Help patients to **understand** their clinical notes with automatic **text simplification.**

Patient-Friendly Clinical Notes: Towards a new Text Simplification Dataset

UNIVERSITÄT DUISBURG ESSEN

*Open-*Minded

Universitätsmedizin Essen
Universitätsklinikum

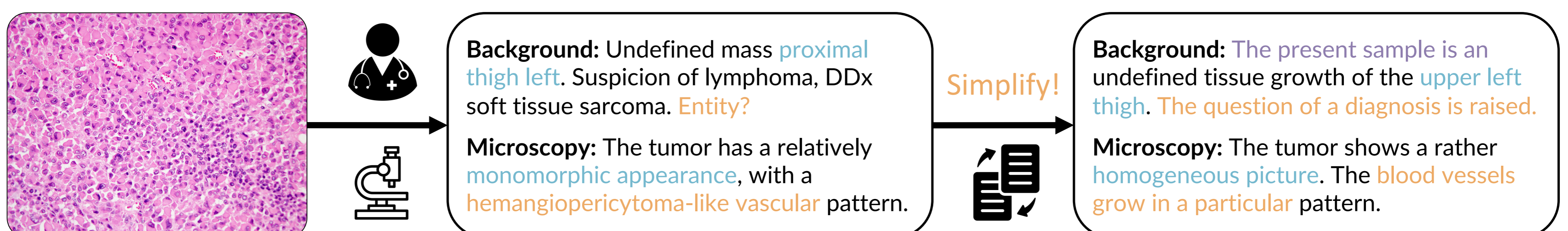👤 **Jan** Trienes, Jörg Schlötterer, Hans-Ulrich Schildhaus, Christin Seifert

✉ jan.trienes@uni-due.de

WHY?

- Hospitals around the world **share clinical notes** with patients
- However, **notes are hard to read** for patients (jargon, detailed analyses, unfamiliar syntax)
- Automatic text simplification can help, but **we lack good datasets** to develop those methods

METHOD

- We **create a parallel dataset of professionally simplified** German pathology **reports**
- Hired 9 medical students: *"how would you intuitively explain this to a patient?"*
- Analyze **surface-level characteristics** of the text and train **paragraph-level baselines**

Background: Undefined mass proximal thigh left. Suspicion of lymphoma, DDx soft tissue sarcoma. Entity?

Microscopy: The tumor has a relatively monomorphic appearance, with a hemangiopericytoma-like vascular pattern.

Simplify!

Background: The present sample is an undefined tissue growth of the upper left thigh. The question of a diagnosis is raised.

Microscopy: The tumor shows a rather homogeneous picture. The blood vessels grow in a particular pattern.

RESULTS

- **851 notes, 790k words**, >800 hours of effort
- Simplifications are on average **41% longer**, with less **lexical diversity** (TTR ↓ 10%), and a **higher readability** (FRE ↑ 22%)
- Background section is hardest to simplify

CONCLUSION

- **The new dataset will allow us to study how to simplify clinical notes**
- Several operations needed: **summarization, explanation generation, lexical simplification**
- Dataset is of **sufficient size to fine-tune** large language models
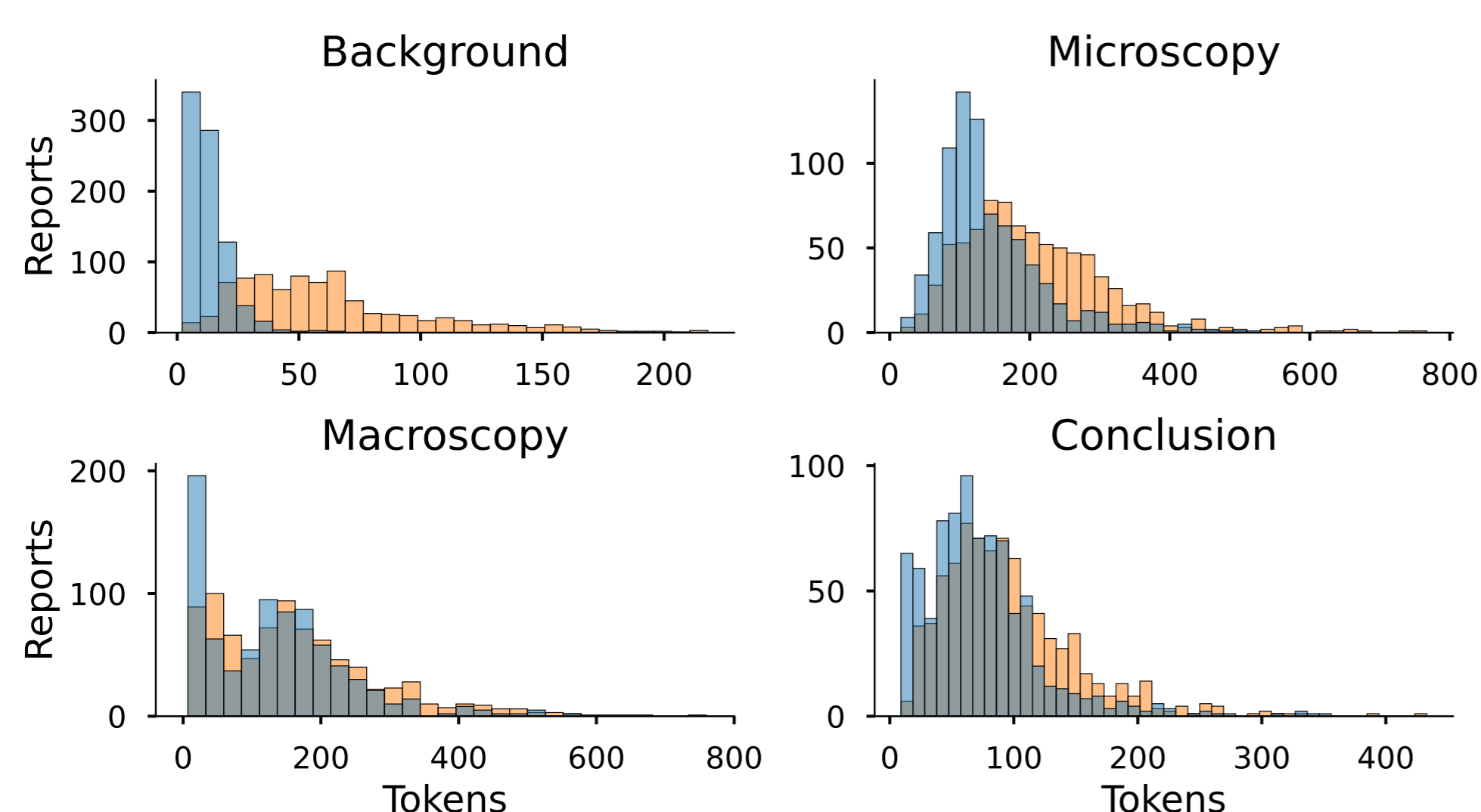
Model	R-1	R-2	R-L	BLEU	SARI	Len.
Identity	29.6	14.3	28.6	10.8	11.2	92
Bert2Bert	26.5	8.3	25.0	7.3	41.4	103
Bert2Share	28.3	9.5	26.6	8.2	42.7	102
mBART	**35.2**	**15.3**	**33.4**	**14.2**	**46.2**	129
mBART by report section						
Background	25.1	7.7	23.0	6.5	47.9	86
Macroscopy	**45.5**	**24.8**	**43.7**	**18.0**	**48.2**	131
Microscopy	36.3	13.2	34.9	13.0	44.3	213
Conclusion	33.8	15.2	32.0	13.6	43.6	88