Prototype of a robotic system to assist the learning process of English language with text-generation through DNN

Carlos Morales-Torres[1], Mario Campos-Soberanis[1,2], Diego Campos-Sobrino[1,2]

[1] Universidad Politécnica de Yucatán, Ucú, Yucatán, México
danniel.torres99@gmail.com
[2] SoldAI Research, Mérida, Yucatán, México
{mcampos,dcampos}@soldai.com

Abstract. In the last ongoing years, there has been a significant ascending on the field of Natural Language Processing (NLP) for performing multiple tasks including English Language Teaching (ELT). An effective strategy to favor the learning process uses interactive devices to engage learners in their self-learning process. In this work, we present a working prototype of a humanoid robotic system to assist English language self-learners through text generation using Long Short Term Memory (LSTM) Neural Networks. The learners interact with the system using a Graphic User Interface that generates text according to the English level of the user. The experimentation was conducted using English learners and the results were measured accordingly to International English Language Testing System (IELTS) rubric. Preliminary results show an increment in the Grammatical Range of learners who interacted with the system.

Keywords: Robotic Systems · Natural Language Processing · Text Generation · Long Short Term Memory Networks.

1 Introduction

As Artificial Intelligence (AI) becomes more equipped to comprehend human communication, more institutions will adopt this technology for areas where Natural Language Processing (NLP) would make a difference. AI technology is already being used in smart home and office assistants, customer service, healthcare, and human robotics, among others.

There are multiple aspects of AI and NLP that generate the opportunity of having machines offering engaging, interactive capabilities. However, the current state of the art in NLP lacks reasoning and empathy capabilities, making complex interactions difficult. One way to exploit NLP technology engagement potential is the application of assistive technology. A particularly interesting field is the use of such systems in interactive robotics.

Humanoid robots are useful with tedious and risky errands for people, including tasks that can result in exhausting for human beings. Jobs that require a lot

of concentration and feedback, like tutoring and guidance, can benefit from incorporating autonomous robotic systems to let the students interact with learning about a specific field. Robotic systems will require the capacity to understand human lexis to achieve these goals, making characteristic language handling more significant.

In the educational context, there are systems capable of teaching or assisting individuals in a self-learning process, such as Conversational Intelligent Tutoring Systems. However, they are still not optimal enough to automatically provide knowledge to help students in the learning process of a language without the need of human assistance [2]. Also, there have been interesting studies that show that interactive robotic systems are beneficial for learning [3]. The previous characteristics devise a synergy opportunity of a robotic system that incorporates an NLP component to be helpful in the self-learning process [20].

This article presents a functional prototype of a robotic system to assist the English language learning process through text-generation using Deep Neural Networks (DNN). A humanoid robot was designed and manufactured to promote learners' engagement with the assisting tool. The interaction was conducted using a Graphical User Interface (GUI) incorporated in the robot. A text-generation component was included to allow the users to interact with the system and generate language using different English levels. The experimentation was conducted with English learners and measured using the International English Language Testing System (IELTS) rubric. Preliminary results show an improvement of the subjects' current English level through regular usage of the system. However, there is a need for further and deeper experimentation to generalize the findings in this work.

The article is structured as follows: Section 2 describes the state of the art of robotic systems implemented to assist self-learning; Section 3 presents the research methodology; Section 4 describes the experimental work carried out, presenting its results in Section 5. Finally, conclusions and lines of experimentation for future work are provided in Section 6.

2 Background

A humanoid robot is a robotic system capable of presenting similar features to resemble human anatomy. These robots are usually presented and utilized as a research tool in scientific fields aimed to understand the human body structure and behavior to build. It has been proposed that robotics will be helpful in various education scenarios [3].

Previous studies indicate that robotics is providing benefits as a teaching tool in particular in the STEM fields [16], and English learning [10]. Robotic systems also provide a learning environment that seeks to improve the interdisciplinary process of learning, promoting the engagement of students in their learning activities [9, 17]. There are examples where the use of a robot for assisting the learning process is appropriate to use in language skill development as it allows a richer interaction than digital platforms [15, 17].

A significant challenge to incorporate robots as a tool to assist the self-learning process of a language is to design an engaging experience tightly related to the language the learner is using. NLP is particularly well suited to close this gap. NLP has evolved from simple classification methods like logistic regression to more complex language statistical methods and DNN [14]. Neural Networks are the dominant paradigm in NLP and have increased the research of end-to-end systems for understating human language, leading to complex applications as conversational chatbots [21].

The current and approachable theory of already-existing NLP models makes extensive use of transformers, which are topologies that use an encoder-decoder architecture incorporating an attention mechanism [26]. Many state of the art results make use of this architecture training with vast amounts of information. Models like BERT [7], T5 [22] and GPT-3 [6] are examples of big transformers delivering state-of-the-art results for various NLP tasks. Nevertheless, the field of NLP is still underdeveloped in terms of using low data quantities to perform fine-tuning in big transformers models.

One way to deal with low quantity data for NLP tasks is using RNNs. These models are effective for predicting sequence analysis tasks [12], as they store the information for the current feature based on previous information, including within the model forecasting and conditioned output capabilities [19].

Recurrent architectures learn the relative importance of different parts of the sequence; nevertheless, transformers substitute recurrent mechanisms with attention mechanisms [26], which allows the capture of longer size dependencies while reinforcing training.

There exist studies that favor traditional models like Conditional Random Fields (CRF) and LSTM networks over big transformers models in settings where the amount of data is not enough to perform fine-tuning, or the language specificity makes generalization difficult [13,23]. Additionally, LSTM runs faster, making it well suited for real-time systems interaction [4].

Language models (LM) also have been used for text-generation either using large transformers [25] or LSTM like in [5,18]. In this research, an LM is generated using an LSTM trained on a specific dataset, and it is used to predict the succeeding word. The predicted output word is then appended with the existing input words and given as new input. This process is continuously repeated by shifting the window to generate text.

In the presented work, a humanoid robotic system was designed and manufactured to help engage in the self-learning process of English language students. A text-generation module to expose users to a variety of vocabulary and sentences was developed, thorough the experimentation, selection, and fine-tuning of LSTM models, transformers, and encoder-decoder architectures. The best model is selected to perform text-generation using a lower seed-text as shown in [24].

3 Methodology

This section presents the tools, methodologies, and development approaches used for corpus creation, text-generation module training, humanoid robotic system design, and the system integration to allow students to interact with it.

3.1 Corpus creation

The dataset consisted on different English sentences divided into three categories: basic, intermediate, and advanced. A human expert IELTS evaluator assisted in the creation of sentences with different levels of English proficiency, considering variation in grammatical range and lexical resources according to each level.

The corpus is structured in sentences, divided by punctuation signs that are further cleaned and omitted to individual process words in the text-generation model. It contains 4,785 sentences and 150,000 words.

3.2 Text generation module

Most advanced models for text-generation make use of deep learning models, including LSTM networks and transformer architectures [8]. Different DNN models were trained using the dataset described in the previous section to develop the text-generation component. The researched models were: Simple LSTM model, BERT fine-tuned model, Encoder-Decoder LSTM model, Bidirectional LSTM model.

To process the text, the input sentences were tokenized and passed through the input layer of each model, then to an embedding layer, and subsequently fed to the RNN substructure that processes the tokens. Finally a softmax layer is used to predict the probability of the next word. The general architecture of the networks are depicted in the figure 1

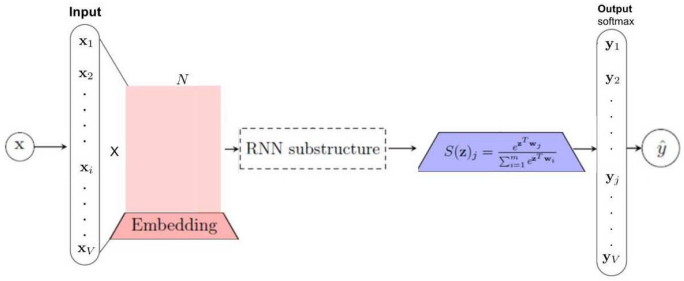

Fig. 1: General architecture of the text-generation network

Each model was implemented using the Keras framework and trained using the same dataset split with 80% for training and 20% for testing. Also, at train-

ing time, a development set proportion of 10% was used for Keras to compute validation loss and accuracy.

After experimenting with the mentioned models, the model with the best performance accuracy is selected and fine-tuned to perform the text-generation.

3.3 Robotic system design

The methodology used to design the humanoid robotic system consisted of three main phases: requirement definition, specification, and design. In the requirement definition phase, an analysis of the functionality requirements of the robot was made, and the functional structures were defined. Then, through the specification stage, the robot and general guidelines for the project were carried on. In the design stage, specifications and guidelines were measured quantitatively, including the kinematics analysis and the definition of mechanical structures.

To favor student engagement with the robot, it was decided to use an anthropomorphic system bearing kinematics considerations. Regardless, the presented robotic system does not attempt to include mechanical components; the mechanical design was made to adopt mechanical actuators further to let the system move and increase interaction with users. The parameters that represent kinematics configuration in general terms were based on Denavitt Hardenberg [11] motion equations.

After the design stage was done, the system was drawn using the 3D drawing software fusion360. The manufacturing stage consists on printing and assembling a 3D sketch of the entire robotic system with the appropriate parameters obtained from the previous analysis.

3.4 System implementation

The implementation includes an embedded system that captures the user's speech and uses Google's Text to Speech (TTS) web service to get the transcription of the user utterance. The embedded system sends the transcription to a web service implemented in Flask to consume the best text-generation model found in the experimentation. The implemented service uses the TTS transcription as a seed to predict the following text using a fixed number of 5 words. After the model predicts the text, the Flask server sends the predicted text to the embedded system using a webhook. The embedded system uses Google's Speech to Text (STT) service to generate an audio file with the predicted text and play it using a speaker. The system is attached to the robot's body, and the user initiates the interaction. Alternatively, a Graphic User Interface (GUI) was implemented using the Gradio library [1], which can consume the service using a tablet incorporated into the robot. The GUI was intended to include users with speech or hearing disabilities. The communication architecture is depicted in the figure 2.

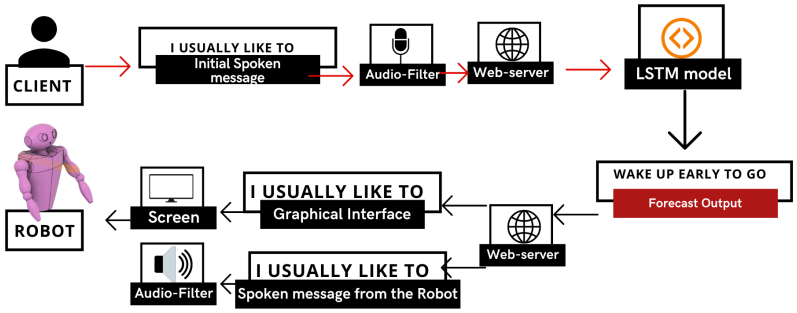

Fig. 2: System communication architecture.

4 Experimentation

This section shows the methods used for the text-generation module training, the manufacturing process of the robotic system, and the experimental process to measure system's effectivity to assist self-learning process for English students. The implemented mechanisms are illustrated, as described in section 3 of the document.

4.1 Corpus data

The corpus consisted of sentences with 3 different English levels: elemental (IELTS accuracy level 1-2.5), pre-intermediate (IELTS accuracy level 3-4.5), and upper intermediate (IETLS accuracy level 6+). Each set contained different sequence-to-sequence compound-complex sentences. This was recommended by the IELTS evaluator to optimize three specific levels of English to tackle fluency levels in different scenarios. The corpus included 171,461 tokens, 150,356 words, and 4,785 sentences.

4.2 Text-generation module

The different models were trained using the corpus described in section 4.1 divided into random partitions for training, validation, and test. Four different models were trained: Simple LSTM model, BERT fine-tuned model, Encoder-Decoder LSTM model, and a Bidirectional LSTM model. Each model was trained for 20 epochs, and the validation metrics were reported using the validation set. Different models were iterated using dropout regularization (*dropout*) with different probability parameters. Once the best model was obtained in the validation set, it was evaluated in the test data to report the metrics presented in section 5.1. The models were implemented using Tensorflow 2.0 and Keras on a Debian GNU/Linux 10 (buster) x86_64 operating system, supplied with an 11 GB Nvidia GTX 1080 TI GPU.

After the first experiments were conducted the best performance model found was the Bidirectional LSTM measured in terms of accuracy and validation. Once the best model was found further experimentation was done using a grid search strategy to find the best hyper-parameters of the model resulting in the following topology: LSTM layer (100 units), Dropout Layer (0.6 drop rate), LSTM layer (100 units), Dense layer (100 units, ReLU activation), Dense layer (125 units, softmax activation).

The best parameters found were the following: Embedding vocabulary-size: 70, dropout layer: 0.6, activation function: softmax, trainable parameters: 180,275, loss function: categorical cross entropy, batch size: 150.

4.3 Robotic system manufacturing

The whole manufacturing design was approached under engineering methods to allow time-optimization and cost reductions to be considered. The process involves the following stages: Material Printing (Through a 3-D printing machine, segments from the material were printed to further treating and assembly), Material purification (Through chemical components, the segments of materials are purified through a specific epoxy designed to purify the material extracting impurities while adding brightness, Assembly of materials. (Through engineering glue, segments are assembled properly).

Each of the previous stages was divided in three segments: head-manufacturing segment, arm-manufacturing segment, body-manufacturing segment respecting each of the previously presented stages. Final configurations of the robot using the tablet and embedded system are presented in figure 3

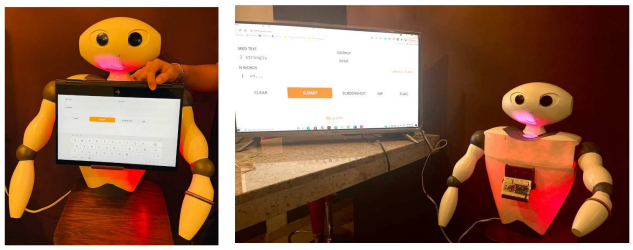

Fig. 3: Robot configurations using embedded system and tablet.

4.4 System Evaluation

To evaluate the system's effectiveness to help learners, they were evaluated using an IELTS rubric before interacting with the system. After that, the learners were exposed to interact with the system for 5 days and a new evaluation using the same rubric was made to asses the performance of the students. The evaluation was conducted with three subjects, one for each English level in the corpus.

5 Results

This section shows the results obtained from the experimentation described in section 4. The improvement of the subjects is analyzed from 250 recorded minutes of training with the system by each subject, including quantitative and qualitative evaluation from IELTS instructors. The system's performance was measured to determine the progress of the subjects.

5.1 LSTM text-forecast model with encoded-decoded attention mechanism.

Four different models were considered and evaluated to obtain the one with the best performance. Table **??** shows the accuracy obtained with the four different models when evaluated with the test dataset.

Model Type	Accuracy
Simple LSTM	80%
BERT fine tuned	80%
Encoder-Decoder LSTM	89%
Bidirectional LSTM	95%

Table 1: Model accuracy results.

The most suitable model that provided results to be used on experimental subjects was the Bidirectional LSTM model. Figure 4 shows the training accuracy and loss for the 20 epochs of training of the Bidirectional LSTM model.

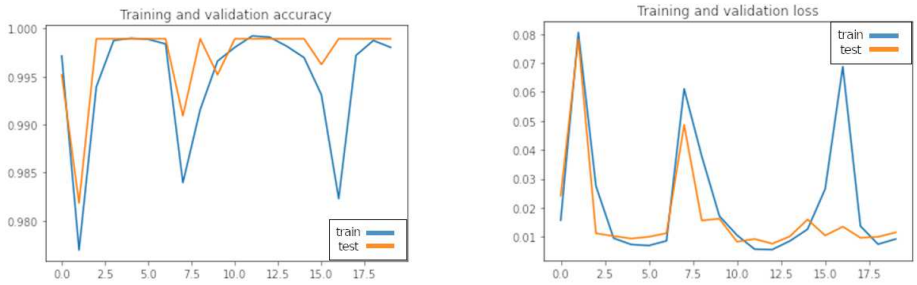

Fig. 4: Accuracy and loss validation for Bidirectional LSTM model

5.2 Fluency improvement on subjects

This section presents the outcome for the fluency analysis in each of the three experimental subjects after 250 minutes of interaction (50 minutes per day for five consecutive days) with the robotic system.

The grammatical range and accuracy and marked by using a determined number of grammatical structures (6 types) in a percentage rate of accuracy and error-mistake (1-100%). The assigned instructors included the number of grammatical sentence usage in terms of accuracy percentage.

After elementary training, an increase in grammatical range and accuracy, lexical resources, and fluency is observed, while pronunciation and language-idiomatic terminology doesn't show improvement. From the pre-intermediate level training, a sustained increase overall dimensions was observed, except for pronunciation. The upper-intermediate level attempted to evaluate fully understanding of complex ideas generated from the advanced corpus previously trained. The idea is to oversee a different set of more compound-complex sentences generated by the robotic system. The results before and after the training are showed in figure 5.

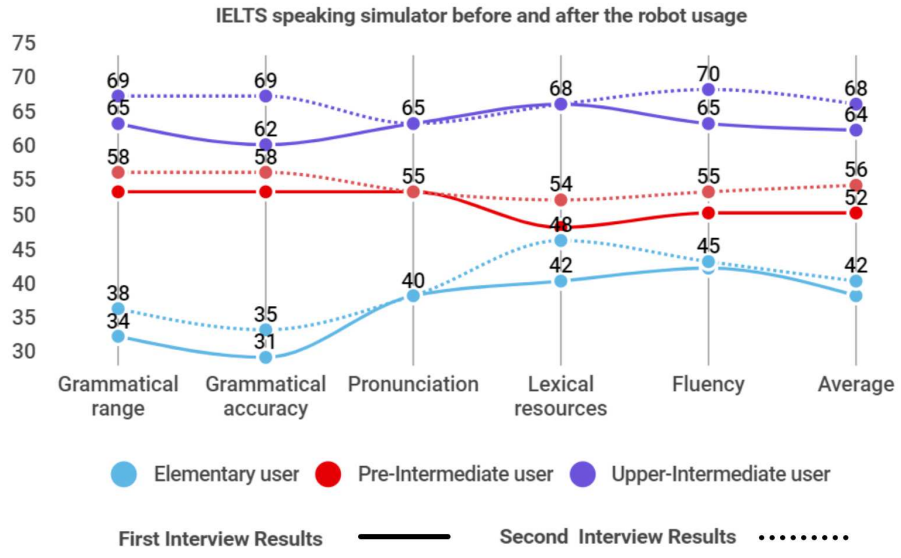

Fig. 5: IELTS metrics comparison before and after training.

5.3 Qualitative results

The qualitative data obtained in this section was collected from IELTS instructors who evaluated and listened a set of questions from one specific context of

coherence for each subject, to determine a mark in grammatical range and accuracy based on IELTS rubric. Finally, instructors who listened the same ideas in the second interview attached written feedback shown in the figure 6.

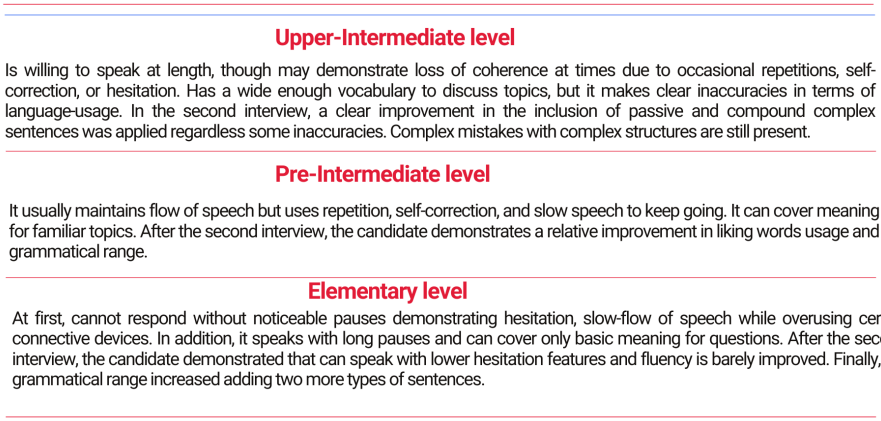

Fig. 6: Qualitative feedback from IELTS instructor after training.

The results express that the instructors perceived noticeable enhancement in the English abilities of the subjects after the interaction with the robot.

6 Conclusions and future work

This work presented the design, development, and manufacturing of a humanoid robotic system to assist English language students in a self-learning process. The robotic system was developed using a three-phase methodology (requirement analysis, specification, and design) which yields good results since the system is articulated and ready to add further interaction using actuators.

Various models were tested to implement the text-generation module; a particularly interesting observation is related to the relatively poor results (80% accuracy) obtained when using a fine-tuned BERT model. This occurs due to the relatively small amount of data used to perform the fine-tuning; in this regard, the bidirectional LSTM model performs better, achieving a 95% of accuracy in the test set.

The bidirectional LSTM text-generation model was useful to predict text using a seed given by the user; nevertheless, noticeable irregular fluctuations were reported on the validation accuracy and loss chart, which can be produced from irregularities in the English levels used within the corpus.

The experimentation was carried on with three English students of elementary, pre-intermediate, and upper-intermediate English levels, and their progress

was measured according to the IELTS rubric. After 250 hours of training, comparative results demonstrated an average improvement of 4% in their grammatical range, 4% in grammatical accuracy, and 3.33 % in their fluency. No difference was observed in their pronunciation abilities.

Quantitative and qualitative data obtained from the experimentation depicted a positive result on how a robotic system can provide aid while tackling a specific ability from a foreign language. In this case, the main improvements were reported in terms of fluency and grammatical range skills. Qualitative results show a favorable opinion both from IELTS instructors and students. In general, they perceived the system as a beneficial tool for the progress of the students.

The experimental results were limited by time constraints and the reduced number of subjects, so further research is needed to generalize the observed results.

The future work regarding this project includes: robust experimentation using more subjects and more structured training sessions, revision of other learning techniques and the overall effect on the English language improvement, experiment with variations on the composition of the corpus to measure its impact in the learning process. Also, interesting research can be conducted regarding pronunciation improvement using a more controlled spoken interaction with the users and the effect of dynamic movement adding actuators to the robot and measuring the impact in the self-learning process.

References

1. Abid, A., Abdalla, A., Abid, A., Khan, D., Alfozan, A., Zou, J.: Gradio: Hassle-free sharing and testing of ml models in the wild. arXiv preprint arXiv:1906.02569 (2019)
2. Akkila, A., Almasri, A., Ahmed, A., Al-Masri, N., Abu, Y., Mahmoud, A., Zaqout, I., Abu-Naser, S.: Survey of intelligent tutoring systems up to the end of 2017. International Journal of Academic Information Systems Research **vol. 3**, 36–49
3. Anwar, S., Bascou, N., Menekse, M.: A systematic review of studies on educational robotics. Journal of Pre-College Engineering Education Research (J-PEER) **9(2)** (2019)
4. Breuel, T.M.: Benchmarking of LSTM networks. CoRR **abs/1508.02774** (2015), http://arxiv.org/abs/1508.02774
5. Cho, K., Merrienboer, B.v., Gulcehre, C., Bougares, F., Schwenk, H., Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In: EMNLP (2014), https://hal.archives-ouvertes.fr/hal-01433235
6. Dale, R.: Gpt-3: What's it good for? Natural Language Engineering **27**(1), 113–118 (2021). https://doi.org/10.1017/S1351324920000601
7. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the NAACL-HLT Volume 1. pp. 4171–4186. Association for Computational Linguistics (2019), https://doi.org/10.18653/v1/n19-1423
8. Ezen-Can, A.: A comparison of lstm and bert for small corpus. ArXiv **abs/2009.05451** (2020), http://arxiv.org/abs/2009.05451

9. Grubbs, M.: Robotics intrigue middle school students and build stem skills. Technology and Engineering Teacher **72(6)**, 12–16 (2013)
10. Han, J., Kim, D.: r-learning services for elementary school students with a teaching assistant robot. In: 2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 255–256 (2009)
11. Hayat, A.A., Chittawadigi, R.G., Udai, A.D., Saha, S.K.: Identification of denavit-hartenberg parameters of an industrial robot. In: Proceedings of Conference on Advances In Robotics. p. 1–6. AIR '13, Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2506095.2506121
12. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF Models for Sequence Tagging (2015), Retrievedfromhttp://arxiv.org/abs/1508.01991
13. Joselson, N., Hallén, R.: Emotion classification with natural language processing (comparing bert and bi-directional lstm models for use with twitter conversations) (2019), student Paper
14. Khurana, D., Koli, A., Khatter, K., Singh, S.: Natural language processing: State of the art, current trends and challenges. CoRR **abs/1708.05148** (2017), http://arxiv.org/abs/1708.05148
15. Lai Poh, E.T., Albert, C., Pei-Wen, T., I-Ming, C., Song Huat, Y.: A review on the use of robots in education and young children. Journal of Educational Technology & Society **19**(2), 148–163 (2016), http://www.jstor.org/stable/jeductechsoci.19.2.148
16. Melchior, A., Cohen, F., Cutter, T., Leavitt, T.: More than robots: An evaluation of the first robotics competition participant and institutional impacts. In: Brandeis University Center for Youth and Communities Heller School for Social Policy and Management. (2005)
17. Menekse, M., Higashi, R., Schunn, C.D., Baehr, E.: The role of robotics teams collaboration quality on team performance in a robotics tournament. Journal of Engineering Education **106(4)**, 564–584 (2017)
18. Merity, S., Keskar, N.S., Socher, R.: Regularizing and optimizing LSTM language models. 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings (2018)
19. Muzaffar, S., Afshari, A.: Short-Term Load Forecasts Using LSTM Networks. Energy Procedia **158**, 2922–2927 (2019), https://www.sciencedirect.com/science/article/pii/S1876610219310008
20. Newton, D.P., Newton, L.D.: Humanoid robots as teachers and a proposed code of practice. Frontiers in Education **4**, 125 (2019), https://www.frontiersin.org/article/10.3389/feduc.2019.00125
21. Peters, F.: Master thesis: Design and implementation of a chatbot in the context of customer support (2018), http://hdl.handle.net/2268.2/4625
22. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.: Exploring the limits of transfer learning with a unified text-to-text transformer. CoRR **abs/1910.10683** (2019), http://arxiv.org/abs/1910.10683
23. Rosvall, E.: Comparison of sequence classification techniques with bert for named entity recognition. Electrical Engineering and computer science faculty (2019)
24. Santhanam, S.: Context based text-generation using LSTM networks. CoRR **abs/2005.00048** (2020), https://arxiv.org/abs/2005.00048
25. Topal, M.O., Bas, A., van Heerden, I.: Exploring transformers in natural language generation: Gpt, bert, and xlnet. CoRR **abs/2102.08036** (2021), https://arxiv.org/abs/2102.08036

26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. CoRR **abs/1706.03762** (2017), http://arxiv.org/abs/1706.03762