

Corresponding author(s): Tzu Tsun Luk
 Last updated by author(s): 29-05-2026

Machine Learning Checklist v 1.1

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form is intended to provide structure for consistency and transparency in reporting of works using or developing Machine Learning models. Some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

1. Availability and reproducibility of Code and Data

Please select all that apply regarding the availability of the data and code used in the study.

- ☐ Code will be included in a CodeOcean capsule.
- ☐ The **source code** is included in the submission or available in a public repository:

- ☐ A **compiled standalone version** of the software is included in the submission or available in a public repository:

- ☐ A **test dataset** and instructions/scripts for replicating the results are included in the submission or available in a public repository:

- ☐ A **Readme file** with instructions for installing and running the code is included in the submission or available in a public repository:

- ☐ The code is made available to reviewers during review.
- ☒ **Pretrained models** are used in the study and accessible through:
<https://oai.azure.com/resource/models/gpt-4o/version/2024-11-20/registry/azure-openai>

- ☐ **Pretrained models** are used in the study and are not accessible.
- ☐ The paper contains information on how to obtain code and data after publication.

2. Datasets

- A. All data sources are listed in the paper.
 - ☐ Yes
 - ☒ No

- B. The train, test and validation datasets are publicly available, and links/accession numbers have been provided in the manuscript or supplementary materials.
 - ☐ Yes
 - ☒ No

- C. We have reported and discussed potential dataset biases in the paper. Where applicable, appropriate mitigation strategies were used.
- ☐ Yes _____
- ☒ No Not applicable as a publicly available pretrained commercial large language model was used
- D. The data cleaning and preprocessing steps are clearly and fully described, either in text or as a code pipeline.
- ☐ Yes _____
- ☒ No Not applicable as a publicly available pretrained commercial large language model was used
- E. Instances of combining data from multiple sources are clearly identified, and potential issues mitigated.
- ☐ Yes _____
- ☒ No Not applicable as a publicly available pretrained commercial large language model was used

3. Model and training

- A. What model architecture is the current model based on? Transformer (Azure OpenAI GPT-4o)
- B. A Model Card is provided¹.
- ☐ Yes
- ☒ No
- C. The model clearly splits data into different sets for training (model selection), validation (hyperparameter optimization), and testing (final evaluation).
- ☐ Yes
- ☒ No
- D. The method of data splitting (e.g. random, cluster- or time-based splitting, forward cross-validation) is clearly stated.
- ☐ Yes _____
- ☒ No Not applicable as a publicly available pretrained commercial large language model was used
- E. The data splitting mimics anticipated real-world applications.
- ☐ Yes _____
- ☒ No Not applicable as a publicly available pretrained commercial large language model was used
- F. The data splitting procedure has been chosen to avoid data leakage.
- ☐ Yes _____
- ☒ No Not applicable as a publicly available pretrained commercial large language model was used

¹ <https://huggingface.co/docs/hub/model-cards>

G. The interpretability of the model has been studied and clearly validated.

- ☐ Yes _____
- ☒ No Not applicable as a publicly available pretrained commercial large language model was used

4. Evaluation

A. The performance metrics used are described and justified in the paper.

- ☐ Yes _____
- ☒ No Not applicable as a publicly available pretrained commercial large language model was used

B. Cross-validation of the results is included.

- ☐ Yes
- ☒ No

C. Community-accepted benchmark datasets/tasks are used for comparisons.

- ☐ Yes _____
- ☒ No No community-accepted datasets/ tasks for GenAI chatbot for smoking cessation are available for benchmarking or comparisons

D. Baseline comparisons to simple/trivial models (for example, 1-nearest neighbour, random forest, most frequent class) are provided.

- ☐ Yes _____
- ☒ No _____

E. Benchmarks with current state-of-the-art are provided.

- ☐ Yes _____
- ☒ No No current state-of-the-art for GenAI chatbot for smoking cessation exists for benchmarking

F. Ablation experiments are included.

- ☐ Yes _____
- ☒ No Not applicable as we did not fine-tune the Large Language Model

G. The model has been tested on a fully independent dataset.

- ☐ Yes
- ☒ No

5. Computational resources

A. The paper contains information on hardware/computing resources that were used.

- ☐ Yes
- ☒ No

B. The paper includes information on the computational costs in terms of computation time, parallelization or carbon footprints estimates.

- ☐ Yes
- ☒ No