BriLLM: Brain-inspired Large Language Model

Anonymous Author(s)
Affiliation
Address
email

Abstract

This paper reports the brain-inspired large language model (BriLLM). This is a non-Transformer, non-GPT, non-traditional machine learning input-output controlled generative language model. The model is based on the Signal Fully-connected flowing (SiFu) definition on the directed graph in terms of the neural network, and has the interpretability of all nodes on the graph of the whole model, instead of the traditional machine learning model that only has limited interpretability at the input and output ends. In the language model scenario, the token is defined as a node in the graph. A randomly shaped or user-defined signal flow flows between nodes on the principle of "least resistance" along paths. The next token or node to be predicted or generated is the target of the signal flow. As a language model, BriLLM theoretically supports infinitely long n-gram models when the model size is independent of the input and predicted length of the model. The model's working signal flow provides the possibility of recall activation and innate multi-modal support similar to the cognitive patterns of the human brain. At present, we released the first BriLLM versions in Chinese and English, with 4000 tokens, 32-dimensional node size, 32-token sequence prediction ability, model sizes around 2B and 1B respectively, bringing language model prediction performance comparable to GPT-1 [1].

1 Introduction

Large language models (LLMs) are igniting the prospect of AGI (artificial general intelligence). However, even SOTA LLMs are still in terms of Transformer architecture and GPT training scheme unlikely to laugh at the final termination of AGI due to the huge difficulties in their scalability and interpretability, let alone the way Transformer or GPT-based LLM works is a far cry from the human brain, the alternative intelligence machine already existing in nature for millions of years, showing how a true AGI must be.

The Transformer (Vaswani et al., 2017) has been a fundamental and indispensable framework for building SOTA LLM backbones. Although Transformers have demonstrated remarkable generalization capabilities across diverse tasks and scalability to achieve higher intelligence, the quadratic computational complexity of the attention mechanism over input sequences poses significant efficiency challenges, particularly for long sequences. This computational bottleneck has spurred research into more efficient attention variants, such as linear attention mechanisms, and RNN-like Transformers. Although these studies focus on preserving model performance and lowering computational costs, they merely mitigate the issue without resolving the computational bottleneck at its core, since they remain dependent on attention-based mechanisms or attention variants.

[1] We have released our code and models publicly. The links are not disclosed here due to the double-blind review policy.

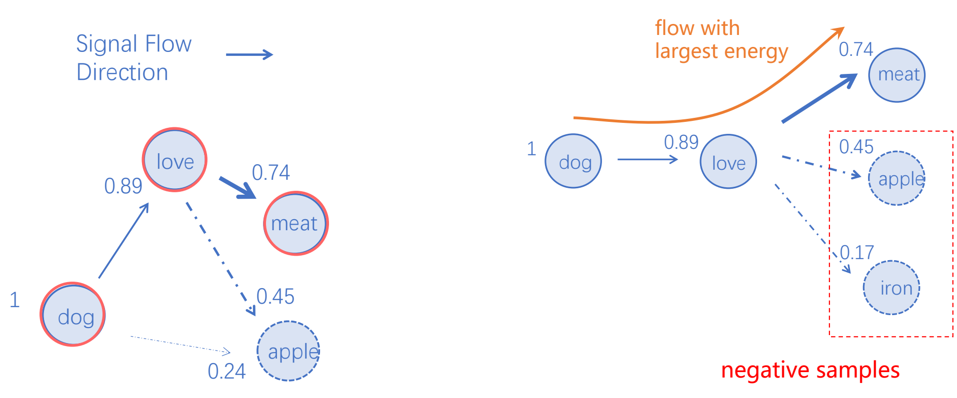
Furthermore, the Transformer architecture exhibits limited parameter-level interpretability due to its complex self-attention mechanisms and opaque parameter interactions, a characteristic that renders it functionally analogous to a black-box system. Many studies attempt to reveal the black box by interpreting the intrinsic mechanism of self-attention or enhancing the interpretability of the model through visualization, attribution methods, and probing tasks. However, the complicated interaction of attention between hidden states remains poorly understood.

To address these challenges, we propose BriLLM, a novel architecture for language modeling that is inspired by signal propagation among neurons in the brain. The BriLLM architecture is structured as a bi-directional graph with multiple nodes and edges. Each node (currently set as a hidden layer of neurons) represents a token, and BriLLM leverages fully-connected neural networks as edges to construct the relationship between these nodes. Like neural signal propagation through biological pathways, BriLLM predicts subsequent tokens by identifying the optimal pathway for energy tensor propagation across nodes. Central to this process is the energy tensor — a dynamic signal representation within BriLLM — which guides the selection of the next node (token). At each step, the model evaluates candidate edges (transitions) and selects the one that maximizes the energy tensor's value, ensuring coherent and contextually relevant token generation.

The proposed mechanism termed Signal Fully-connected Flowing (SiFu) systematically models the entire signal propagation process. This SiFu architecture comprises three core components: (1) a fully-connected directed graph topology where each node maintains bidirectional connections with all other nodes, (2) a dynamic weighting system that modulates signal transmission intensity between nodes based on their functional correlations, and (3) a nonlinear activation module that enables hierarchical relationship extraction during signal propagation.

2 SiFu Mechanism

Inspired by the working mode of the brain, we propose *Signal Fully-connected Flowing (SiFu)* on the Directed Graph, a novel input-output stream control mechanism for machine learning, serving as the core design of BriLLM. As shown in Figure 1a, *SiFu* model is a graph composed of multiple nodes, which are sparsely activated and utilize tensors to transmit a nominal signal. Each node (ideally, a layer of neurons) represents a certain concept or word, e.g., a noun, a verb, etc. Each edge models the relationship between every pair of nodes. The signal is transmitted by the magnitude of the energy. The energy will be strengthened, i.e., maximized, if it is in the right route. Or, at least, the right path always keeps the maximal energy for the transmitted signal. Each node is sequentially activated in terms of the maximized energy. The route or path is determined in a competitive way, i.e., the next node will be activated only if the energy can be maximally delivered in this node.



(a) Forward (b) Optimization

Figure 1: An illustration of SiFu Directed Graph (Numbers by the node denote energy scores).

SiFu model works in a straightforward way, after choosing a series of tokens as input, let a signal continuously transmit from the the beginning node in order, all the tokens represented by each node along the right path that the signal energy keeps the maximal compared to other alternative paths will be collected as the output.

For example, as shown in Figure 1a, the path “dog $\rightarrow$ love $\rightarrow$ meat" has the highest energy. As shown in Figure 1b, the correct sequence should yield the highest energy. For example, to calculate the loss for the sequence “love $\rightarrow$ meat": multiple negative samples in the vocabulary, such as “apple" and “iron," are selected. Energy tensors are computed for both the ground-truth node (“meat") and negative nodes (“apple", “iron"). A chosen loss function maximizes the energy associated with the node “meat" while minimizing energies from the negative nodes.

3 BriLLM Formulation

BriLLM implements *SiFu* neural network for language modeling, as shown in Figure 3. Each token in the vocabulary is modeled as a node, which is defined by a hidden layer of neurons with GeLU activation function and a bias $b \in \mathbb{R}^{d_{node}}$, where d_{node} denotes node size, i.e., how many neuron in a node. An edge connecting nodes u and v is modeled as a fully-connected matrix $W_{u,v} \in \mathbb{R}^{d_{node} \times d_{node}}$. Two fully-connected matrices $W_{u,v}$ and $W_{v,u}$ play the roles of the bidirectional edges between nodes. The signal tensors are fitted into matrices. The forward process begins with an initial signal shape:

$$e_0 = [1, 1, \dots, 1]^\top \in \mathbb{R}^{d_{node}} \tag{1}$$

Suppose we have a token sequence, $u_1, ..., u_{L-1}, v_{predict}$, as a training sample. When the signal flows from a node u_i to its next node u_{i+1}, the energy tensor $e_{i+1} \in \mathbb{R}^{d_{node}}$ will be computed:

$$e_{i+1} = \begin{cases} \mathrm{GeLU}(W_{u_i,u_{i+1}} e_i + b_{u_i,u_{i+1}} + PE_i) & \text{if } \ i > 0 \\ \mathrm{GeLU}(e_0 + b_{u_1} + PE_0) & \text{if } \ i = 0 \end{cases}$$

where PE represents the sine and cosine positional encoding. Note that we have an edge sensitive bias setting for each node taking inputs. When a node starts a sequence, there is no edge difference, i.e., node u_1 has an edge independent bias b_{u_1} in this case.

To predict a token (node), an expanded signal tensor $\mathcal{E}_i \in \mathbb{R}^{d_{node}}$ is computed as a linear weighted sum of previous signals using learnable weights $w \in \mathbb{R}^{L-1}$:

$$\mathcal{W} = \mathrm{softmax}(w_{1:L-1}) \tag{2}$$

$$\mathcal{E}_{L-1} = \sum_{k=1}^{L-1} \mathcal{W}_k e_k, \tag{3}$$

where L is sequence length and $\mathcal{W}$ represents the softmax-normalized weights. The learnable weights w let the predicted token pay “attention" to all previous tokens other than the directly connected one.

At last, the final energy tensor for next token prediction is computed by:

$$E_{u,v} = \mathrm{GeLU}(W_{u_{L-1},v} \mathcal{E}_{L-1} + b_{u_{L-1},v} + PE_{L-1}),$$

During inference, the model finds the right predicted token $v_{predict}$ which has the largest energy:

$$v_{predict} = \arg\max_v \|E_{u,v}\|_2 \tag{4}$$

where the L2 norm of the signal tensor computes its energy score or magnitude.

To train a token sequence sample in BriLLM, every time we build an individual common neural network to perform the regular BP training. This network consists of two parts, in which the front part connects all input nodes (i.e., tokens), then it follows the rear parts which connect all possible paths in order. At last, a softmax layer collects all paths' energy tensors to indicate the right path with a 0-1 ground truth vector. We adopt a cross-entropy loss for training.

4 Experiments

We released BriLLM-Chinese and BriLLM-English models.

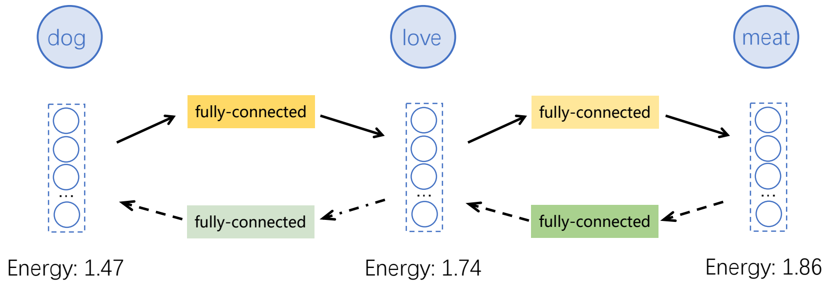


Figure 2: The architecture of BriLLM.

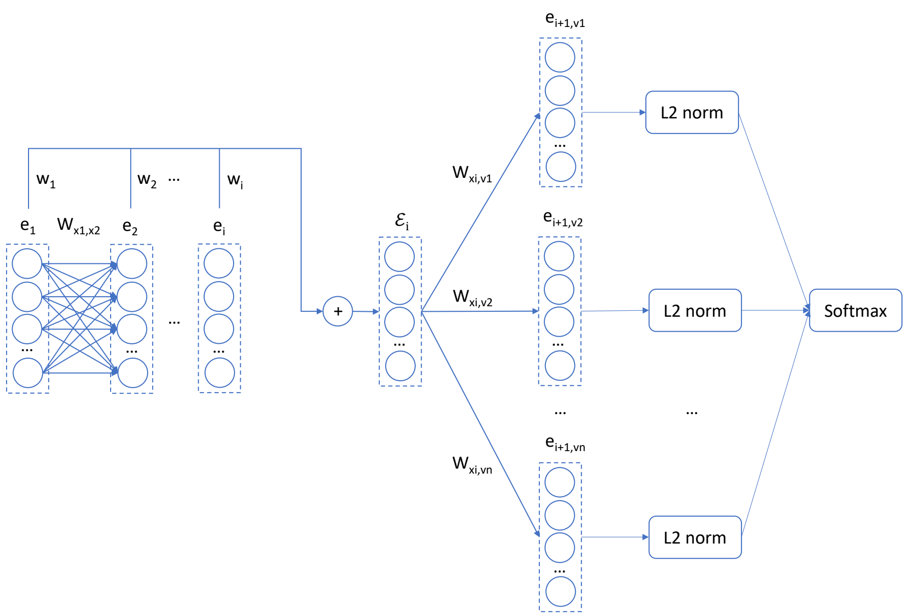


Figure 3: The training network of BriLLM for one training sample .

Datasets For BriLLM-Chinese and BriLLM-English, we use the Chinese and English versions of Wikipedia respectively, each containing over 100M tokens. We truncate the long sentences into small sentences with a maximum length of 32. We select a vocabulary of 4,000 tokens for both languages.

Implementation Details. BriLLM is implemented using PyTorch. It uses sine and cosine positional encoding, GeLU as the activation function, cross-entropy loss for next-token prediction, and a node size of $d_{node} = 32$. We used the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The original model size is about $512 + 4000 * 4000 * (32 * 32 + 32) \approx 16B$. We trained our models on one machine with 8 NVIDIA A800 GPUs for 1.5k steps. The training loss is shown in Figure 4.

Sparse Training BriLLM enables sparse training, where the occurrence probability of most bigrams is very low or even zero, allowing us to leverage this characteristic for sparse training. We set the connection weights corresponding to low-frequency bigrams (those not appearing in the training set) to be shared and update them randomly. After applying sparse training, the actual size of BriLLM-Chinese and BriLLM-English is reduced to 2B and 1B, respectively, as shown in Table 1. This approach reduces the model size to approximately 10% of the original while significantly accelerating the training speed.

Complexity Let L be the sequence length, n the vocabulary size, and d_{node} the node size (dimension), then the forward computational complexity of BriLLM is $O(L \cdot n \cdot d_{node}^2)$.

Figure 4: The training loss.

Table 1: Model sizes before and after sparse training.

	BriLLM-Chinese	BriLLM-English
original	16.90B	16.90B
sparse	2.19B	0.96B
ratio	13.0%	5.7%

Case Study Tables 2 and 3 present some of the decoding results, including both training samples and test samples for Chinese and English, respectively.

5 Conclusion, Limitation and the Future

BriLLM introduces a novel framework for language modeling by replacing attention-based architectures with a brain-inspired dynamic signal propagation mechanism over a fully connected graph. By representing tokens as nodes and leveraging energy tensor dynamics to identify optimal pathways, the model is capable of doing non-autoregressive generation, full node-level interpretability, and theoretically infinite n-gram modeling. Its biologically plausible design decouples model size from sequence length, enabling efficient resource utilization while simulating neurocognitive processes like memory formation. This work challenges the dominance of attention mechanisms, offering a scalable, transparent alternative aligned with neural signaling principles.

Currently, due to our quite limited computational power for this work, we just reach early model checkpoints with a moderate hyperparameter setting. However, the current released models have demonstrated promising performance compared to GPT-1 (Radford et al., 2018).

To precisely understand the SiFu learning mechanism or BriLLM, one must realize that their biggest difference from traditional machine learning is that the former supports multiple concurrent multiple input and multiple output streams, while the latter can only physically accept one input at a time while managing one output. We envision an embodied intelligent implementation of BriLLM, where nominal signals can be multiple, and multiple signal streams can propagate independently along different paths without interference inside the BriLLM, guided by the principle of energy maximization, thereby achieving synchronous multiple inputs and outputs. According to the definition of SiFu learning, this means that each signal stream path represents a pathway of thought. This is very similar to how the human brain synchronously engages in multiple thoughts and multitasks, which is difficultly implemented in terms of traditional single input single output stream of machine learning methods.

In addition to the differences and connections between SiFu learning and traditional machine learning, another interesting topic that naturally arises is whether SiFu learning can be considered a form of representation learning in the context of deep learning. However, it seems difficult to draw such a

Input	Completion
Training samples	
《幽明录》，亦作	《幽明录》，亦作《幽冥录》、《我
《罗马》描述了	《罗马》描述了古罗马从共和国走下
众多哥萨克领导人开始	众多哥萨克领导人开始改变阵营运动
阿根廷探戈是起源于	阿根廷探戈是起源于阿根廷或乌拉圭
惠安第一中学是	惠安第一中学是位于福建省泉州市天
提琴本泛指	提琴本泛指有琴杆，需一手持杆另一
塞人定义为	塞人定义为阿塞拜疆人或南高加索地
袁乃宽旧居是	袁乃宽旧居是清朝及中华民国军事情
飞翔公园站是	飞翔公园站是广州地铁二号线的一。
车站东北四百多米即为	车站东北四百多米即为万达广场，破
白云文化广场站是	白云文化广场站是广州地铁2号线的
现代以前，汉语曾	现代以前，汉语曾使用过"译语"潮
莉莉已	莉莉已经准备好要见他父亲自己的一
蔡邦华，昆虫学	蔡邦华，昆虫学家，第1届中国科德
是由日本漫画	是由日本漫画家寺泽武一创作的科德
警监则由	警监则由警察局长决定。另外两岸气
，他的小提琴	，他的小提琴演奏技巧备受关注。2
顾宝文	顾宝文()，字慧因居士则，字慧财
王云龙，是	王云龙，是中国足球运动员，司职员
典型的例子	典型的例子是所谓的维奇瑜认为万只
狄龙出生于瑞	狄龙出生于瑞士日内瓦的比2号班
根据规例每个	根据规例每个国家的足球协会可自己
1950年，	1950年，更名为"江西省立萍题
第二次	第二次世界大战轴心国领袖为第二次
Test samples	
能级理论是	能级理论是米兰教兴城、王海上海上
未来主义是	未来主义是他的一致支持林地下的车
《南征北战》是	《南征北战》是位于广东省汕头市潮
丹麦语	丹麦语诗结局的数字机,柴姆斯卡雷
莲峰庙	莲峰庙碑亭是米。",设立为那亚州
他也不认为	他也不认为一个地区()是一个地区
卓越工程师	卓越工程师评量大陆的固的选择权—
群众只能够	群众只能够喷嘴能随即在宗,每年去
晚些时候	晚些时候阮惠安岭林斯·罗力发的第
他是	他是日返自行车特的一部,但没有的

Table 2: Case study of BriLLM-Chinese decoding results.

Input	**Completion**
Training samples	
In frogs, the hind legs are larger	In frogs, the hind legs are larger than taxation arrangements and terms, misconstd Paris Academy members of Portals
The requirement for the Sun angle was	The requirement for the Sun angle was arguments from Intr proposed: documentary directed by employing hundreds reduced by employe 11 September 1972
The English biologist Thomas Henry Huxley	The English biologist Thomas Henry Huxley coined World C that ADE XaZul 30 Ars lead singular shipb more smaller im
Physicist Richard Feynman was noted for facility	Physicist Richard Feynman was noted for facility in him increasingly holding six countries, misconstd atomic freedom before
Elements heavier than iron were	Elements heavier than iron were retreatywriter 10th worked (ital magnitude, misconstd atomic Music freedom
Typically, when an algorithm is associated with	Typically, when an algorithm is associated with Achill declaraus, misconceptions presented at Irraditional emotunday Prich
Plants are used as herbs	Plants are used as herbs and Earth Day of Portals working on recent years of Portals working on recent genocots only marked serious risk that
The term vestibular	The term vestibular at Texas variable Spec struggathological ideal remains the division of value of value cannot be supern2
Knight's criticism greatly damaged van	Knight's criticism greatly damaged vanand soon to: examples are 'to looked identity said to: accounts reduced by employe
Atlas-Imperial, an American	Atlas-Imperial, an American Advideo game), December with Achill declar between 2003, misconstd atomic freedom in
Test samples	
The islands have	The islands have been cultivated less than form of value and 1969 via the division of value, miscons lead to non-ane rock
The blue whale (Balaenoptera musculus)	The blue whale (Balaenoptera musculus) order in him responsibility of Portals working on recent gene 11 September 197
The Vincent Price film, House of Wax	The Vincent Price film, House of Waxi theorem approached the sequel strikend across the sequel strikend across
The Jewish Encyclopedia reports, In February	The Jewish Encyclopedia reports, In February 11th worked in him increasingly holds reduced by employe 11 September 1972
The Bermuda Triangle	The Bermuda Triangle, Azerbaijani official letters) markeditors), highest number of Portals working on recent years, misconcept of

Table 3: Case study of BriLLM-English decoding results.

conclusion. Currently, in the implementation of the BriLLM model, the only learnable weight at the most critical node definition is the bias vector b. However, b itself does not carry any motivation for representation learning, because according to the original design of SiFu learning, the role of b is merely to filter the same signal flow into different shapes. Therefore, even if we view the bias vector b as some form of embedded representation for a node like deep learning, it is still a very weak form of representation, far from the strong representation forms that are directly and clearly defined in representation learning.

Our current BriLLM implementation has a size of $(n \times (1 + d_{node}))^2$+$n^2 \times d_{node}$, where n is the number of tokens (nodes) and d_{node} is the node size. This quadratically increasing model size is indeed inconvenience. However, as most model parameters come from the fully-connected matrices, we have shown that it is possible to adopt a sort of sparse representation or shared parameters for those less active tokens, i.e., set a default non-updated matrix for all these inactive tokens. Our empirical results in Table 1 show such strategy may save up to about 90% or more parameters for BriLLM.

Both our BriLLM training practice and the SiFu mechanism show BriLLM is hard to efficiently trained in parallel as every time the training has to be conducted in a different individual neural network. In addition, theoretically accurate training objective needs the right predicted token has to compared its energy to all the other tokens. When token set is large, such ranking may result in a very wide softmax output layer, which further slows the training down and requires much larger training memory. It is lucky that such inconvenience may be alleviated by some sort of approximated ranking strategy. Namely, BriLLM training may be done locally only within those 'necessary' compared counterpart tokens. When all these locally trained networks does not overlap, then all these local network can be trained parallelly, so that the entire BriLLM model training can be done in a good parallel way.

Full model interpretability of BriLLM theoretically facilitates BriLLM to serve as a multi-modal model by nature. Each node in BriLLM does not have to be defined as tokens from languages, they are surely capable of being defined as alternative modal units or jointly defined among different modalities. It is different from LLM, in the case of node-redefinition, no matter one or many nodes, the BriLLM does not need to be re-trained from the very beginning. In one word, the full model interpretability enables BriLLM a natural multi-modal model design, helping the machine learning model closer to the cognition mode as the human brain.

Note that even though BriLLM theoretically supports infinite-gram language model without increasing model size, in practice, the model during training has to cover long enough input sequences, otherwise BriLLM decoding cannot give good enough sequence prediction beyond the training sample length. However, facilitating longer sequence prediction in terms of BriLLM just depends on longer training without resizing the model itself.

So far, we adopt a uniform signal vector like Eq. (1). However, this shape of the signal is not necessary. We tried a randomly initialized signal, the BriLLM can be stably trained. According to the definition of BriLLM, the signal is indeed exploited nominally, however, it may differ the way for activating the input of BriLLM. In the future, we may explore the function of the signal as that of the pre-filled prompt in LLM. If the shape of the signal can be properly used as the primary scenario setting to specify the working of BriLLM, then this should be a much more natural way against in-context learning in the current LLM.

The last but not the least issue we need to explore about BriLLM is the possibility of supervised finetuning (SFT) like LLM. Note that as BriLLM does not need to resize the model for any sized input or output sequences and the size BriLLM has to be quadratically correlated to the node size and token numbers, it is not in an advantageous position when the model sizes are the same 'small' or moderate as LLM. As we reported in this paper, a 1-2B BriLLM (our current released checkpoints) only gives comparable performance as 0.1B GPT-1. Thus, we have reasons to speculate that BriLLM has a very high emergent ability threshold. What's more, now we even do not know how to do SFT over BriLLM, which leaves a big future work.

References

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. *OpenAI blog*, 2018.

Table 4: Comparison of LLM and BriLLM.

	LLM	BriLLM
model size	correlated to input context length	independent
interpretability	only in input & output	all nodes throughout the model
multi-modal implementation	limited to be joined from input/output	all nodes throughout the model

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017. URL `https://arxiv.org/abs/1706.03762`.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

   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: [Yes]

   Justification: We discuss the limitations of the work in Section 5.

   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: [NA]

Justification: The paper does not include theoretical results.

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: [Yes]

   Justification: We disclose the information needed to reproduce the main experimental results in Section 4.

   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: [Yes]

Justification: We will open data and code.

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, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We disclose experimental setting in Section 4.

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: [No]

Justification: We don't report error bars.

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: [Yes]

   Justification: We provide sufficient information in Section 4.

   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: [Yes]

   Justification: The research conforms with the NeurIPS Code of Ethics.

   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: [NA]

    Justification: There is no societal impact of the work performed.

    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: [NA]

    Justification: The paper poses no such risks.

    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: [Yes]

    Justification: These assets are properly credited.

    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: [Yes]

Justification: New assets introduced in the paper are well documented.

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: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    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: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    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.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.