

906 **NeurIPS Paper Checklist**

907 **1. Claims**

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

910 Answer: [\[Yes\]](#)

911 Justification: As shown in Section 1.

912 Guidelines:

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

922 **2. Limitations**

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

924 Answer: [\[Yes\]](#)

925 Justification: The limitations of the work are shown in Section 6.3.

926 Guidelines:

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

953 **3. Theory Assumptions and Proofs**

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

956 Answer: [\[Yes\]](#)

957 Justification: As shown in Section 3.1 and Appendix E.

958 Guidelines:

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

#### 969 4. Experimental Result Reproducibility

970 Question: Does the paper fully disclose all the information needed to reproduce the main ex-

971 perimental results of the paper to the extent that it affects the main claims and/or conclusions

972 of the paper (regardless of whether the code and data are provided or not)?

973 Answer: [Yes]

974 Justification: We provide the dataset details in Appendix D and implementation details in F

975 to reproduce the main experimental results.

976 Guidelines:

- 977 • The answer NA means that the paper does not include experiments.
- 978 • If the paper includes experiments, a No answer to this question will not be perceived
- 979 well by the reviewers: Making the paper reproducible is important, regardless of
- 980 whether the code and data are provided or not.
- 981 • If the contribution is a dataset and/or model, the authors should describe the steps taken
- 982 to make their results reproducible or verifiable.
- 983 • Depending on the contribution, reproducibility can be accomplished in various ways.
- 984 For example, if the contribution is a novel architecture, describing the architecture fully
- 985 might suffice, or if the contribution is a specific model and empirical evaluation, it may
- 986 be necessary to either make it possible for others to replicate the model with the same
- 987 dataset, or provide access to the model. In general, releasing code and data is often
- 988 one good way to accomplish this, but reproducibility can also be provided via detailed
- 989 instructions for how to replicate the results, access to a hosted model (e.g., in the case
- 990 of a large language model), releasing of a model checkpoint, or other means that are
- 991 appropriate to the research performed.
- 992 • While NeurIPS does not require releasing code, the conference does require all submis-
- 993 sions to provide some reasonable avenue for reproducibility, which may depend on the
- 994 nature of the contribution. For example
- 995 (a) If the contribution is primarily a new algorithm, the paper should make it clear how
- 996 to reproduce that algorithm.
- 997 (b) If the contribution is primarily a new model architecture, the paper should describe
- 998 the architecture clearly and fully.
- 999 (c) If the contribution is a new model (e.g., a large language model), then there should
- 1000 either be a way to access this model for reproducing the results or a way to reproduce
- 1001 the model (e.g., with an open-source dataset or instructions for how to construct
- 1002 the dataset).
- 1003 (d) We recognize that reproducibility may be tricky in some cases, in which case
- 1004 authors are welcome to describe the particular way they provide for reproducibility.
- 1005 In the case of closed-source models, it may be that access to the model is limited in
- 1006 some way (e.g., to registered users), but it should be possible for other researchers
- 1007 to have some path to reproducing or verifying the results.

#### 1008 5. Open access to data and code

1009 Question: Does the paper provide open access to the data and code, with sufficient instruc-

1010 tions to faithfully reproduce the main experimental results, as described in supplemental

1011 material?

1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063

Answer: [Yes]

Justification: we have released the code and experiment setting details in our supplemental material.

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 dataset details in Appendix D, implementation details in F and hyperparameter details in Section 4 and Appendix F.3.

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 report the average performance of 5 different random seeds for finetuning procedures, as shown in Section 5.2, 5.3, Figure 5 and Table 4. Besides, we report the average performance when merging different numbers of tasks, as shown in Appendix C.1 and Table 8.

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).

- 1064 • The method for calculating the error bars should be explained (closed form formula,  
1065 call to a library function, bootstrap, etc.)
- 1066 • The assumptions made should be given (e.g., Normally distributed errors).
- 1067 • It should be clear whether the error bar is the standard deviation or the standard error  
1068 of the mean.
- 1069 • It is OK to report 1-sigma error bars, but one should state it. The authors should  
1070 preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis  
1071 of Normality of errors is not verified.
- 1072 • For asymmetric distributions, the authors should be careful not to show in tables or  
1073 figures symmetric error bars that would yield results that are out of range (e.g. negative  
1074 error rates).
- 1075 • If error bars are reported in tables or plots, The authors should explain in the text how  
1076 they were calculated and reference the corresponding figures or tables in the text.

## 1077 8. Experiments Compute Resources

1078 Question: For each experiment, does the paper provide sufficient information on the com-  
1079 puter resources (type of compute workers, memory, time of execution) needed to reproduce  
1080 the experiments?

1081 Answer: [Yes]

1082 Justification: As shown in Appendix F.1.

1083 Guidelines:

- 1084 • The answer NA means that the paper does not include experiments.
- 1085 • The paper should indicate the type of compute workers CPU or GPU, internal cluster,  
1086 or cloud provider, including relevant memory and storage.
- 1087 • The paper should provide the amount of compute required for each of the individual  
1088 experimental runs as well as estimate the total compute.
- 1089 • The paper should disclose whether the full research project required more compute  
1090 than the experiments reported in the paper (e.g., preliminary or failed experiments that  
1091 didn't make it into the paper).

## 1092 9. Code Of Ethics

1093 Question: Does the research conducted in the paper conform, in every respect, with the  
1094 NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

1095 Answer: [Yes]

1096 Justification: This research is conducted in the paper conform, with the NeurIPS Code of  
1097 Ethics.

1098 Guidelines:

- 1099 • The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- 1100 • If the authors answer No, they should explain the special circumstances that require a  
1101 deviation from the Code of Ethics.
- 1102 • The authors should make sure to preserve anonymity (e.g., if there is a special consid-  
1103 eration due to laws or regulations in their jurisdiction).

## 1104 10. Broader Impacts

1105 Question: Does the paper discuss both potential positive societal impacts and negative  
1106 societal impacts of the work performed?

1107 Answer: [Yes]

1108 Justification: As shown in Section 1 and Appendix A.

1109 Guidelines:

- 1110 • The answer NA means that there is no societal impact of the work performed.
- 1111 • If the authors answer NA or No, they should explain why their work has no societal  
1112 impact or why the paper does not address societal impact.

- 1113 • Examples of negative societal impacts include potential malicious or unintended uses  
1114 (e.g., disinformation, generating fake profiles, surveillance), fairness considerations  
1115 (e.g., deployment of technologies that could make decisions that unfairly impact specific  
1116 groups), privacy considerations, and security considerations.
- 1117 • The conference expects that many papers will be foundational research and not tied  
1118 to particular applications, let alone deployments. However, if there is a direct path to  
1119 any negative applications, the authors should point it out. For example, it is legitimate  
1120 to point out that an improvement in the quality of generative models could be used to  
1121 generate deepfakes for disinformation. On the other hand, it is not needed to point out  
1122 that a generic algorithm for optimizing neural networks could enable people to train  
1123 models that generate Deepfakes faster.
- 1124 • The authors should consider possible harms that could arise when the technology is  
1125 being used as intended and functioning correctly, harms that could arise when the  
1126 technology is being used as intended but gives incorrect results, and harms following  
1127 from (intentional or unintentional) misuse of the technology.
- 1128 • If there are negative societal impacts, the authors could also discuss possible mitigation  
1129 strategies (e.g., gated release of models, providing defenses in addition to attacks,  
1130 mechanisms for monitoring misuse, mechanisms to monitor how a system learns from  
1131 feedback over time, improving the efficiency and accessibility of ML).

## 1132 11. Safeguards

1133 Question: Does the paper describe safeguards that have been put in place for responsible  
1134 release of data or models that have a high risk for misuse (e.g., pretrained language models,  
1135 image generators, or scraped datasets)?

1136 Answer: [Yes]

1137 Justification: As shown in Appendix D.

1138 Guidelines:

- 1139 • The answer NA means that the paper poses no such risks.
- 1140 • Released models that have a high risk for misuse or dual-use should be released with  
1141 necessary safeguards to allow for controlled use of the model, for example by requiring  
1142 that users adhere to usage guidelines or restrictions to access the model or implementing  
1143 safety filters.
- 1144 • Datasets that have been scraped from the Internet could pose safety risks. The authors  
1145 should describe how they avoided releasing unsafe images.
- 1146 • We recognize that providing effective safeguards is challenging, and many papers do  
1147 not require this, but we encourage authors to take this into account and make a best  
1148 faith effort.

## 1149 12. Licenses for existing assets

1150 Question: Are the creators or original owners of assets (e.g., code, data, models), used in  
1151 the paper, properly credited and are the license and terms of use explicitly mentioned and  
1152 properly respected?

1153 Answer: [Yes]

1154 Justification: As shown in Section 5.1 and Appendix D.

1155 Guidelines:

- 1156 • The answer NA means that the paper does not use existing assets.
- 1157 • The authors should cite the original paper that produced the code package or dataset.
- 1158 • The authors should state which version of the asset is used and, if possible, include a  
1159 URL.
- 1160 • The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- 1161 • For scraped data from a particular source (e.g., website), the copyright and terms of  
1162 service of that source should be provided.
- 1163 • If assets are released, the license, copyright information, and terms of use in the package  
1164 should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has  
1165 curated licenses for some datasets. Their licensing guide can help determine the license  
1166 of a dataset.

- 1167
- 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.
- 1168
- If this information is not available online, the authors are encouraged to reach out to the asset’s creators.
- 1169
- 1170

1171 **13. New Assets**

1172 Question: Are new assets introduced in the paper well documented and is the documentation  
1173 provided alongside the assets?

1174 Answer: [Yes]

1175 Justification: As shown in Section 5.1 and Appendix F.2.

1176 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.
- 1177
- 1178
- 1179
- 1180
- 1181
- 1182
- 1183
- 1184

1185 **14. Crowdsourcing and Research with Human Subjects**

1186 Question: For crowdsourcing experiments and research with human subjects, does the paper  
1187 include the full text of instructions given to participants and screenshots, if applicable, as  
1188 well as details about compensation (if any)?

1189 Answer: [No]

1190 Justification: This paper does not involve crowdsourcing nor research with human subjects.

1191 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.
- 1192
- 1193
- 1194
- 1195
- 1196
- 1197
- 1198
- 1199

1200 **15. Institutional Review Board (IRB) Approvals or Equivalent for Research with Human  
1201 Subjects**

1202 Question: Does the paper describe potential risks incurred by study participants, whether  
1203 such risks were disclosed to the subjects, and whether Institutional Review Board (IRB)  
1204 approvals (or an equivalent approval/review based on the requirements of your country or  
1205 institution) were obtained?

1206 Answer: [No]

1207 Justification: This paper does not involve crowdsourcing nor research with human subjects.

1208 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.
- 1209
- 1210
- 1211
- 1212
- 1213
- 1214
- 1215
- 1216
- 1217
- 1218