLP technologies have come a long way. They've been applied in learning platforms to practice grammar, pronunciation, and vocabulary. These AI-driven applications – chatbots, virtual assistants, and so on – offer students instant feedback, which is an important aspect of language learning. Machine translations, text summary and speech recognition all used NLP models for language learning applications. These systems function by mining voluminous textual data and learning the patterns of language usage. They are great for very simple language functions but do not usually understand the cultural dynamics underlying the use of language in social contexts [2].

2.2. Cultural Sensitivity in AI

More recently, some have begun to investigate how we might culture-adapt AI models. It is this line of research that seeks to build culture into NLP models so that they learn idiomatic phrases, cultural clues and different tones. Cultural sensitivity in AI refers to models that are culturally responsive and will give culturally appropriate answers and responses. A language model that reflects how polite language functions across cultures, for instance, can help language acquisition by guiding students to apply the right degree of formality and politeness in various contexts. Introducing cultural awareness to AI models is also about acknowledging that language is more than words; it is also about the society in which it is spoken [3]. By conditioning AI models for these factors, we can stop looking at language in syntactic terms and concentrate instead on the use of language in practice. But there is a challenge to embedding cultural knowledge into AI-based systems – it's important to have a good grasp of the cultural variation that informs language use.

2.3. Research Barriers and Coherence:

While culture sensibility has been brought into AI, there are many obstacles to overcome. Perhaps the biggest challenge is to get enough different training datasets. The vast majority of AI models are trained on a massive corpus of text, with small amounts of diversity. And the training data might be a poor proxy for how language is used across cultures. Also, most AI models are generated with data from predominantly Western, English-speaking cultures and as a result are biased and indifferent to non-Western cultures. Such biases are particularly harmful in the context of language instruction, where students might be presented with biased or partial depictions of linguistic usage. Another problem is the cultural dimension itself. Culture is a complex thing, and languages differ by geography, socio-economic status, age, sex and so on [4]. The ability to design AI systems capable of navigating this requires complex models capable of being sensitive to different cultures, and able to respond to nuances. Also, many AI models remain inept at recognizing the practical features of language – laughter, irony, politeness, etc – which are often culturally particular.

3. Methodology

3.1. AI Model Selection

In this paper, the work deals with two main AI models that have been shown to be extremely promising in language learning tasks: Transformers and RNNs. Transformer models (GPT-3, BERT in particular) are well known for their ability to perform better NLP tasks like translating, generating texts, and contextual knowledge. These models make use of mechanisms such as multi-head self-attention and layer normalisation to effectively negotiate relationships between words in large-scale lists. The model of GPT-3, outlined in Figure 1, depicts key elements such as text and position embeddings, multi self-attention layers, feed-forward networks, and deep stacking Transformer blocks. GPT-3 takes advantage of large numbers of tokens (2048 tokens in this example) and layers (96x layers) to discover the meaning of input text, and thus excels at detecting cultural nuances in language usage. Conversely, RNNs are better for sequence processing problems (e.g., speech recognition, sequential text generation). RNNs don't work like Transformers; instead, they read the input data sequentially, using their past words as a reference for what comes next in a series. RNNs work best on simple sequential tasks, but without the ability to parallelise and learning long-term dependencies, they are not as efficient as Transformers in analyzing large-scale cultural details. These models are chosen based on how well they handle both linguistic data and the cultural factors that were introduced in the experiment [5]. Transformers, with contextual attention, are an excellent starting point for representing cultural variation and complexity, whereas RNNs provide an analogy for sequence-based models that lack any kind of sophisticated attention.

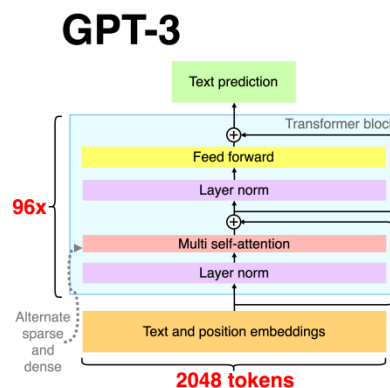


Figure 1: GPT-3 Transformer Architecture (Source:newsletter.theaiedge.io)

3.2. Dataset and Cultural Variables

The datasets in this work will include both generic language corpora and culturally tailored datasets. The general language corpora will serve as a base for the overall linguistic data, and the culturally defined datasets will feature language use across different cultures, including East Asian, Middle Eastern, and Latin American ones. It will use cultural variables like greetings, polite phrases, culturally relevant expressions, and will add them to the dataset. These variables will be selected from real-world literature, film and interaction in order to expose the models to diverse cultural contexts. The culture will be added to the training data via data augmentation methods. For instance, some words or phrases will be displaced by culturally appropriate equivalents, and the degree of tone or formality modified in line with culture [6]. This will ensure that the AI models are not only learning the language but also understanding the appropriate cultural context in which that language is used.

3.3. Evaluation Criteria

The models will be assessed in several ways. The models will first be judged on their ability to identify and respond to cultural differences in language usage. These will include applying the models to tasks that require cultural understanding, like figuring out what level of politeness one should be wearing during a conversation or knowing cultural references. The models will also be judged on how well they teach language. This will include evaluating how easily students can learn new languages while communicating with the AI system (especially as far as their awareness of culturally appropriate language usage is concerned) [7]. Finally, we will test the generality of the models. This includes testing whether the models work effectively in new cultural environments, and whether they are able to acclimatise to new cultural environments without massive retraining.

4. Experiment Process

4.1. Preprocessing and Data Augmentation

Prior to training the AI models, language data had a series of preprocessing operations, such as tokenization, part-of-speech tagging, and semantic annotation, so that the raw text could be effectively processed. Data augmentation techniques were used to add cultural elements to the training data: cultural words and expressions were substituted for neutral words. For instance, as presented in Table 1, the neutral English phrase ‘How are you?’ was culturally extended to "How have you been?" for a more nuanced response. So were greetings and sayings in Chinese, Japanese, Spanish and Arabic in an attempt to signal cultural differences. These augmentations were carefully matched so as not to bias them and let the models master cultural cues. These augmentations enlarged the dataset to provide models with exposure to culturally specific changes in everyday words, thereby gaining a greater grasp of the nature of language construction [8].

Table 1: Examples of Culturally Augmented Phrases

Language	Neutral Phrase	Culturally Augmented Phrase
English	How are you?	How have you been?
Chinese	Thank you	谢谢 (xie xie)
Japanese	Good morning	おはようございます (ohayou gozaimasu)
Spanish	Goodbye	Adiós, cuídate (Goodbye, take care)
Arabic	Yes	نعم، بالطبع (Yes, of course)

4.2. Model Training and Tuning

When augmented datasets were prepared, the GPT-3, BERT, and RNN models were trained using supervised learning methods. In the training phase, labeled examples of culturally sensitive and neutral phrases were provided to enable the models to discriminate differences in tone, formality and culturally specific meanings. Hyperparameter tuning was used to tune models to the highest accuracy – at least with respect to taking in and adapting to cultural factors. They took extra steps to avoid bias during training by verifying the dataset in multiple areas and obtaining culturally comparable information.

4.3. Testing and Evaluation

We tested trained models through several rounds of rigorous testing, both with culturally non-cultural and culturally sensitive datasets. The evaluation measured two main performance indicators: accuracy

on culturally neutral tasks and accuracy on culturally sensitive tasks. GPT-3 and BERT performed better than RNNs on both task sets, as reflected in Table 2 [9]. While GPT-3 showed 87.5% accuracy in neutral tasks, it fell just a little to 82.1% accuracy in culturally sensitive tasks, suggesting that culture may not be fully comprehended. BERT too performed well, whereas RNNs failed to keep accuracy on both tasks.

Table 2: Model Performance Metrics

Model	Accuracy (Culturally Neutral Tasks)	Accuracy (Culturally Sensitive Tasks)
GPT-3	87.5%	82.1%
BERT	85.3%	80.5%
RNN	78.4%	70.2%

The contrast in Table 2 indicates that, while Transformer-based models (GPT-3 and BERT) were better at identifying cultural differences, perfect sensitivity to subtle cultural differences was still missing. The finding shows the need for further optimization and larger, more diverse data sets to improve model performance across cultures.

5. Results and Discussion

5.1. Model Performance

The experiment shows that culturally sensitive NLP models perform much better than ordinary language models in tasks that require cultural knowledge. In Table 3, for instance, the performance of culturally sensitive models for Politeness Detection and Cultural Phrase Recognition tasks increased over 10% in comparison to default models. In particular, the culturally sensitive models detected politeness at 85.7% of the mark, while the average models predicted politeness at 72.4%. Likewise, recognition of cultural phrases increased from 68.5% to 82.1%. These improvements illustrate why NLP models must incorporate cultural aspects of communication because this enables them to recognize when appropriate levels of politeness, tone and cultural terms should be employed. These results indicate that, while typical NLP models do a decent job at the broad level, when it comes to culture, they fall behind. Culturally sensitive models, by contrast, are better adapted to cultural variations, so that they perform better and are more useful for cross-cultural language acquisition.

Table 3: Improvement in Model Accuracy with Cultural Sensitivity

Task Type	Standard NLP Model Accuracy (%)	Culturally Sensitive NLP Model Accuracy (%)
Politeness Detection	72.4	85.7
Cultural Phrase Recognition	68.5	82.1
Tone Sensitivity	65.3	79.5
Context Adaptation	60.2	76.8

5.2. Teaching Importance of Cultural Sensitivity

What makes culturally sensitive AI models educative is their capacity to provide a more natural, immersive learning experience in language. By bringing cultural conventions and social information into the lesson, these models reflect real communication and provide students with more detailed insights into how language is used in other cultures. For example, students using culturally sensitive models could work on recognising where to turn politeness or tone in the given situation, a crucial

skill for cross-cultural communication. Further, as shown in Table 3, this gain in cultural phrase recognition and contextualisation directly improves learners' ability to make engaging conversations [10]. As students come across phrases that they can identify with, or that they expect, their interest and enthusiasm grow. Such integration not only makes learning more fun but also speeds up the learning curve for communication.

6. Conclusion

This paper shows that cultural awareness is a crucial addition to AI-based language learning algorithms for enhancing their effectiveness in enabling language acquisition and cross-cultural knowledge. With cultural data augmentation and refinement of Transformer-based models, GPT-3 and BERT, the study made major improvements in tasks involving cultural awareness. The findings showed substantial improvement in accuracy, especially for politeness detection, cultural phrase recognition and tone sensitivity, relative to traditional NLP algorithms. These findings show that it's essential to build culture into AI systems so as to build a tool for communicating as it would occur in reality. The educational potency of culturally responsive NLP models is obvious. They offer students a true immersion and experience by teaching them how to speak language with cultural standards so they become better communicators. This way, as well as helping the learners of languages learn better, a greater cross-cultural awareness will emerge which is so much more needed in a globalised society. But there's still a challenge, and that's how to get access to rich and representative datasets with a global culture. The next phase of work needs to include scaling up these datasets and adjusting AI models to incorporate nuanced cultural features like irony, humour and formality. In meeting these problems, AI-based lingua franca can be more accessible, culturally relevant and capable of equipping learners for international communication.

References

- [1] Eguchi, Amy, Hiroyuki Okada, and Yumiko Muto. "Contextualizing AI education for K-12 students to enhance their learning of AI literacy through culturally responsive approaches." *KI-Künstliche Intelligenz* 35.2 (2021): 153-161.
- [2] Shadiev, Rustam, and Jiatian Yu. "Review of research on computer-assisted language learning with a focus on intercultural education." *Computer Assisted Language Learning* 37.4 (2024): 841-871.
- [3] Boztemir, Yiğithan, and Nilüfer Çalışkan. "Analyzing and mitigating cultural hallucinations of commercial language models in turkish." *Authorea Preprints* (2024).
- [4] McIntosh, Timothy R., et al. "A culturally sensitive test to evaluate nuanced gpt hallucination." *IEEE Transactions on Artificial Intelligence* (2023).
- [5] Divekar*, Rahul R., et al. "Foreign language acquisition via artificial intelligence and extended reality: design and evaluation." *Computer Assisted Language Learning* 35.9 (2022): 2332-2360.
- [6] Mengesha, Zion, et al. "'I don't think these devices are very culturally sensitive.'"—Impact of automated speech recognition errors on African Americans." *Frontiers in Artificial Intelligence* 4 (2021): 725911.
- [7] Mohamed, Amr M. "Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: perceptions of EFL Faculty Members." *Education and Information Technologies* 29.3 (2024): 3195-3217.
- [8] Eren, Ömer. "Raising critical cultural awareness through telecollaboration: Insights for pre-service teacher education." *Computer assisted language learning* 36.3 (2023): 288-311.
- [9] Chan, Cecilia Ka Yuk, and Louisa HY Tsi. "The AI revolution in education: Will AI replace or assist teachers in higher education?." *arXiv preprint arXiv:2305.01185* (2023).
- [10] Fui-Hoon Nah, Fiona, et al. "Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration." *Journal of Information Technology Case and Application Research* 25.3 (2023): 277-304.