Automated Deep Learning Based Answer Generation to Psychometric Questionnaire: Mimicking Personality Traits

Anirban Lahiri[1], Shivam Raj[2], Utanko Mitra[2], Sunreeta Sen[3], Rajlakshmi Guha[2], Pabitra Mitra[2], Partha P. Chakrabarti[2] and Anupam Basu[2]

[1]*Kainos Ltd., London, U.K.*

[2]*Indian Institute of Technology, Kharagpur, India*

[3]*Arndit Ltd., Cambridge, U.K.*

Keywords: Deep Learning, Natural Language Processing, Psychometric Testing, Personality Trait Emulation.

Abstract: Questionnaires have traditionally been used for psychometric testing and evaluation of personality traits. This work explores if personality traits or characteristics can be emulated by a computer through the responses to a questionnaire. A state-of-art Deep Learning model using natural language processing techniques coupled to a personality prediction model has been exploited. A standard OCEAN – Five-Factor evaluation questionnaire was used as the test bench for this novel study combining psychometry and machine learning. This article explains the design details of the emulation framework, the obtained results and their significance. The obtained results look promising and the framework can potentially find commercial or academic application in the near future.

1 INTRODUCTION AND BACKGROUND

The use of Artificial Intelligence (AI), especially Deep Learning(DL) for personality detection has been researched only very recently(Mehta et al., 2020). It has been decades since (Mcculloch and Pitts, 1943) proposed the Perceptron and implemented by (Rosenblatt, 1958) as a mechanism mimicking human neurological processes. The science and technology for neural networks has progressed ever since and we are now in the era of Deep-Learning. However, there is still much work remaining in terms of modelling human behaviour through neural networks and deep-learning. Recent work in modelling human behaviours and traits through Artificial Intelligence(AI)/ Machine Learning (ML) include the following. An experiment on face based personality detection of OCEAN - Big Five Factors was done by (Al Moubayed et al., 2014). Predicting personality traits from physical behaviour like everyday eye movements has been researched (Hoppe et al., 2018). It has also been possible to predict private traits and attributes from people's online presence (e.g. Facebook Likes) using AI / Machine Learning (ML) as outlined by (Stillwell et al., 2013). Computer based games for psychometric analysis have been proposed by (Lahiri et al., 2020). Work has also been done on modelling behaviour of groups through inverse planning (Shum et al., 2019) and predicting group behaviour.

Along with rapid advancements in Artificial Intelligence, Machine Learning, and Deep Learning, computers have become increasingly more efficient at tasks that were formerly considered to be the forte of humans. These tasks even involve artistic skills, (Wang et al., 2020) reviews image synthesis using generative adversarial networks. Recent advances in semantic capture of natural languages using DL (Wu et al., 2016) have paved the way for this work. Deep-learning has been proven to capture not only the syntax but also the semantics of sentences through sentence-embeddings (Reimers and Gurevych, 2019). Transformers have proved the ability to capture the human attention mechanism for language translation tasks (Vaswani et al., 2017).

However, not much work has been done towards psychometric emulation using AI especially the advance in DL and the authors are unaware of any studies using Deep Learning for emulating personality traits or characteristics through questionnaires. Very little literature exists in this domain and it can become key area for nascent research in the near future.

Lahiri, A., Raj, S., Mitra, U., Sen, S., Guha, R., Mitra, P., Chakrabarti, P. and Basu, A.
Automated Deep Learning Based Answer Generation to Psychometric Questionnaire: Mimicking Personality Traits.
DOI: 10.5220/0011772200003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 737-743
ISBN: 978-989-758-623-1; ISSN: 2184-433X

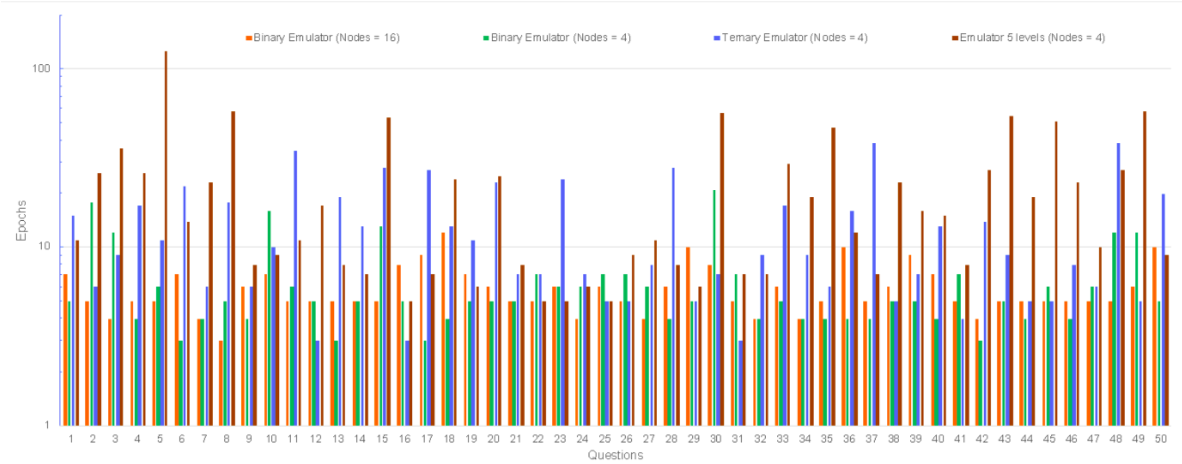

Figure 1: Neural Network training duration plots corresponding to various emulators.

2 OCEAN FIVE FACTOR

Alan Turing once conjectured that at some point in time machines would be able to mimic human thought and interaction manifested through an “imitation game” (Turing, 1950). This work attempts to take one more step in that direction by exploring whether it is possible for computers to imitate human personalities.

The OCEAN Five-Factor model is well known to psychologists(Goldberg, 1992) and will be described briefly later in this section. This work explores whether it is possible for a computer to impersonate a particular combination of the five factors defining a person’s characteristics. A set of 50 questions (in the form of a questionnaire) is answered by the computer while imitating that particular set of OCEAN values.

Although it is theoretically possible for a computer to mimic a particular personality and answer the questionnaire using an exhaustive set of look-up tables it is not an elegant solution due to the following pitfalls:

- If there are 50 questions and the value of each of the 5 personality traits ranges from 1 to 40, the number of rows on the look-up table would be 50×40^5=512×10^7. Though this technically achievable it is not a prudent solution for a much higher number of questions.
- If the questions which are part of the questionnaire were rephrased, then the computer would be clueless about how to answer the questions to reflect a particular personality.

Therefore, an approach using look-up tables though theoretically possible is not a recommended solution for practical purposes. Deep-Neural Networks(DNN) have shown promise in being able to generalize their learning and make effective decisions (Mehta et al., 2020). The minimum theoretical neural network for answering each question for a given combination of OCEAN personality traits is shown in Figure 2. This work investigates whether it is possible for such a network to impersonate a set of personality traits while answering a psychometric questionnaire. The next paragraph gives brief outline of the OCEAN five factor model which is often used for psychometric evaluation. The OCEAN model has been chosen since it freely available and does not require any licensing.

For evaluating the OCEAN factors, a questionnaire/form with 50 questions (Goldberg, 1992) needs to be answered by the test subject. Each question has 5 possible answers numbered 1 to 5, out of which the subject needs to choose one. Finally, the subject for the personality test is given a score between 0 to 40 (total of 41 values) for each of the 5 OCEAN personality traits.This questionnaire was used as a basis for building the DNN framework.

OCEAN - Five Factors: The Big Five personality

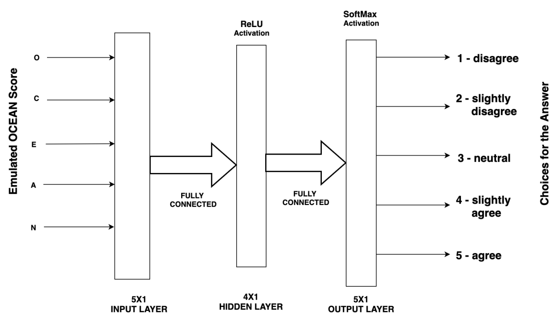

Figure 2: Quinary emulator.

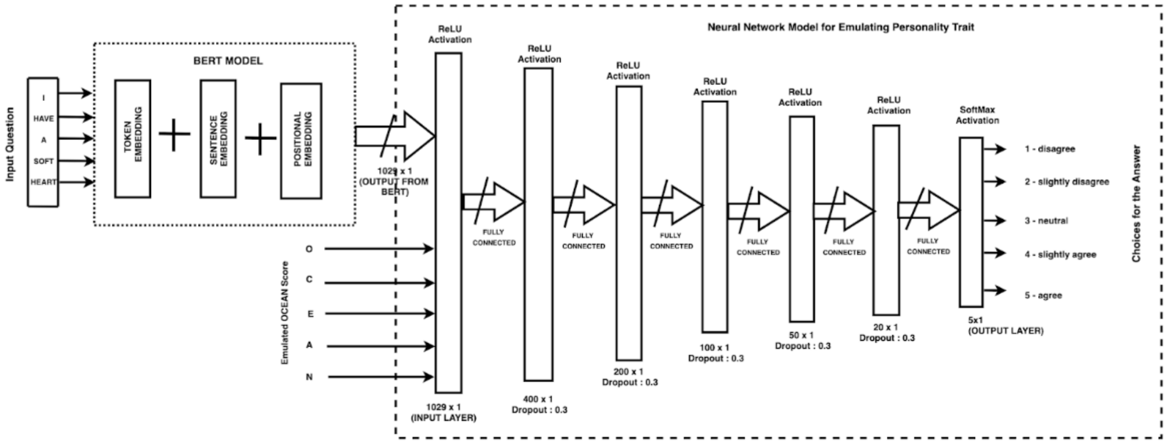

Figure 3: Deep Neural Network framework Personality emulation using BERT.

traits, also known as the five-factor model (FFM), is a taxonomy for personality traits (Goldberg, 1992). It is based on common language descriptors. The five factors have been defined as follows:

- Openness to experience (inventive/curious vs. consistent/cautious).
- Conscientiousness (efficient/organized vs. easy-going/careless).
- Extroversion: (outgoing/energetic vs. solitary/reserved)
- Agreeableness (friendly/compassionate vs. challenging/detached)
- Neuroticism (sensitive/nervous vs. secure/confident).

The objective was to train a Deep Neural Network which would be able to answer the questions in the questionnaire so as to reflect a particular set of of OCEAN characteristics(E.g set of OCEAN scores : 20, 30, 35, 15, 20). A set of minimalistic Neural networks(NN) were created, where each NN corresponded to one question in the questionnaire. These were then trained through rote learning and minimized in size. The summarized results are shown in Figure 1. The NNs were first tried with binary response to the questions (either 'yes' or 'no') and each output for the 5 factors would correspond to a high or a low. It was then progressively made more rich to capture the nuances of the answers (ternary, quinary - 5 levels) as per the standard set by (Goldberg, 1992). The generic structure of the quinary emulator is shown in Figure 2).

The results summarized in Figure 1 indicate the order of the resources required by the NNs for capturing the spectrum of answers. The training time in terms of epochs varied between 3 and 110 for each

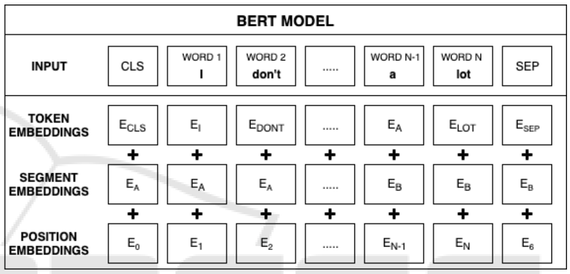

Figure 4: BERT language representation model for questions.

question with an average of 5.9 as shown in Figure 1. Validation accuracy of 100% was required for the training to be completed. It was found that a single hidden layer with 4 nodes was able to produce the correct output to 100% accuracy and the number of epochs required for training were fairly low. An NN any smaller than 4 nodes in the hidden layer failed to reach 100% accuracy even after 1000 epochs. More nodes in the hidden layer did not contribute much both in terms of the output spectrum or learning time.

However, as mentioned before such simplistic NN models would not be able to capture the nuances of the language therefore a robust language model could be necessary, otherwise the results obtained would seem meaningless. The next section describes the incorporation of a language model to the DNN and the corresponding validation mechanism.

3 LANGUAGE MODEL AND SUITABILITY OF BERT

A language model is required for capturing the semantics of the questions in the questionnaire. The BERT model (Devlin et al., 2018) claims improve-

#	Set 1	Set 2	Set 3
1	I Am the life of the party.	I am talkative	I like to go unnoticed in the party
2	I Feel little concern for others.	I am not worried about other people's lives	I am sympathetic towards others
3	I Am always prepared.	I am always ready for any situation	I am unplanned
4	I Get stressed out easily.	I worry a lot when faced with a difficulty	I have a cool head
5	I Have a rich vocabulary.	I use a lot of different words	I am usually out of words
6	I Don't talk a lot.	I am reserved	I am talkative
7	I Am interested in people	I take interest in other people	I am indifferent to others
8	I Leave my belongings around.	I am somewhat careless	I take care my belongings
9	I Am relaxed most of the time.	I can easily handle stress	I am usually stressed
10	I Have difficulty understanding abstract ideas	I cannot understand complicated concepts easily	I comprehend abstract ideas easily

Figure 5: Examples of Rephrased questions.

#	Set 1- Negative	Set2 Negative
1	I Am not the life of the party.	I am not talkative
2	I don't Feel little concern for others.	I am worried about other people's lives
3	I Am always unprepared.	I am not always ready for any situation
4	I don't Get stressed out easily.	I don't worry a lot when faced with a difficulty
5	i dont have a rich vocabulary	I don't use a lot of different words
6	I talk a lot.	I am not reserved
7	I Am not interested in people	I do not take interest in other people
8	I don't Leave my belongings around.	I am not careless
9	I Am not relaxed most of the time.	I can not easily handle stress
10	I don't Have difficulty understanding abstract ideas	I can understand complicated concepts easily

Figure 6: Examples of Negative Rephrased questions.

ments in natural language inferencing, paraphrasing and enhanced ability for question-answer tests. This study leverages the capabilities of such a pre-trained language representation model to answer psychometric questions while mimicking a particular personality.

In this case the BERT(Bidirectional Encoder Representations from Transformers) model was pre-trained and only required minor fine-tuning (HuggingFace, 2018). The BERT model is capable of mapping sentences (questions in this case) to a set of embeddings capturing both the semantics of the words and their positions in the sentence as illustrated in Figure 4. Additional Neural Network layers were added to the output layer of the BERT model to convert the positional embeddings from the BERT model to the final question response. The resulting DNN as shown in Figure 3 was then trained iteratively. The input to the DNN is the question text. The pre-trained BERT model outputs a set of 1024 embeddings which are then provided as input to the second phase of the DNN. The second phase of the DNN accepts the embeddings corresponding to a question from the BERT models as well the combination of the OCEAN scores to emulate and produces as output one of the following five answers:

- disagree
- slightly disagree
- neutral
- slightly agree
- agree

The Personality Emulator section of the NN has 5 dense layers in between its input and output layers. Note, that this size-optimized NN was arrived at empirically after rigorous experimentation with NNs of varying depths and number of nodes in each layer.

BERT is fundamentally a transformer language model with a number of encoder layers that may be changed. A transformer is a deep learning model that uses the self-attention process and weighs the importance of each component of the input data differently. It has been used in effectively in the past for natural language processing (NLP) (Devlin et al., 2018). It captures the concept of attention in neural networks since it tries to mimic cognitive attention in humans. It does so by enhancing some parts of the input data while diminishing other parts which insignificant to the semantics.

Every input embedding is a combination of three embeddings as depicted in Figure. 4. Positional em-

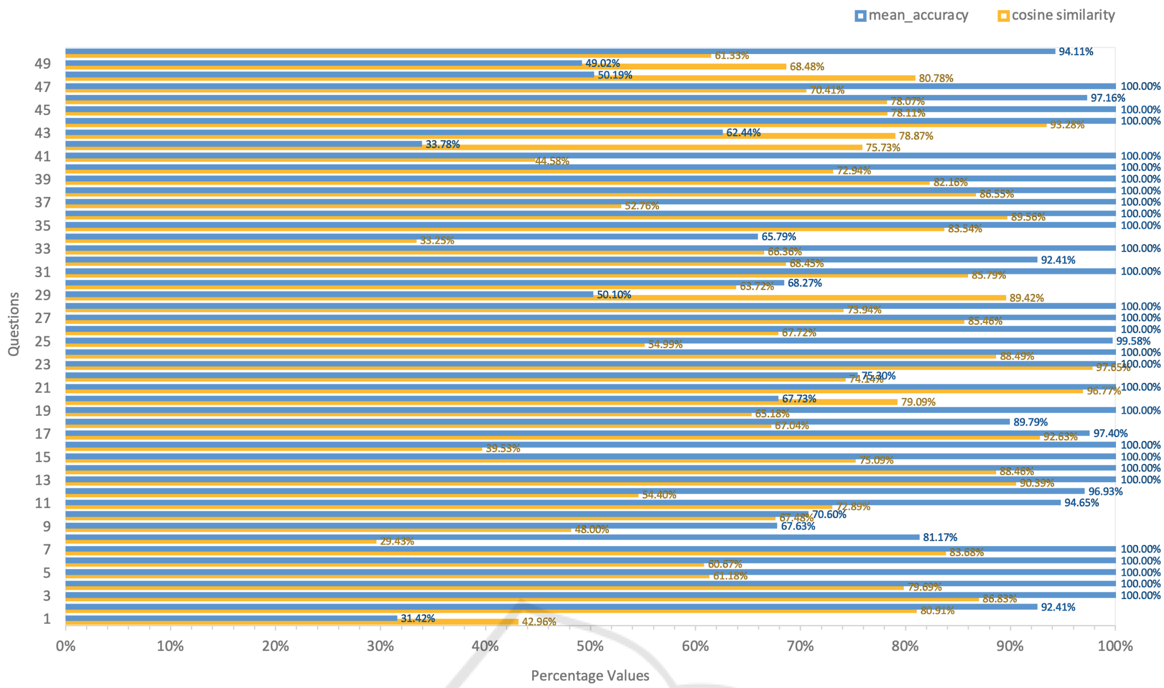

Figure 7: Mean accuracy of the pair of questions with corresponding cosine similarity.

bedding which is learned by BERT to express the position of the words in a sentence which overcomes the limitation of transformer, unlike recurrent neural network(RNN) where it is unable to capture "sequence" or "order" information in case of NLP. Segment embedding is able to take a pair of sentences or phrases of the same sentence as inputs for tasks(E.g. Question-Answering) and can differentiate between them. Lastly, there is the token embeddings learned for the specific token from the WordPiece token vocabulary (Wu et al., 2016). The input representation for a particular question is represented by the relevant token, segment, and position embeddings. Therefore, it is an effective tool for deriving the meanings of the questions in a psychometric test and finding suitable answers. An additional spelling checker which uses Levenshtein distance (Levenshtein et al., 1966) for checking and correcting spellings can be used if needed but is not necessary for the current scope.

It is desirable to generalize the emulation framework such that the results can be reproduced even if the questions in the questionnaire are rephrased. In order to achieve this a few variants of the original set of questions were created as shown in Figures 5 and 6. Set 1 refers to the original set of questions from (Goldberg, 1992). Set 2 is semantically identical to Set 1, i.e. each questions have been rephrased while keeping the meaning the same. This was done to validate whether the DNN is able to answer the questions correctly though text of the question has changed. The cosine similarities (Li et al., 2004) between Set 1 and Set 2 as output by the BERT network are shown in figure 8. Set 3 represents the semantic negative of the corresponding question in Set 1 without using a strong negative word like"not' or "don't". Therefore it is expected to be more complicated for the DNN to interpret it as the semantically opposite to the corresponding question in Set 1. In contrast, Figure 6 shows two additional sets of questions which are semantically opposite to the corresponding question in Set 1, but use a strong negative word like "not" or "don't". Set 1 was used for training the DNN while all the other sets were used for test and validation.

The computational platform used for this work was a server with 40 virtual Xeon processor cores, 96GB RAM, and an Nvidia K80 GPGPU (Nvidia, 2014) for accelerating the DNN computations. For each iteration, a total of approximately 10 hours of machine time was required for training the DNNs using BERT for all 50 questions.

4 RESULTS AND DISCUSSION

Figure 7 shows the average accuracy of the emulator for all questions across all possible personality types which is numerically $(41)^5$. The mean accuracy for the questions ranges from 31.4% to 100% with a com-

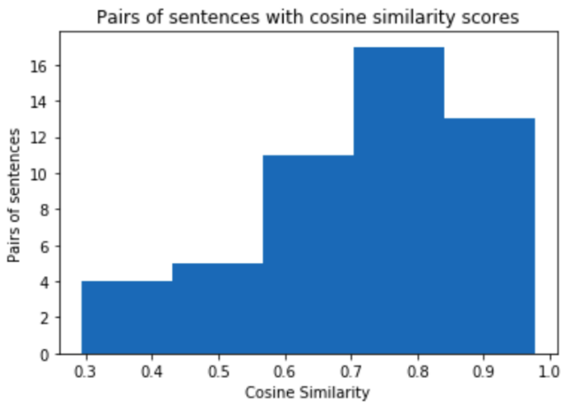

Figure 8: Distribution of cosine similarities between Set 1 and Set 2.

bined average of 88.55%. On the other hand the Cosine similarity of the pairwise question embeddings (taken from combinations of 2 sets) range from 42.9% to 93.8% with a combined average of 72.1%. This refers to cosine similarity between the original question in the personality questionnaire and its rephrased version from another set.

Although the cosine similarity between the embeddings of the same question number from different sets might not be an accurate measure, it gives an indication that the BERT model is able to find a semantic similarity between the two questions. It is observed that apart from a couple of outliers most of the questions are answered accurately. Interestingly, we found that some of the questions which seem to have been answered incorrectly have actually been answered using a very close alternative choice. For example instead of answering with the choice 5 the personality emulator answered with the choice 4. This is manifested in the heatmap plot in Figure 9. The Heatmap in Figure 9 shows a distribution of the expected results vs the obtained result given by the DNN. Given that answers to personality questionnaires are not always exact and self-assessments can vary between persons, a tolerance of +/-1 is often considered by psychologists. Therefore, if the DNN comes up with a result that only differs by +/- 1 it can be considered to be correct.

It is worthwhile to note that the NN has been able to answer correctly even those questions which have moderate cosine similarities. This indicates that the personality emulation section of the DNN correctly amplifies or attenuates the significance of various embeddings.

Obtained Value

Expected Value	1	2	3	4	5
1	64.25	15.92	5.18	3.70	10.94
2	3.55	53.59	37.72	4.67	0.47
3	0.42	4.75	87.43	6.01	1.39
4	0.16	10.29	34.29	42.75	12.51
5	11.72	4.22	11.70	9.63	62.72

Figure 9: Distribution of Obtained values vs Expected Values.

5 CONCLUSIONS AND FUTURE WORK

This study explored how personality traits can potentially be emulated by computers using deep neural networks. The results seem promising and confirm that the framework can emulate human personality through answering a psychometric questionnaire . A potential future work might include extending the framework to also answer various other psychometric questionnaires for example, MBTI(Myers Briggs Type Indicator)(Myers, 1962). In the future, such personality emulation system can find commercial applications in chatbots(Lokman and Ameedeen, 2018) which can optimize their communication styles based on customer preferences.

ACKNOWLEDGEMENTS

The authors would like to thank Arndit Ltd.for the computing resources especially GPGPU infrastructure provided for this work.

REFERENCES

Al Moubayed, N., Vazquez-Alvarez, Y., McKay, A., and Vinciarelli, A. (2014). Face-based automatic personality perception. In *Proceedings of the 22nd ACM International Conference on Multimedia*, MM '14, page 1153–1156, New York, NY, USA. Association for Computing Machinery.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26.

Hoppe, S., Loetscher, T., Morey, S. A., and Bulling, A. (2018). Eye movements during everyday behavior predict personality traits. *Frontiers in human neuroscience*, page 105.

HuggingFace (2018). RoBERTa.

Lahiri, A., Mitra, U., Sen, S., Chakraborty, M., Kleiman-Weiner, M., Guha, R., Mitra, P., Basu, A., and Chakraborty, P. P. (2020). Antarjami: Exploring psychometric evaluation through a computer-based game. *arXiv preprint arXiv:2007.10089*.

Levenshtein, V. I. et al. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710. Soviet Union.

Li, M., Chen, X., Li, X., Ma, B., and Vitányi, P. M. (2004). The similarity metric. *IEEE transactions on Information Theory*, 50(12):3250–3264.

Lokman, A. S. and Ameedeen, M. A. (2018). Modern chatbot systems: A technical review. In *Proceedings of the future technologies conference*, pages 1012–1023. Springer.

Mcculloch, W. and Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:127–147.

Mehta, Y., Majumder, N., Gelbukh, A., and Cambria, E. (2020). Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, 53(4):2313–2339.

Myers, I. B. (1962). The myers-briggs type indicator: Manual (1962).

Nvidia (2014). Nvidia K80 Tesla.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084.

Rosenblatt, F. (1958). Rosenblatt solved the problem with his perceptron. *Psychological Review*, 65(6):386–408.

Shum, M., Kleiman-Weiner, M., Littman, M. L., and Tenenbaum, J. B. (2019). Theory of minds: Understanding behavior in groups through inverse planning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6163–6170.

Stillwell, D., Kosinski, M., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*.

Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236):433–460.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, L., Chen, W., Yang, W., Bi, F., and Yu, F. R. (2020). A state-of-the-art review on image synthesis with generative adversarial networks. *IEEE Access*, 8:63514–63537.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.