

Evaluating Webcam-based Gaze Data as an Alternative for Human Rationale Annotations

**Stephanie Brandl¹, Oliver Eberle², Tiago Ribeiro³,
Anders Søgaard¹, Nora Hollenstein^{1,4}**

¹ University of Copenhagen, Denmark

² Technische Universität Berlin, Germany

³ IT University of Copenhagen, Denmark

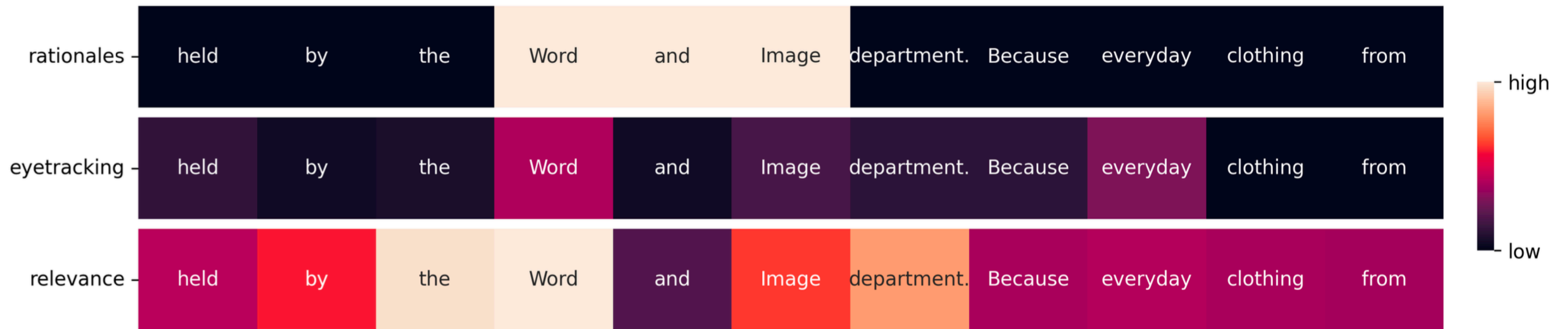
⁴ University of Zurich, Switzerland

Evaluating Webcam-based Gaze Data as an Alternative for Human Rationale Annotations

- ▶ XAI is often evaluated with human annotations as the ground truth
- ▶ How can we use human gaze, in particular webcam-recorded, to evaluate XAI?
 - ▶ Pros:
 - more intuitive
 - additional information: continuous scores, regression, information about task difficulty and accuracy
 - ▶ Cons:
 - webcam-based data does not reach lab quality (yet!) and lab data is too expensive

Question Answering

Which department houses the works on paper of the costume collection?



Experiments

- ▶ total reading times, decoding accuracy and entropy
- ▶ 4 mono- and multilingual models
- ▶ 3 languages: en, es, de
- ▶ our pipeline can easily be applied to other datasets/tasks/languages

Conclusion

- ▶ XAI methods and eye-tracking show similar token rankings
- ▶ this is influenced by
 - ▶ relative position of answer in text
 - ▶ text length
 - ▶ number of tokens in correct answer
- ▶ longer texts and shorter answers lead to higher agreement
- ▶ webcam-based eyetracking can not yet fully replace high-quality rationale annotations, but they do provide useful information, in particular, when collecting rationales is not feasible

 Thank you for your attention 

Reach out:

Stephanie Brandl

✉ brandl@di.ku.dk

🐦 @StephanieBrandl