

Evaluation Event II/III/Final - *Agent Account Name*

1st Full Name (*User Account Name*)

2nd Full Name (*User Account Name*)

October 27, 2025

1 Introduction

In this section, please briefly summarize your agent’s current capabilities and key features.

Recommended length: 50 to 200 words.

2 Capabilities

In this section, please first provide an overview diagram of your agent, depicting the agent’s components and the inputs/outputs of the components.

Then, please elaborate on your agent’s capabilities regarding answering the specified five types of questions, including factual questions, embedding questions, multimedia questions, recommendation questions, and crowdsourcing questions. For intermediate evaluation events, you can disregard question types that are not considered.

Example:

- Factual Questions

Our agent answers factual questions in three steps: 1) ..., 2) ..., and 3) ... In the first step, we employ BERT [1] to perform ... We made the following changes to adapt BERT to the task ...

Recommended length: 100 - 300 words.

3 Adopted Methods

In this section, please report any adopted methods, techniques, tools, and packages with proper URL references or citations, e.g., NLTK¹ and BERT [1]. You

¹<https://www.nltk.org/>

also need to provide the rationale for using these methods.

Example:

- BERT [1]
We used BERT for ... The reason is that ...

Recommended length: 100 - 300 words.

4 Examples

In this section, regarding each question type reported above, please provide the answering process of one example question.

Example:

- Factual Questions
Given the question “*who is the director of Good Will Hunting*,” our agent first ... then ... In the end, we can obtain the answer ..., which is correct as ... or false due to ...

Recommended length: 100 - 200 words.

5 Additional Features

In this section, please report additional features of your agent, e.g., mechanisms that improve the timeliness or “humanness” of your agent.

Also, you can add code snippets or screenshots to showcase any strengths of your agent or any challenges you managed to address.

Recommended length: within one page.

6 Conclusions

For intermediate submissions, please report your plan for the next steps.

For the final report, please provide a summary of your work in the whole course assignment and report your individual contributions.

References

- [1] J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, and T. Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN*,

USA, June 2-7, 2019, Volume 1 (Long and Short Papers), pages 4171–4186.
Association for Computational Linguistics, 2019.