VlogQA Dataset User Agreement

1. User Information

- Full name:

- Organization:

- Email:

- Telephone (optional):

  Note: If you are a student, this user agreement should be signed by your supervisor.

2. Dataset Information

- Dataset Name: **VlogQA**
- Size: 1,230 transcript documents and 10,076 human-annotated question-answer pairs.

3. Terms and Conditions

The user of VlogQA developed by the NLP@UIT research group must respect the following terms and conditions:

1. The dataset is only used for **non-profit research** for natural language processing and education.

2. The dataset is **not allowed to be used in commercial systems**.

3. Do not redistribute the dataset. This dataset may be modified or improved to serve a research purpose better, but the edited dataset may not be distributed.

4. Summaries, analyses, and interpretations of the properties of the dataset may be derived and published, provided it is not possible to reconstruct the information from these summaries.

5. Published research works that use the dataset must cite the following paper:

Thinh Ngo, Khoa Dang, Son Luu, Kiet Nguyen, and Ngan Nguyen. 2024. **VlogQA: Task, Dataset, and Baseline Models for Vietnamese Spoken-Based Machine Reading Comprehension**. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1310–1324, St. Julian's, Malta. Association for Computational Linguistics*.

By the User

Signature:

Date:

Name (please print):

*Forward the signed agreement by PDF to this email sonlt@uit.edu.vn