A Broader Impacts and Limitations

A.1 Broader Impacts

Learning before Interaction provides grounded answers to complex multi-agent decision-making problems through the generation of simulators and trial-and-error learning. This can benefit those seeking to make decisions through long-term planning. With significant technological advancements, exploring the use of this technology may be crucial for enhancing existing human decision-making capabilities. For instance, negotiators could describe the opponent's personality traits and their decision-making limits to generate better negotiation strategies.

At the same time, we recognize that current generative simulators still cannot reliably generate state transitions across multiple domains, and learning joint multi-agent strategies still faces convergence difficulties. Therefore, Learning before Interaction may lead to incorrect decisions in specific fields. If humans intentionally follow the generated answers instead of using them as references, it could lead to unsafe or worse consequences. On the other hand, it could also have negative impacts when Learning before Interaction is misused in harmful applications if the generated environments and answers are sufficiently accurate.

A.2 Limitations

Although we have already seen significant improvements in reasoning capabilities for complex multi-agent tasks with Learning before Interaction, performance may be affected by the simulator's accuracy and the multi-agent policy learning performance. Unqualified simulators and difficult-to-converge multi-agent policies may lead to erroneous simulation results, which could be more misleading than the vague answers generated by existing visual language models. For example, the world model has limited out-of-domain generalization for domains that are not represented in the training data, e.g., unseen unit types. Further scaling up training data could help, as the parser can quickly and automatically generate images based on a given state.

While the learned reward functions can enhance the speed of multi-agent policy learning compared to other inverse reinforcement learning and online interaction learning methods, it still requires considerable waiting time to obtain a converged policy and the final answer. Such long waiting time is unacceptable in applications requiring real-time feedback, such as chatbots. One possible solution is to replace multi-agent reinforcement learning with planning methods based on the learned rewards and dynamics models, thereby accelerating the reasoning process. We will leave this issue in future work.

In addition, this paper is confined to scenarios within the game StarCraft II. This is an environment that, while complex, cannot represent the dynamics of all multi-agent tasks. Evaluation of multi-agent reinforcement learning algorithms, therefore, should not be limited to one benchmark but should target a variety with a range of tasks.

Map Name	Return Distribution	Map Name	Return Distribution
3s5z	19.43 ± 1.86	5m_vs_6m	19.83 ± 2.16
1c3s5z	19.66 ± 1.25	6h_vs_8z	18.84 ± 2.09
10m_vs_11m	19.75 ± 1.03	3s5z_vs_3s6z	19.76 ± 1.26
2c_vs_64zg	19.98 ± 0.71	corridor	19.69 ± 1.48
3s_vs_5z	19.88 ± 1.40	MMM2	19.63 ± 2.07

Table 6: Return distribution on training maps.

B Dataset Preparation

The training maps include 3s5z, 1c3s5z, 10m_vs_11m, 2c_vs_64zg, 3s_vs_5z, 5m_vs_6m, 6h_vs_8z, 3s5z_vs_3s6z, corridor, MMM2 in StarCraft Multi-Agent Challenge (SMAC) (Samvelyan et al., 2019). We use EMC (Zheng et al., 2021) and IIE (Liu et al., 2024) to collect 50000 trajectories

18

for each map and save these data as NPY files. The data includes the states, the observations, the terminated signals, the actions, the available actions, and the rewards. The return distribution on training maps is shown in Table 6. The average return is 19.64 ± 1.63 across ten training maps.

In Figure 6, we have presented the whole procedure of converting a state vector into an image for simulation and parsing a trajectory to produce a textual task description. First, as shown in Figure 5, we collect the element images that appear in the game and affect the state, including units and background terrains of training maps.

Figure 5: Images of units and terrains.
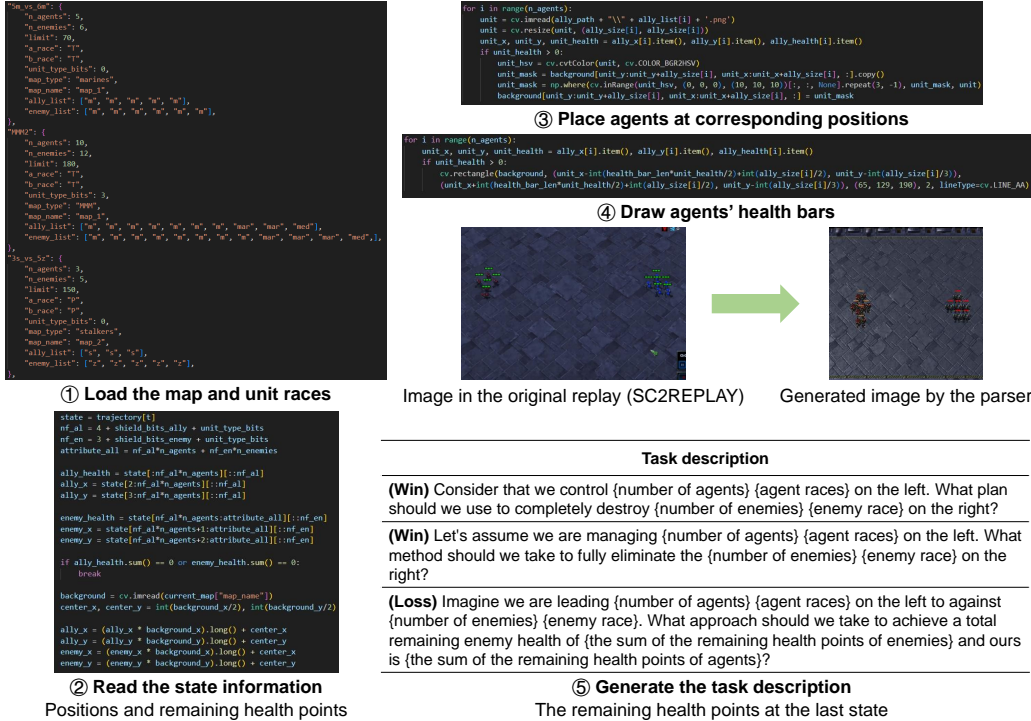
Figure 6: The whole pipeline of how the parser generates the image and the task description for a given state. Here, we only show three task descriptions the parser produces for demo purposes.

19

C StarCraft Multi-agent Challenge

StarCraft II is a real-time strategy game featuring three different races, Protoss, Terran, and Zerg, with different properties and associated strategies. The objective is to build an army powerful enough to destroy the enemy's base. When battling two armies, players must ensure army units are acting optimally. StarCraft Multi-Agent Challenge (SMAC) (Samvelyan et al., 2019) is a partially observable reinforcement learning benchmark built in StarCraft II. An individual agent with parameter sharing controls each allied unit, and a hand-coded built-in StarCraft II AI controls enemy units. The difficulty of the game AI is set to the "very difficult" level.

On the SMAC benchmark, agents can access their local observations within the field of view at each time step. The feature vector contains attributes of both allied and enemy units: `distance`, `relative x`, `relative y`, `health`, `shield`, and `unit_type`. In addition, agents can observe the last actions of allied units and the terrain features surrounding them. The global state vector includes the coordinates of all agents relative to the center of the map and other features present in the local observation of agents. The state stores the energy of Medivacs, the cooldown of the rest of the allied units, and the last actions of all agents. Note that the global state information is only available to agents during centralized training. All features in state and local observations are normalized by their maximum values. After receiving the observations, each agent is allowed to take action from a discrete set which consists of `move[direction]`, `attack[enemy_id]`, `stop` and `no-op`. Move direction includes north, south, east, and west. Note that the dead agents can only take `no-op` action while live agents cannot. For health units, Medivacs use `heal[agent_id]` actions instead of `attack[enemy_id]`.

Depending on different scenarios, the maximum number of actions varies between 7 and 70. Note that agents can only perform the `attack[enemy_id]` action when the enemy is within its shooting range. At each time step, agents take joint action and receive a positive global reward based on the total damage dealt to the enemy units. In addition, they can receive an extra reward of 10 points after killing each enemy unit and 200 points after killing all enemy units. The rewards are scaled to around 20, so the maximum cumulative reward is achievable in each scenario.

D Experiment Setting

In this section, we describe the ground-truth environment that agents interact, the implementation details of online learning methods, offline learning methods, and our model Learning before Interaction.

D.1 Online Learning

We adopt the same architectures for QMIX, QPLEX, CW-QMIX[1], RODE[2], MAVEN[3], EMC[4] as their official implementations (Samvelyan et al., 2019; Wang et al., 2020a; Rashid et al., 2020; Wang et al., 2020c; Mahajan et al., 2019; Zheng et al., 2021). Each agent independently learns a policy with fully shared parameters between all policies. We used RMSProp with a learning rate of 5e-4 and $\gamma = 0.99$, buffer size 5000, and mini-batch size 32 for all algorithms. The dimension of each agent's GRU hidden state is set to 64.

For our experiments, we employ an ϵ-greedy exploration scheme for the joint policy, where ϵ decreases from 1 to 0.05 over 1 million timesteps in `6h_vs_8z`, `3s5z_vs_3s6z` and `corridor`, and over 50 thousand timesteps in other maps. The implementation of MAPPO is consistent with their official repositories[5] (Yu et al., 2022). As shown in Table 7, all hyperparameters are left unchanged at the origin best-performing status. For CW-QMIX, the weight for negative samples is set to $\alpha = 0.5$ for all scenarios.

---

[1]https://github.com/oxwhirl/wqmix
[2]https://github.com/TonghanWang/RODE
[3]https://github.com/AnujMahajanOxf/MAVEN
[4]https://github.com/kikojay/EMC
[5]https://github.com/zoeyuchao/mappo

20

Hyperparameter	Value	Hyperparameter	Value
critic lr	5e-4	actor lr	5e-4
ppo epoch	5	ppo-clip	0.2
optimizer	Adam	batch size	3200
optim eps	1e-5	hidden layer	1
gain	0.01	training threads	32
rollout threads	8	γ	0.99
hidden layer dim	64	activation	ReLU

Table 7: Hyper-parameters in MAPPO.

All figures in online learning experiments are plotted using mean and standard deviation with confidence internal 95%. We conduct five independent runs with different random seeds for each learning curve.

D.2 Offline Learning

We adopt the same architectures for MA-AIRL[6], MADT[7], MAPT[8], ICQ[9], OMAR[10], and OMIGA[11] as their official implementations (Yu et al., 2019; Meng et al., 2023; Zhu et al., 2024; Fujimoto et al., 2019; Kumar et al., 2020; Yang et al., 2021; Pan et al., 2022; Wang et al., 2024). We implement MA-TREX, BCQ-MA and CQL-MA based on TREX (Brown et al., 2019), BCQ (Fujimoto et al., 2019), and CQL (Kumar et al., 2020), respectively. In particular, we add the task description into MADT's target sequence because it deprecates the reward-to-go term.

D.3 Learning before Interaction

We train our image tokenizer for 100k steps using the AdamW optimizer, with cosine decay, using the hyperparameters in Table 8. The batch size is 32, and the learning rate is 1e-4.

Component	Hyperparameter	Value
Encoder	num_layers	5e-4
	num_res_layers	2
	num_channels	(256,256)
	num_res_channels	(256,256)
	downsample	(2,4,1,1)
Decoder	num_layers	5e-4
	num_res_layers	2
	num_channels	(256,256)
	num_res_channels	(256,256)
	upsample	(2,4,1,1,0)
Codebook	num_codes	256
	latent_dim	32
	commitment_cost	0.25

Table 8: Hyper-parameters in VQ-VAE.

We build our dynamics model implementation based on Decision Transformer[12] (Chen et al., 2021). The complete list of hyperparameters can be found in Table 9. The dynamics models were trained using the AdamW optimizer.

---

[6]https://github.com/ermongroup/MA-AIRL

[7]https://github.com/ReinholdM/Offline-Pre-trained-Multi-Agent-Decision-Transformer

[8]https://github.com/catezi/MAPT

[9]https://github.com/YiqinYang/ICQ

[10]https://github.com/ling-pan/OMAR

[11]https://github.com/ZhengYinan-AIR/OMIGA

[12]https://github.com/kzl/decision-transformer

21

Hyperparameter	Value	Hyperparameter	Value
number of layers	6	grad norm clip	1.0
attention heads	8	weight decay	0.1
embedding dims	64	Adam betas	(0.9,0.95)

Table 9: Hyperparameters in the transformer model.

The reward shares the same architecture as the dynamics model, but the attention mask in the transformer model is modified in order to receive the whole trajectory as input rather than the tokens that have come before the current one. Here are some tricks for reward learning: (1) we control the gap between the rewards of the expert behavior and the policy action - we stop the gradient for the reward of the expert behavior at a given state if it is greater than the one of the policy action, where $beta$ is the margin and set to 2; (2) we also set the target of unavailable actions' rewards to 0; (3) we alternate between k-step of policy update and reward update to avoid completely solving the policy optimization subproblem before updating the reward parameters, where $k = 5$.

In this paper, all experiments are implemented with Pytorch and executed on eight NVIDIA A800 GPUs.

E   Additional Results

Using a Text-to-Code Converter can generate scenarios with the original game engine and then learn the joint policy. Therefore, we also consider the comparison with online MARL methods including CW-QMIX (Rashid et al., 2020), QPLEX (Wang et al., 2020a), MAVEN (Mahajan et al., 2019), EMC (Zheng et al., 2021), RODE (Wang et al., 2020c), QMIX (Rashid et al., 2018), MAPPO (Yu et al., 2022). Figure 7 demonstrates a significant improvement in the sample efficiency of LBI compared to the online MARL methods, suggesting that a pre-trained world model is necessary to reduce the waiting time for generating grounded answers for multi-agent decision-making problems.
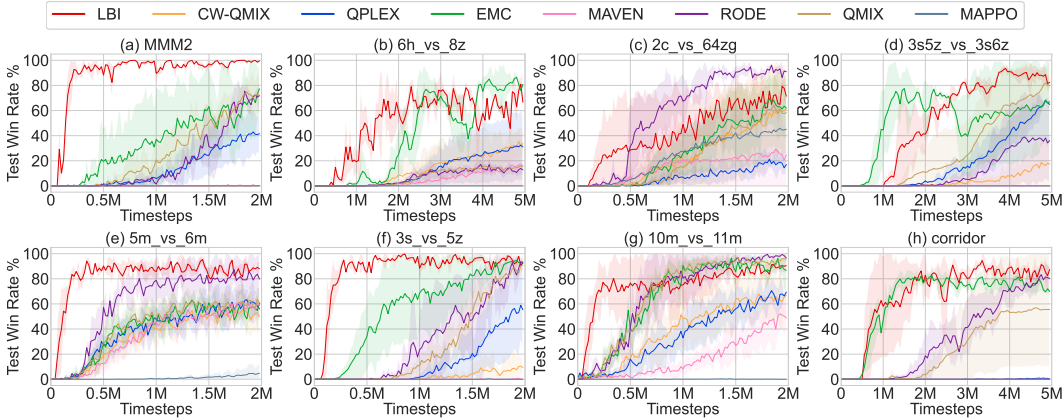
Figure 7: Performance comparisons between online learning methods using ground-truth rewards on the SMAC benchmark and LBI using the learned reward functions on the imagined world model.

In addition, we also show the qualitative comparison between the target and the generated sequences in Figure 8. Both trajectories are collected by running the same policy. We can see that the generated sequence can resemble the target one in most frames, but some differences exist in positions and health bars. However, compounding errors in the single-step model, which lead to physically implausible predictions, are not observed in the dynamics model generated by the causal transformer. For example, at the timestep of 10 in the MMM2 scenario, the generated frame does not contain the ally's Medivac, but we can see it in the following frames.

22
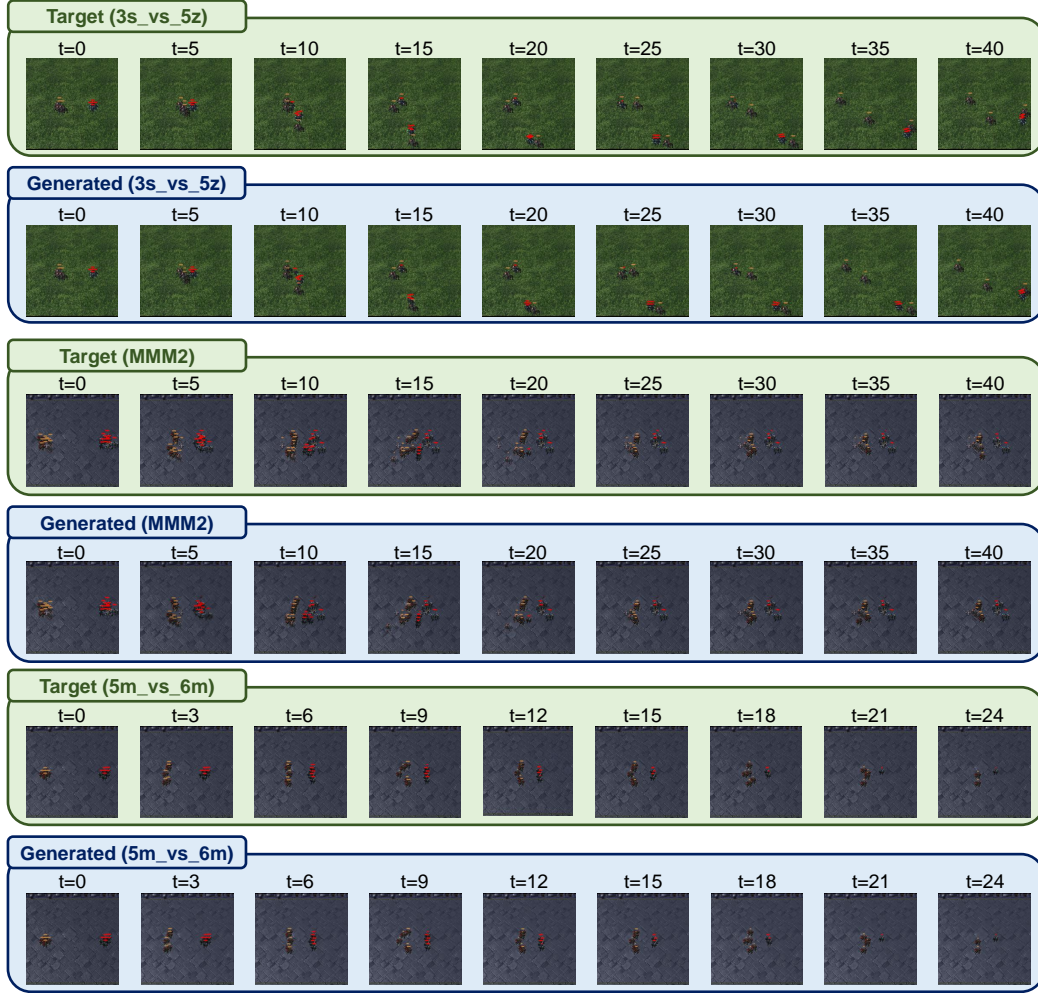
Figure 8: Comparisons of the target and the generated sequences across three different maps.

F Additional Related Work

Offline Q-Learning Offline Q-learning learns a policy from a fixed dataset where the reward is provided for each transition sample. Most off-policy reinforcement learning (RL) algorithms are applicable in offline Q-learning. However, they typically suffer from the overestimation problem of out-of-distribution (OOD) actions due to the distribution shift between the action distribution in the training dataset and that induced by the learned policy (Fujimoto et al., 2019). Several constraint methods are proposed to restrict the learned policy from producing OOD actions by leveraging importance sampling (Sutton et al., 2016; Nachum et al., 2019), incorporating explicit policy constraints (Kostrikov et al., 2021; Fakoor et al., 2021; Fujimoto & Gu, 2021; Tarasov et al., 2024), penalizing value estimates (Kumar et al., 2020; An et al., 2021; Shao et al., 2024), and uncertainty quantification (Wu et al., 2021; Zanette et al., 2021). Another branch resorts to learning without querying OOD actions and thus constrain the learning process within the support of the dataset (Bai et al., 2021; Lyu et al., 2022).

Transformer Model Several works have explored the integration of transformer models into reinforcement learning (RL) settings. We classify them into two major categories depending on the usage pattern. The first category focuses on representing components in RL algorithms, such as policies and value functions (Parisotto et al., 2020; Parisotto & Salakhutdinov, 2021). These methods rely on standard RL algorithms to update policy, where the transformer only provides a large representation capacity and improves feature extraction. Conversely, the second category aims to

23

replace the RL pipeline with sequence modeling. They autoregressively generate states, actions, and rewards by conditioning on the desired return-to-go during inference (Chen et al., 2021; Lee et al., 2022; Reed et al., 2022). Due to its simplicity and potential generalization ability, this category is widely used in various domains, such as robotics control (Brohan et al., 2023a; Padalkar et al., 2023; Driess et al., 2023) and multi-agent reinforcement learning (Meng et al., 2023; Liu et al., 2024).

Multi-agent Reinforcement Learning   This section briefly introduces recent related work on cooperative multi-agent reinforcement learning (MARL). In the paradigm of centralized training with decentralized execution (CTDE), agents' policies are trained with access to global information in a centralized way and executed only based on local histories in a decentralized way (Oliehoek et al., 2008; Kraemer & Banerjee, 2016). One of the most significant challenges in CTDE is to ensure the correspondence between the individual Q-value functions and the joint Q-value function Q_{tot}, i.e., the Individual-Global Max (IGM) principle (Son et al., 2019). VDN (Sunehag et al., 2018) and QMIX (Rashid et al., 2018) learn the joint Q-values and factorize them into individual Q-value functions in an additive and a monotonic fashion, respectively. Several works (Yang et al., 2020b,a; Wang et al., 2020b,c) have been proposed to improve the performance of QMIX, but as many previous studies pointed out, monotonic value function factorization limits the representational capacity of Q_{tot} and fails to learn the optimal policy when the target Q-value functions are non-monotonic (Mahajan et al., 2019; Son et al., 2019; Rashid et al., 2020). To solve this problem, some recent works (Wang et al., 2020a; Mahajan et al., 2021) try to achieve the full representational capacity of Q_{tot}, while others prioritize the potential optimal joint action and learn a biased Q_{tot}.

Some independent learning algorithms have also proven robust in solving multi-agent cooperative tasks. Distributed Q-learning (Lauer, 2000) and Hysteretic Q-learning (Matignon et al., 2007) place more importance on positive updates that increase a Q-value estimate, which is similar to the weighting function in WQMIX. However, Wei & Luke (2016) prove that these methods are vulnerable towards misleading stochasticity and propose LMRL2, where agents forgive the other's miscoordination in the initial exploration phase but become less lenient when the visitation of state-action pair increases. MAPPO (Yu et al., 2022) applies PPO (Schulman et al., 2017) into MARL and shows strong empirical performance. However, Kuba et al. (2021) points out MAPPO suffers from instability arising from the non-stationarity induced by simultaneously learning and exploring agents. Therefore, they introduce the sequential policy update scheme to achieve monotonic improvement on the joint policy.

Learning communication protocols to solve cooperative tasks is one of the desired emergent behaviors of agent interactions. It has recently become an active area in MARL, such as learning to share observations (Das et al., 2019; Wang et al., 2019; Liu et al., 2020) and intentions (Kim et al., 2020; Böhmer et al., 2020; Wen et al., 2022; Liu et al., 2023).

24

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 abstract and introduction include the contributions made in the paper. See Section 1 for more information.

   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 this work in Appendix A.2, such as limited out-of-domain generalization and considerable cost time.

   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?

25

Answer: [NA]

Justification: NA.

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 describe the steps taken to construct the dataset in Appendix B, and the implementation details of our model and baselines in Appendix D.

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**

26

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: [No]

Justification: We choose not to release the data and code at present. We would like to have the opportunity to further engage with the research community and to ensure that any future such releases are respectful, safe, and responsible.

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: [Yes]

Justification: We provide the training and test details in Appendix D.

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: [Yes]

Justification: We conduct five independent runs with different random seeds for each result. The results are accompanied by standard deviations.

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.

27

- 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 the information on the computer resources in Appendix D.

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: We conduct the research 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: [Yes]

Justification: We discuss both potential positive societal impacts and negative societal impacts in Appendix A.1.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

28

- 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: [Yes]

    Justification: We does not use a pre-trained model (which may generate unsafe images), and we construct the image dataset through a parser. See Appendix B for more information.

    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: We cite the original paper and provide the URLs for the assets in Appendix D.

    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.

29

- 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: [NA]

    Justification: We describe the steps taken to construct the dataset in Appendix B, and the implementation details of our model and baselines in Appendix D. However, We choose not to release the data and code at present. We would like to have the opportunity to further engage with the research community and to ensure that any future such releases are respectful, safe and responsible.

    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: NA.

    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: NA.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

30

- 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.

31