NSUCEC NeuraHack 2026

Official Rulebook

Organizer:	North South University Computer & Engineering Club (NSUCEC)
Preliminary round:	24 hours, Kaggle Community Competition
Final:	Hackathon, 4th December 2026
Team size:	2 to 3 members

1. Event Timeline

All times are Bangladesh Standard Time (BST, UTC+6). Kaggle uses UTC, so 6:00 PM BST equals 12:00 PM UTC.

Date	Activity
15th November 2026 (night)	Preliminary round registration starts
23rd November 2026	Preliminary registration closes (1-day extension follows)
24th November 2026, 12:00 PM	Preliminary registration finally closes
24th November 2026, 9:00 PM	Guidance email sent to all teams
25th November 2026, 8:00 PM to 10:00 PM	Mock Prelims
26th November 2026, 6:00 PM	**Prelims start**
27th November 2026, 6:00 PM	**Prelims end**
28th November 2026, 2:00 PM	Prelims results published; final registration starts
1st December 2026, 12:00 PM	Top 35 teams' final registration ends; teams are called from the waiting list
3rd December 2026	Final registration ends
4th December 2026	**Hackathon (Final)**

Reminders are sent 6 hours, 1 hour, and 15 minutes before the Prelims close. The final ZIP package (Section 4) is due on 27th November 2026, 10:00 PM BST.

2. Eligibility and Teams

1. Teams have 2 to 3 members.
2. Each person may belong to only one team and may use only one Kaggle account. Duplicate accounts or teams lead to disqualification.
3. Everyone in the team must be a current undergraduate student of any university in Bangladesh.

3. Scoring and Leaderboards

Submissions are scored with Macro F1 over all classes, as described on the Kaggle competition page.

- **Public leaderboard:** 30% of hidden test rows, visible during the competition.
- **Private leaderboard:** the other 70%, hidden until the competition closes. It decides the final ranking.
- The public score is only a rough guide. The test set differs from the training data (new phrasings, brands, and class mix), so build a robust validation scheme and avoid overfitting to the public leaderboard.
- Other evaluation data may also be used when selecting the Top 35 (Section 7), so aim for a system that generalizes well, not just one that scores high on the leaderboard.

Tie-breaks

1. Higher private score.
2. Earlier time of the submission that achieved the final selected score.

If teams are still tied at the boundary of the Top 60 cutoff after both tie-breaks, the organizers may advance all tied teams to the notebook evaluation in Section 7.

4. Submission Format

4.1 Kaggle submission

Submit a CSV file named `submission.csv` on Kaggle with exactly two columns, in lowercase:

```
id,label
20001,0
20002,5
20003,2
```

1. Include the header row.

2. Include every `id` from `test.csv` exactly once.
3. Do not change the `id` values or add extra columns, including the message text.
4. `label` must be an integer from 0 to 5. Strings, decimals such as `2.0`, and blanks are rejected.

Submissions that fail these checks are rejected and may still count against your daily limit.

4.2 Final package (ZIP)

After the competition closes, each team must submit one ZIP file by the deadline in Section 1. The ZIP file must be named:

```
NeuraHack2026_Team_Name_Team_Code.zip
```

and must contain exactly the following four items:

Item	Description
`.csv`	The final prediction file (`submission.csv`) matching your selected Kaggle submission
`.ipynb`	The reproducible notebook that generates the predictions, including the declaration section required by Section 6
`.env`	Environment file containing every API key your notebook uses (see Section 5)
`README.md`	A short guide to reproducing your result (see Section 6)

Example: `NeuraHack2026_MissionX_NH344.zip`, where `MissionX` is the team name and `NH344` is the team code.

Further guidance on the package and the upload process will be sent by email to all teams.

5. Permitted and Prohibited Conduct

Permitted

- **All AI tools are allowed.** This includes open-source pretrained models (for example XLM-R, mBERT, BanglaBERT, and multilingual sentence embeddings, subject to their licenses), hosted LLM and inference APIs (for example OpenAI, Gemini, or Claude, paid or free), and AI coding assistants.
- Hosted LLM and inference APIs may be used for training, data labeling, data generation, augmentation, feature extraction, and inference, provided the calls are made from code inside your submitted notebook and declared as described in Section 6.
- Public external datasets, if declared (see Section 6).
- Feature engineering, fine-tuning, ensembling, data augmentation, and use of public code, if disclosed.
- Unsupervised or semi-supervised use of the unlabeled `test.csv` text (for example pseudo-labeling or domain adaptation) is allowed, provided it is done by code in your notebook, uses only model predictions, and is declared.

- Kaggle Notebooks and the resources available under Kaggle's current notebook settings.

API keys and the `.env` file

- Every API key used by your notebook must be placed in a `.env` file inside your final ZIP package (Section 4) so that organizers can re-run your notebook.
- The notebook must read keys from the environment (for example with `python-dotenv` or `os.getenv`). Do not hardcode keys inside the `.ipynb`.
- Use a dedicated key for this competition with a spending limit, and make sure it stays valid and funded until the verification period ends. A key that is invalid, expired, or out of credit may be treated as a failed reproduction.
- Organizers will use the keys only to reproduce your submission and will not use them for any other purpose. We recommend revoking the key once verification is complete.

Prohibited

1. **Test-label recovery.** Any attempt to recover, look up, or infer the test labels by any means other than your model's predictions leads to disqualification.
2. **Manual labeling or editing.** Hand labeling, manual annotation, or manual editing of test predictions is prohibited. This includes pasting answers from a chat interface into a file; every test prediction must be produced by code in your notebook.
3. **Hardcoded predictions.** Submitting predictions that are hardcoded, copied, or stored in a manually created label file is prohibited.
4. **Private sharing.** Private sharing of code, data, predictions, or API keys between teams is prohibited. Sharing publicly through the official channel is allowed.
5. **Multiple identities.** Using more than one account, or submitting for another team, is prohibited.
6. **Undeclared resources.** Using any external dataset, model, or API that is not declared in the notebook's declaration section (Section 6) is prohibited.

Teams are responsible for the cost of their own API usage.

6. Reproducibility and Verification

All teams must submit the final ZIP package (Section 4) by the deadline in Section 1. No files other than the four items listed in Section 4 may be added to the ZIP. The ZIP must include a `README.md` with short instructions for reproducing your result, such as the run order, approximate runtime, and any notes on API usage. In addition, the notebook must begin with a declaration section (markdown cells) containing:

1. The libraries and versions needed to run it (or an install cell that sets them up).
2. A list of all external datasets, pretrained models, and APIs used, with their sources and licenses or providers.
3. A short description of the method and validation strategy.

4. The random seeds used, and any instructions needed to reproduce the result.

Organizer verification. Teams shortlisted under Section 7 are verified as follows:

- The organizers may run the notebook's pipeline on additional evaluation data (see Section 7). Your notebook should be written so that it can be applied to new, unseen text without manual changes, and the 0.01 tolerance above applies only to the original Kaggle test set.
- The notebook is re-run in a clean environment (Kaggle or Colab) with internet access, using the keys in the team's `.env` file and the same seeds where possible.
- A reproduced private score within **0.01** of the submitted score is accepted. A larger difference triggers a review, taking non-deterministic LLM output into account.
- Notebooks are inspected for undeclared resources, undeclared API calls, hardcoded predictions, manually created label files, and hardcoded keys.
- Notebooks may be compared for suspiciously similar code or identical, unexplained predictions.

7. Advancement

1. The **Top 60** teams by private Kaggle score are shortlisted.
2. From these, the **Top 35** are selected by manual notebook evaluation. Notebooks are judged on reproducibility, sound validation without leakage, compliance with Section 5, analysis of generalization (languages, unseen brands, noise, class confusions), and clarity of documentation.
3. **Additional evaluation data.** For the Top 60 shortlist, the organizers may also evaluate notebooks on extra data that is not part of the Kaggle test set (for example, a fresh held-out set with new phrasings, brands, operators, or a different class mix). Performance on this data may be considered alongside the private leaderboard score and the notebook review when selecting the Top 35. Teams should therefore build the most robust and generalizable system they can, rather than one tuned to the public or private leaderboard.
4. After verification, the Top 35 teams advance to the NeuraHack 2026 Final. Prelims results are published on 28th November 2026 at 2:00 PM, when final registration opens.
5. Selected teams must complete final registration by 1st December 2026, 12:00 PM. Slots not claimed by then are offered to teams on the waiting list. Final registration closes on 3rd December 2026.

8. Submission Limits and Final Selection

- **Daily submission limit:** 10 submissions per team per day.
- **Final submissions:** each team may select 2 final submissions before the deadline.
- **If you do not select:** Kaggle automatically uses the two submissions with the best public score. Because the public leaderboard covers only 30% of the test set, choose your final submissions using your own validation results.
- The private score of your best selected submission determines your rank.
- The `.csv` in your final ZIP package must match one of your selected Kaggle submissions.

9. Disqualification and Appeals

The organizers may disqualify a team for violating Sections 2, 5, or 6, including any of the following:

- attempting to recover or look up test labels
- hardcoding or manually editing predictions
- using multiple accounts, or sharing private code, data, or API keys between teams
- using undeclared datasets, models, or APIs
- failing to submit the final ZIP package on time, omitting the `.env` file or the `README.md` file, or providing a notebook or keys that cannot reproduce the score within the tolerance stated in Section 6

Disqualified teams are removed from the standings. Appeals must be submitted through the official channel within 18 hours of the notice. Organizers' decisions are final.

10. Clarifications and Communication

- All clarifications are posted in the official channel and apply to every team.
- Organizers do not answer rule questions privately.
- A launch announcement with the competition link, rules, and deadline in UTC+6 is pinned in the official channel.
- Further guidance on the final package will be sent by email to all teams.

11. Prizes and Taxation

- As per North South University (NSU) policy, a **10% tax will be deducted** from all prize money before it is disbursed to the winning teams.
- The amount received by a team is the announced prize amount minus this deduction.
- Any prize money is distributed at the team level, and the team members are responsible for dividing it among themselves.

12. FAQ

Can we use pretrained models? Yes, open-source and downloadable models such as XLM-R, mBERT, BanglaBERT, or multilingual sentence embeddings are allowed.

Can we use ChatGPT, Gemini, Claude, or other hosted LLMs and APIs? Yes. All AI tools are allowed, including for labeling, data generation, and prediction, as long as the calls come from code in your notebook, are declared, and the API key is included in the `.env` file of your final ZIP.

Do we have to include our API key? Yes, if your notebook uses any API. Put the key in the `.env` file in your ZIP. Do not hardcode it in the notebook.

What should the ZIP be named and what does it contain?
`NeuraHack2026_Team_Name_Team_Code.zip`, containing the final `.csv`, the `.ipynb`, the `.env`, and the `README.md`. Example: `NeuraHack2026_MissionX_NH344.zip`.

Can we use extra data? Yes. Public datasets are allowed and must be declared in the declaration section of your notebook (Section 6).

Can we use the test text for semi-supervised learning? Yes, if done by code in your notebook and declared (see Section 5). Looking up or hand-editing test labels is never allowed.

Can we use Kaggle Notebooks? Yes.

Why does my public score differ from my validation score? The test set contains shifted phrasings, new brands and operators, and a different class distribution. Use robust validation, for example grouping similar messages together, and avoid overfitting to the public leaderboard.

Which submission determines the final result? The private leaderboard score of your selected final submissions (see Section 8).

Is tax deducted from the prize money? Yes. As per NSU policy, 10% tax is deducted from the prize money (see Section 11).

Revision notice (Version 1.1, 16th November 2026): Section 11 (Prizes and Taxation) was added. As per NSU policy, 10% tax is deducted from prize money. No other rules were changed.

Revision notice (Version 1.2, 19th November 2026): Section 7 (Advancement) now states that additional evaluation data may be used when selecting the Top 35, with matching notes in Sections 6 and 3. Section 11 (Prizes and Taxation) was added in Version 1.1. No other rules were changed.

Revision notice (Version 1.3, 22nd November 2026): A `README.md` file is now required in the final ZIP package, which now contains four items. Sections 4 and 6, Section 9, and the FAQ were updated accordingly. No other rules were changed.