# Development of a Korean Counseling Chatbot : Data Augmentation and Stable Training with Transformers

**Anonymous Author(s)**
Affiliation
Address
email

## Abstract

This paper presents the development of a Korean counseling chatbot using Transformer-based models. By augmenting and preprocessing the data, a stable training procedure was devised to mitigate overfitting. The results show that the chatbot provides relevant responses to user queries, including those related to comfort and emotional support.

## 1 Introduction

Can AI be a comfort to someone?

### 1.1 Background and Motivation

Mental health issues are becoming increasingly recognized as a critical global concern, and the demand for accessible mental health care is rising. Traditional therapy and counseling are often not easily accessible due to factors such as cost, time, and stigma. With the rapid development of artificial intelligence (AI), counseling chatbots have emerged as promising tools to provide immediate, accessible, and cost-effective support to individuals seeking mental health assistance.

In particular, Korean counseling chatbots face unique challenges due to the linguistic and cultural complexities of the Korean language. Korean is a high-context language, where much of the meaning in conversation relies not only on the literal interpretation of words but also on understanding the emotional tone, context, and subtle nuances. Therefore, creating an AI-powered chatbot that can effectively engage in therapeutic dialogue in Korean is a complex task that requires advanced natural language processing (NLP) techniques.

The margins in 2024 are the same as those in previous years.

### 1.2 Research Objectives

This research focuses on developing a Korean counseling chatbot that can offer effective emotional support while overcoming the linguistic and cultural barriers inherent in the Korean language. The specific objectives of this study are:

1. Data Augmentation: To enhance the quality and diversity of training data, we explore methods to generate synthetic data, addressing the issue of data scarcity for training deep learning models.
2. Transformer-Based Model: We leverage Transformer models, which have demonstrated state-of-the-art performance in various NLP tasks, to train the chatbot. The goal is to ensure the chatbot can generate natural, contextually relevant, and empathetic responses in Korean.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

3. Stable Training: To ensure that the model can be trained effectively without overfitting, we introduce strategies such as regularization and early stopping, ensuring stable learning on small or imbalanced datasets.

By achieving these objectives, we aim to create a counseling chatbot that is capable of providing high-quality support for mental health in Korean, which could be scaled to a broader audience, offering much-needed services for those unable to access traditional forms of therapy.

## 2 Related Works in Counseling Chatbots

The development of counseling chatbots has been an area of significant interest in recent years. Several approaches and studies have contributed to advancing the capabilities of these systems.

1. Early Models and Approaches: One of the earliest and most well-known chatbot models for mental health was ELIZA (Weizenbaum, 1966). ELIZA was a rule-based chatbot that mimicked a Rogerian psychotherapist by reflecting user inputs. While ELIZA was groundbreaking at the time, its lack of deep understanding and empathy led to limited effectiveness. Since then, many other chatbot systems have been developed, incorporating more advanced techniques such as natural language processing (NLP) and machine learning (ML).
2. Machine Learning and Deep Learning Approaches: More recent chatbot systems have utilized machine learning techniques to improve the interaction quality. For example, Woebot, a mental health chatbot based on CBT, uses deep learning to understand and respond to emotional cues in user inputs. Woebot has been shown to be effective in helping users manage symptoms of anxiety and depression, primarily by delivering therapeutic interventions through conversation. These systems often rely on sequence-to-sequence models or transformer-based architectures, which enable them to generate contextually appropriate responses.
3. Transformer-Based Models in Counseling Chatbots: The advent of Transformer models, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformers), has greatly improved the ability of chatbots to understand and generate natural language. Transformer models have achieved state-of-the-art results in many NLP tasks, including text classification, sentiment analysis, and response generation. These models, with their attention mechanisms, can capture long-range dependencies and nuances in conversation, making them highly effective for building emotionally intelligent chatbots capable of offering counseling support.
4. Challenges in Multilingual Counseling Chatbots: The effectiveness of a counseling chatbot is highly influenced by the language in which it operates. For instance, most of the well-known counseling chatbots have been developed in English-speaking contexts, limiting their accessibility for non-English speakers. Research on multilingual and culturally aware chatbots is still in its early stages. In particular, Korean language counseling chatbots face unique challenges due to the complexities of the Korean language, including its rich honorifics system, nuanced emotional expressions, and cultural norms regarding mental health.
5. Empathy and Sentiment Recognition: Several studies have focused on improving the chatbot's ability to recognize and respond to emotional cues in text. For example, EmoReact, a sentiment analysis model, is trained to detect emotions such as happiness, sadness, and anger in textual input. By combining sentiment analysis with natural language generation (NLG), chatbots can tailor their responses to match the emotional state of the user, helping to create a more empathetic and supportive experience.
6. Data Scarcity and Data Augmentation in Counseling Chatbots: A significant challenge in building effective counseling chatbots is the scarcity of domain-specific data. Mental health conversations are private and sensitive, making it difficult to obtain large datasets for training. Several studies have explored data augmentation techniques to overcome this challenge. For example, back-translation and text synthesis methods can be employed to create synthetic training data, thus enriching the chatbot's understanding of a wide range of user inputs.

**Cultural Considerations in Counseling Chatbots** An often overlooked but crucial factor in counseling chatbot development is cultural context. Mental health and emotional expression vary significantly across cultures, and understanding these differences is vital for creating a chatbot that resonates with users.

- **In Korea**, there is a strong societal stigma surrounding mental health issues, which can lead to reluctance in seeking professional help. As a result, counseling chatbots offer a low-stigma alternative for individuals who may feel uncomfortable seeking traditional therapy.
- **Language-specific nuances** in Korean, such as formal and informal speech, emotional expression, and indirect communication styles, further complicate the development of effective chatbots. A Korean chatbot must be able to engage users while respecting these nuances and delivering supportive responses that reflect a deep understanding of the Korean cultural context.

# 3 Model Architecture and Training Methodology

## 3.1 Transformer

In this study, we implemented a Korean chatbot based on the Transformer architecture. The model adopts an Encoder-Decoder structure, with each component designed as follows:

Encoder: After embedding the input sentence, it extracts contextual information through Multi-Head Self-Attention Mechanisms and Position-wise Feedforward Networks.

Decoder: It receives the encoder's output along with previously generated words to predict the next word. The decoder also consists of Multi-Head Self-Attention Mechanisms and Position-wise Feedforward Networks.

This structure enables the model to effectively understand the meaning of input sentences and generate natural Korean responses.

In addition, the Transformer model was trained directly from the training data, without any pre-training model, and achieved stable convergence and overfitting through proper hyperparameter tuning.

## 3.2 Data preprocessing and augmentation

In this study, we used the publicly available Korean dialog dataset for chatbot training.

`https://github.com/songys/Chatbot_data/blob/master/ChatbotData.csv`

This dataset is widely used in Korean language chatbot research for its simplicity and relevance to daily conversation.

**Overview**: The dataset consists of **11,876 pairs of user questions and chatbot responses**. Each data point includes:

- `Q`: User question (input sentence)
- `A`: Chatbot response (target sentence)
- `label`: Category label (0, 1, or 2) representing the type of question

**Characteristics**:

- Contains short, colloquial Korean sentences suitable for training conversational agents.
- The `label` indicates the conversational intent, but our primary focus was on Q–A pairs for generative chatbot modeling.
- Openly available under a permissive license and commonly used as a baseline dataset.

**Data Preprocessing** :

- Special Character Removal: Unnecessary special characters were removed from the text to improve data quality.
- Tokenization : We used SentencePiece to tokenize sentences into subword units, alleviating the rare word problem and improving the model's ability to handle various words. SentencePiece was trained using the Byte-Pair Encoding (BPE) algorithm.

**Data Augmentation** :

- Lexical Substitution : Lexical Substitution involves replacing words in a sentence with semantically similar alternatives. Using word embedding similarity, we analyzed each input sentence and substituted specific words with their nearest neighbors in the embedding space.

# 4 Experimental Results and Performance Evaluation

In this experiment, we trained and evaluated a Korean counseling chatbot based on the Transformer model. We applied lexical substitution as a data augmentation technique to improve the model's generalization ability.

## 4.1 Experiment Setup

The dataset used in the experiment is based on Songys' Chatbotdata, which consists of question-answer pairs. During the data preprocessing phase, the SentencePiece tokenizer was used to tokenize the text, and the Transformer model with embedding and encoder-decoder structure was trained. The training was conducted in a GPU environment, and hyperparameters such as learning rate, batch size, and epoch were optimized for the best performance.

## 4.2 Metrics

The performance of the model was evaluated mainly using accuracy and BLEU score. Accuracy measures how well the chatbot provides appropriate responses to the given inputs, while the BLEU score is commonly used in machine translation tasks to evaluate the quality of generated text.

## 4.3 Training Results

The training lasted for 20 epochs, and the model was evaluated using a validation set. The results showed that the Transformer model gradually improved its performance from the early stages of training, achieving an accuracy of 87 percents and a BLEU score of 0.44. The model was trained with methods such as early stopping to prevent overfitting and ensure stable learning.

## 4.4 Model Performance

The final trained model was able to generate natural and appropriate responses to actual user queries. For example, when the user asked "I am Boring", the chatbot generated an appropriate response like "Is it frustrating" This shows that the model had learned adequately, and the data augmentation technique contributed to improving its generalization performance.

## 4.5 Result Analysis

The experiment confirmed that the Transformer model performed well in text generation and response creation. Moreover, applying data augmentation helped the model handle various contexts and generate more accurate responses. However, there were cases where the chatbot's responses were not perfect, indicating that further improvements in data augmentation techniques and more epochs are necessary in future studies.

# 5 Insights and Future Directions

The Korean consultation chatbot developed in this study is based on the Transformer model, and we found that it is able to provide natural and empathetic responses in line with the user's emotions.

Through data augmentation techniques, the model learned a variety of expressions, which allowed it to generate consistent responses to new inputs. Nevertheless, there are some limitations and future improvements.The Korean consultation chatbot developed in this study is based on the Transformer model, and we found that it is able to provide natural and empathetic responses in line with the user's emotions. Through data augmentation techniques, the model learned a variety of expressions, which allowed it to generate consistent responses to new inputs. Nevertheless, there are some limitations and future improvements.

### 5.1 Limitations

First, we were limited by the size and diversity of the dataset we used in this study, which is approximately 30,000 samples, which may not adequately cover all situations that a counseling chatbot might encounter. For example, we didn't include data from consultations with different age groups or occupations. Also, in situations where the emotional state is unclear, chatbots may struggle to generate appropriate responses.First, we were limited by the size and diversity of the dataset we used in this study, which is approximately 30,000 samples, which may not adequately cover all situations that a counseling chatbot might encounter. For example, we didn't include data from consultations with different age groups or occupations. Also, in situations where the emotional state is unclear, chatbots may struggle to generate appropriate responses.

Second, the quality of the responses our model generates still has room for improvement. Chatbots can sometimes generate responses that are too general or abstract. For example, when a user types, "I'm so overwhelmed," the chatbot might respond with something more formal like, "I know this is hard, but you'll get through it." A response like this might not be appropriate in a real-world consultation, where more personalized and specific feedback is needed.

Third, the model's understanding and ability to handle multi-turn conversations is limited. While the model works by generating an answer to a single question given, real-world counseling often requires handling multiple conversational turns, which means generating responses that take into account multiple interactions with the user, which is not currently implemented in the model.

### 5.2 Future research directions

This study used a basic Transformer model for a counseling chatbot and can suggest various research directions for future improvements and extensions.

- **Expand and diversify dataset**: We can increase the versatility and reliability of model by building a conversation dataset that includes a more diverse user base. In particular, we can improve model's ability to respond more accurately and flexibly by adding data that reflects different emotional states and situations.

- **Handling multi-turn conversations**: In real-world counseling, multiple conversations take place, requiring skills to understand the context of multi-turn conversations and generate appropriate responses. To this end, it is important to develop model architectures and training methods that can reflect the flow of multi-turn conversations in current models. For example, we can extend Transformer's Encoder-Decoder structure to track the flow of a conversation and improve it to generate responses that reflect previous conversations.

- **Personalized feedback system**: To provide more personalized advice tailored to the user's situation and emotions, research is needed to generate customized responses that reflect the user's previous conversations and emotional state. For example, if a user repeatedly expresses a certain emotional state, we can ensure that the model recognizes this and provides personalized advice or feedback accordingly.

- **Improve efficiency**: Current models require large datasets and computational resources. Therefore, to apply chatbots to real-time consultation systems, we need to improve the lightweight and efficiency of the models. We need to study how to maintain high performance with smaller models and explore how they can operate smoothly in mobile or resource-constrained environments.

### 5.3 Concolusions

In this study, we developed a Korean consultation chatbot using the Transformer model, and improved the model performance through data augmentation techniques and stable learning methodology. However, the current model has some limitations, and future research can develop a more improved consultation chatbot through data augmentation, multi-turn conversation processing, and personalized response system. If these studies are carried out, it will be possible to provide more helpful chatbots for people in real consultation systems.

## References

## References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008. *In this foundational paper, the Transformer architecture is introduced, which has become the core model for many natural language processing tasks. The self-attention mechanism and positional encodings proposed in this paper address the shortcomings of previous sequence-based models like RNNs.*

[2] TensorFlow, "Transformer Tutorial," *TensorFlow Documentation*, 2020. [Online]. Available: https://www.tensorflow.org/text/tutorials/transformer?hl=ko *The TensorFlow Transformer tutorial provides a detailed implementation guide for building transformer models for sequence-to-sequence tasks. This resource includes code examples for tokenization, embedding layers, multi-head attention, and model training, making it a helpful starting point for anyone working with transformer-based architectures.*

[3] Google, "Transformer Model in Trax," *Trax GitHub Repository*, 2020. [Online]. Available: https://github.com/google/trax/blob/master/trax/models/transformer.py *The Trax repository by Google offers an efficient implementation of the Transformer model, focusing on both performance and scalability. It includes various optimization techniques that enhance the model's ability to handle large-scale datasets, making it suitable for research and production applications.*

[4] Kyubyong, L., "Word Vectors," *GitHub Repository*, 2020. [Online]. Available: https://github.com/Kyubyong/wordvectors *This GitHub repository provides various pre-trained word vectors and tools for using them in natural language processing tasks. The repository includes vectors trained on large corpora, such as Wikipedia, and allows easy integration into custom NLP projects.*

[5] Jang, H., & Lee, S. H., "A study on the development of a Korean counseling chatbot using deep learning," *Journal of Korean Society for Computer Application*, vol. 19, no. 4, pp. 12-20, 2019. DOI: 10.5392/JKCA.2019.19.04.012. *This paper discusses the development of a Korean counseling chatbot using deep learning techniques. The authors explore the use of NLP models for understanding user input and delivering meaningful responses in a counseling context, which is particularly challenging in Korean due to language and cultural nuances.*

## A Appendix / supplemental material

Optionally include supplemental material (complete proofs, additional experiments and plots) in appendix. All such materials **SHOULD be included in the main submission.**

# NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

**The checklist answers are an integral part of your paper submission.** They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS paper checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: **[TODO]**

   Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: **[TODO]**

    Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: **[TODO]**

Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.