UE MLPC 2025:
DATA EXPLORATION TASK

Abdalaziz Adil Awad Ayoub
Abdulkarim Al Jamal
Beibarys Abissatov
Jeronim Bašić

JOHANNES KEPLER
UNIVERSITY LINZ

Annotation Quality

- **How precise are the temporal annotations?**

We analysed annotation data to identify overlapping annotation durations among files. We first calculated the annotation duration by subtracting the onset from the offset, then we detected the instances of unique overlapping between annotations using a precision threshold of 100 ms, to achieve the following results: “Total number of files with 2 annotators: 725; Maximum number of overlapping regions: 20 (for audio file: 560530.mp3); Number of files with at least 1 overlapping region with 2 annotators: 446 Total number of files with 3 annotators: 6 Maximum number of overlapping regions: 6 (for audio file: 582364.mp3); Number of files with at least 1 overlapping region with 3 annotators: 4.”, meaning there is an agreement for 61.5% of the cases with two annotations within the given threshold. This is not very precise but perhaps explained by files, such as 560530.mp3, containing complex or multiple sound events, which can be interpreted differently by the several annotators.

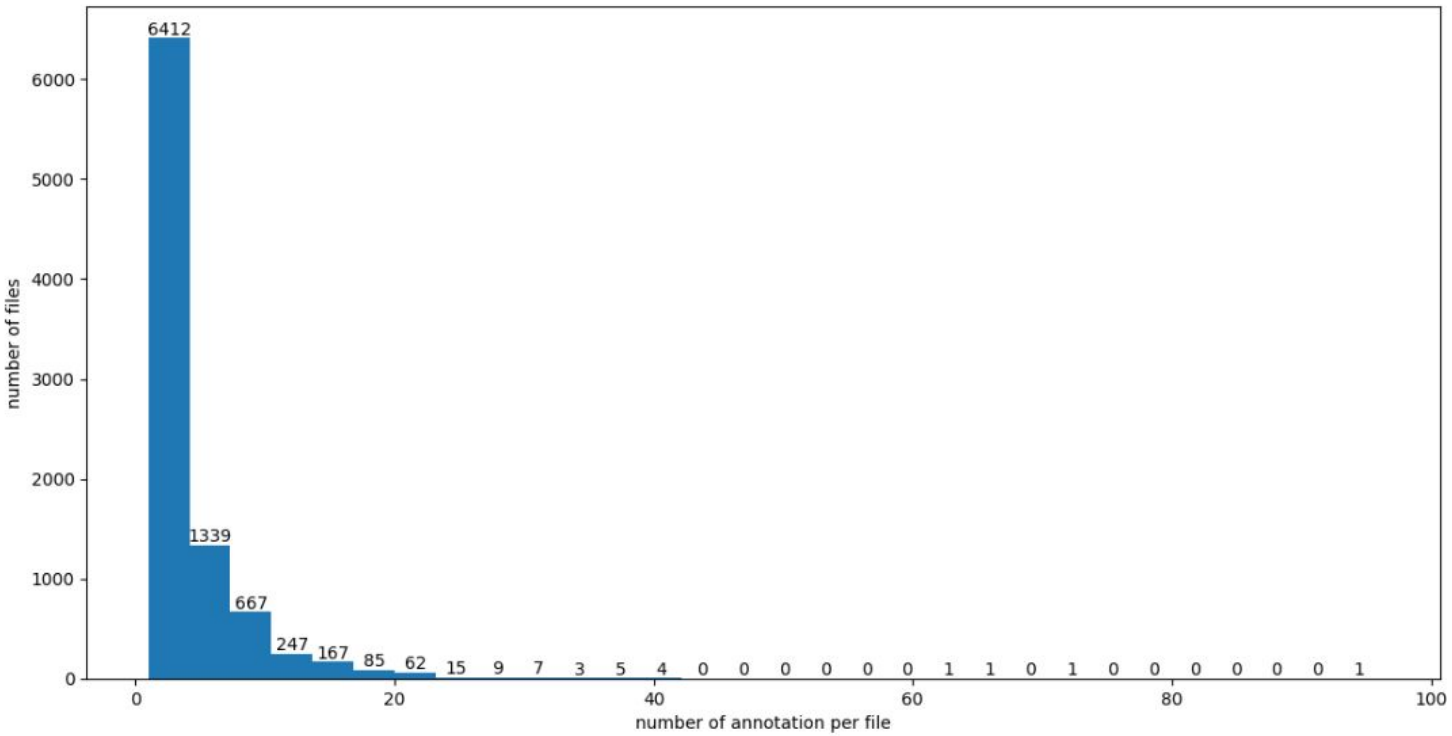
Annotation Quality

- **How similar are the text annotations that correspond to the same region?**

  Here we calculated the cosine similarity of text embeddings for sound event annotations overlapping within the mentioned threshold to get the following results: “mean similarity with 2 annotators: 0.4271726608276367; mean similarity with 3 annotators: 0.3451352119445801”, which makes for some level of agreement, although lesser among triple-annotator files.

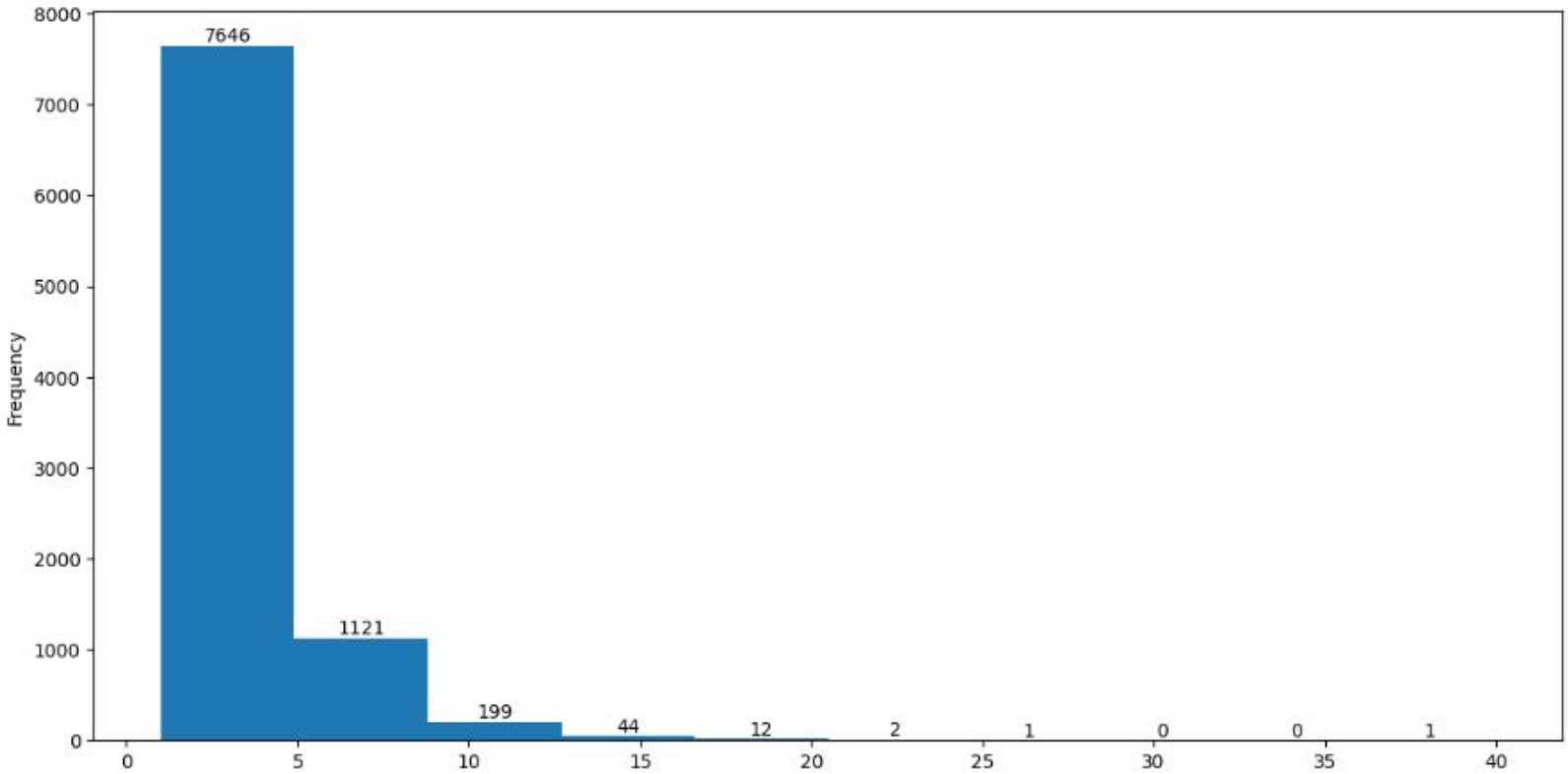
Annotation Quality

- **How many annotations did we collect per file?**



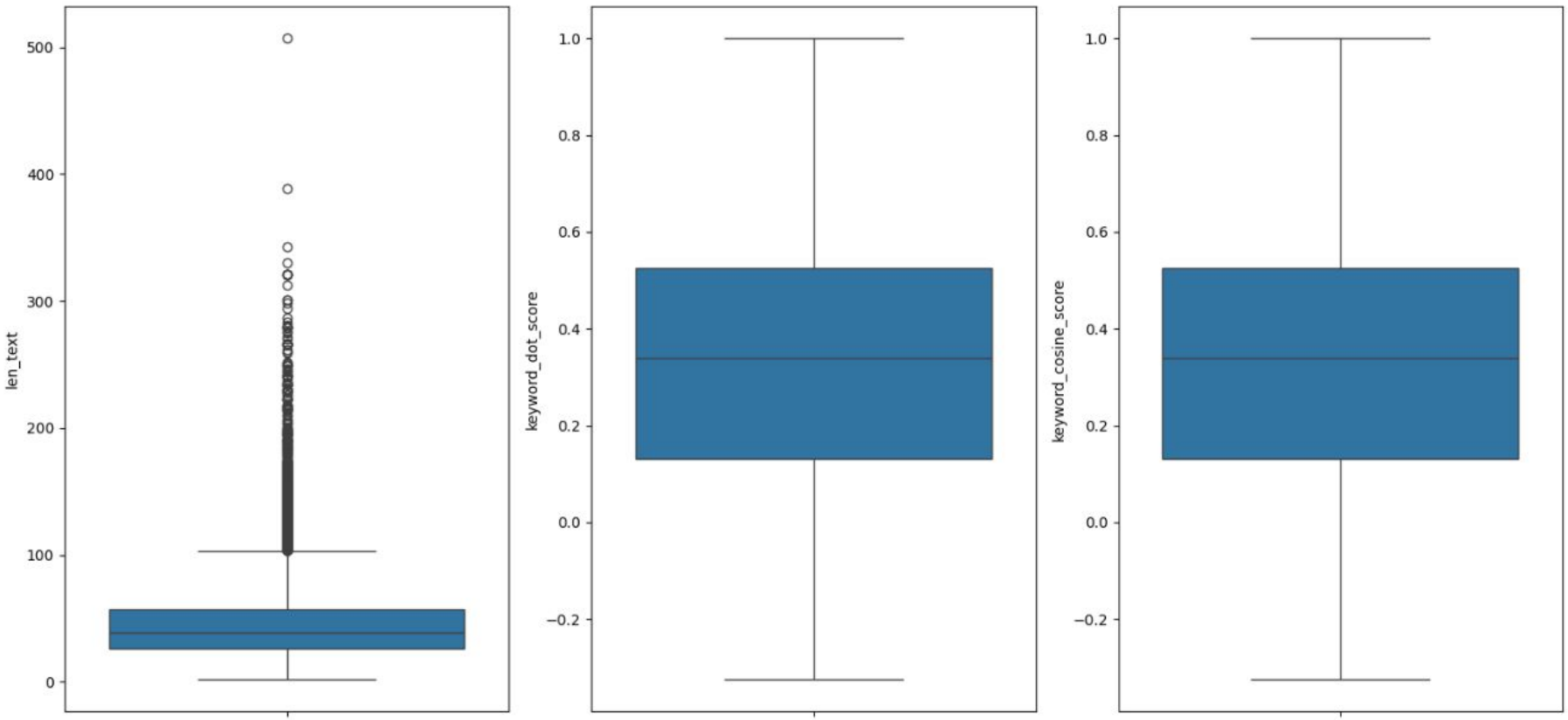
Annotation Quality

- **How many distinct sound events per file?**



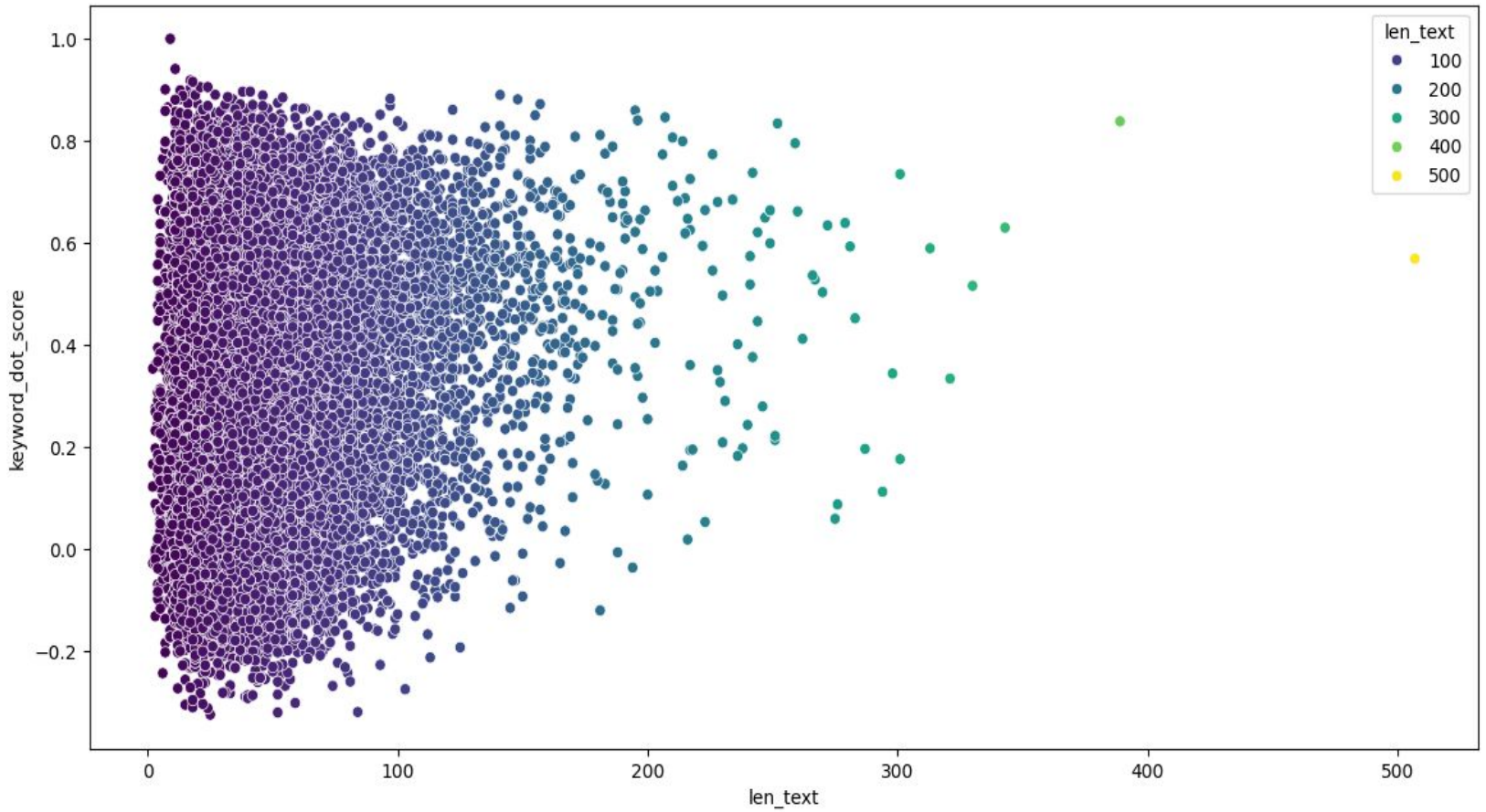
Annotation Quality

- **How detailed are the text annotations?**



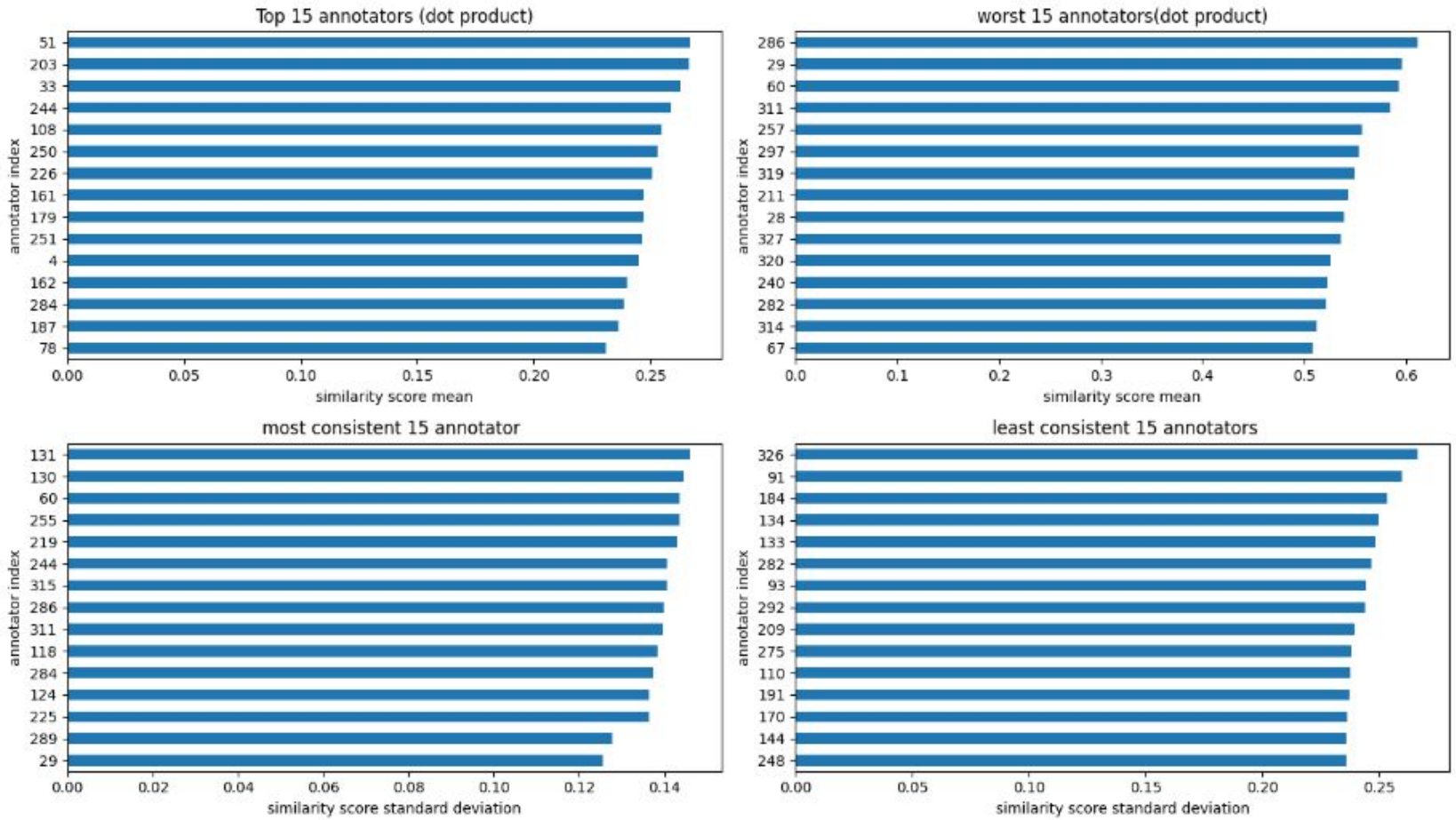
Annotation Quality

- **How detailed are the text annotations?**



Annotation Quality

- **How detailed are the text annotations?**



Annotation Quality

- **Are there any obvious inconsistencies, outliers, or poor-quality annotations in the data? Propose a simple method to filter or fix incorrect or poor-quality annotations (e.g., remove outliers, typos, or spelling errors).**

We notice several outliers in the box-plots, with some annotations containing over 300 characters. However, according to the scatter, these did not necessarily register a higher similarity score. One annotation, consisting of a single word, achieved a perfect similarity score of 1. Upon further inspection, this annotation just so happened to exactly contain the only word listed in the metadata, hence being a perfect match. Most of the lower quality annotations also contained fewer characters. Therefore, a simple method at filtering such instances would be to exclude the quadrant of negative similarity scores containing <100 characters.

Thank you!
Any Questions?