

# Guidelines for QA Creation

You will receive a **spreadsheet** where the QA can be written and annotated. The spreadsheet also contains all the necessary information and materials to complete the task, including a link to the *video*, a *transcript* of the talk, and an *abstract of the talk*.

## What we need

- **(3) General questions** that apply to all papers with answers: the questions are already given, and the answers are created after reading the transcript and watching the video
- **(3) Abstract paper-specific questions** with answers, where the questions are created after reading the abstract, and the answers are written after reading the transcript and watching the video
- **(4) Transcript paper-specific questions**, with both questions and answers that are created after reading the transcript and watching the video

🕒 Each paper will have in total 10 Q-A pairs: 3 General, 3 Abstract paper-specific, 4 Transcript paper-specific.

The *general questions* that apply to each paper are the following:

1. What is the name of the speaker?
2. How many authors are involved in the paper?
3. What are the affiliations of the authors of the paper?

## Which process to create Q and A

- Each person will be assigned X papers, and for each of them, they will need to
  - i) find the A for the general Q (see above),
  - ii) come up with paper-specific Q and corresponding A
- In all cases (i.e. both specific and general Q), while trying to find an A the person needs to keep track of whether:
  - **NA**: the Q is not answerable, neither based on the audio nor video (the information is not available at all)
  - **AV**: the Q is answerable based on either video or audio (the information is present in both video and audio, and either one of them could be sufficient to answer the Q)
  - **A**: the Q is answerable based on audio only (no information present in the video)

- **V:** the Q is answerable based on video only (no information present in the audio)







**!!!** To create the Q-A pairs, you need to both **read the transcript and watch the video**.

- Watching the **video** is necessary to:
  - Assign the correct above-mentioned labels
  - Annotate timestamps. For each Q-A, you need to annotate the start and end time of the video in which the answer is to be found (see instructions below in “Where to create”).
- Reading the **transcript** serves as a verification step of its quality: If you find any errors, please rework the transcript to fix them (see instructions below in “Where to create”). However, the procedure on when to read the transcript and create the questions differs between *abstract paper-specific* vs transcript *paper-specific* Q-A pairs:
  - *abstract paper-specific*: we want to simulate a scenario where a user reads an abstract and then has questions about the paper. For this reason, you need first to read the abstract, and only the abstract to generate the Q. You will read the transcript and watch the video afterwards, and only to find the A.
  - *Transcript paper-specific*: you come up with Q-A pairs while and/or after reading the transcript and watching the video.
- 🎯 **The goal is to obtain 10 Q-A pairs for each paper, following this distribution of Q-A types:**
  - 7 (A/AV type, at least 5 AV) + 2 (V) + 1 (NA).
  - Accomplishing this distribution is the responsibility of each person for their own assigned papers.
  - Ideally, you will accomplish this distribution with the three General Q-A + three Abstract Q-A + by creating four Transcript paper-specific Q-A
  - *Should* the general Q and Abstract Q imply an unbalanced distribution (e.g. two or more NA types), you need to come up with more Transcript paper-specific Q-A to reach the distribution target. You might have more than 10 Q-A pairs at the end of the process.







## How to create

### Type of Q:

- Come up with Q that allow for fairly concise answers (from few words to 2 sentences max). The goal is not to conflate QA with the summarization task. For instance:
  - ❌ Q: What are the conclusions of the work?
  - ✅ Q: Which models are used in the presented experiments?

- Avoid Q that are ambiguous, too broad, vague, or with undefined scope that could be hard to evaluate. For instance, try to ground Q that are more open-ended:
  -  What is the cause?
  -  What is the cause of the results improvement concerning translation quality?
  -  Q: What is the significance of this work?<sup>1</sup>
  -  Q: What is the significance of the [presented empirical results] on [LMs translation quality] for the [modeling of the MT task]?
- Use Diverse Sentence Structures: Avoid repeating patterns in questions, e.g. using only wh-questions (how, what, where..?).
- Avoid “Multi-Questions”, i.e. inserting two questioning steps. For instance:
  -  Is the presentation about a new model? If yes, what is its name?
  -  If the presentation is about a new model, provide its name, else answer that there is no new model.

## Type of A:

- Avoid providing a single-word response. For instance, for the question, *How many institutions participated in the study?* See some possible answers:
  -  This study is the result of the collaboration across five institutions
  -  Five institutions worked on this study
  -  Five institutions worked on this study
  -  There are five institutions
  -  Five institutions
  -  Five
- Ensure the A directly addresses the Q. Avoid irrelevant details.
- For NA type of questions, the only acceptable kind of A is:
  - Not answerable

## General formatting guidelines:

- For English QA, use US spelling
- Avoid using number format (i.e. write *seven researchers*, not *7 researchers*) except for:
  - Dates
  - Numerical results
  - Numbers above 12, e.g. twelve, but 13, not thirteen.

---

<sup>1</sup> In regard to what? Social, technical, economic or scientific?