As words have power, names have power

Maria Irena Szawerna, **Ricardo Muñoz Sánchez**, Simon Dobnik, Therese Lindström Tiedemann, Lisa Södergård, Elena Volodina, Xuan-Son Vu

CLEANUP seminar
Oslo, 08.10.2024

The Team

What is Pseudonymization?

- The process of replacing an individual's personal data with a pseudonym, which is not related to the original data (Volodina et al. 2023)

Jag heter Maria. Jag kommer från Wrocław (det är i Polen), men jag bor i Heidelberg. Jag studerar lingvistik [...]

My name is Maria. I come from Wrocław (that's in Poland), but I live in Heidelberg. I study linguistics [...]

23.10.2019

Jag heter **Maria**. Jag kommer från **Wrocław** (det är i **Polen**), men jag bor i **Heidelberg**. Jag studerar **lingvistik** [...]

*My name is **Maria**. I come from **Wrocław** (that's in **Poland**), but I live in **Heidelberg**. I study **linguistics** [...]*

23.10.2019

Jag heter **Anna**. Jag kommer från **Ostrava** (det är i **Tjeckien**), men jag bor i **Cambridge**. Jag studerar **kemi** [...]

*My name is **Anna**. I come from **Ostrava** (that's in **Czechia**), but I live in **Cambridge**. I study **chemistry** [...]*

23.10.2019

Why pseudonymize?

- Research requires data
  - Language data: linguistics, language technology, digital humanities, etc.
  - Language data can contain personal information
- Data sharing: ethical and legal issues (GDPR)
- Mitigating privacy risks can lead to more open access to data
- "Tampering" with data can have negative effects

Who is Grandma Karl?

Grandma Karl is 27 years old: automatic pseudonymization of research data — VR project

– robust AI methods and tools for pseudonymizing Swedish texts

(University of Gothenburg: *Elena Volodina, Simon Dobnik, Maria Irena Szawerna* **+** *Ricardo Muñoz Sánchez*)

– effectiveness in terms of privacy preservation

(Umeå University: *Xuan-Son Vu*)

– impact on the data and its usability in future research

(University of Helsinki: *Therese Lindström Tiedemann, Lisa Södergård*)

Pseudonymization and AI

- Detection and classification of personal information

- Generation of viable substitutes

- What we've done so far:
  - Can ML models generalize what personal information is? (Szawerna et al. 2024a)
  - How do people categorize personal information and how does it differ across domains? (Szawerna et al. 2024b)

Input

Model

Which parts of the text are sensitive?

Output
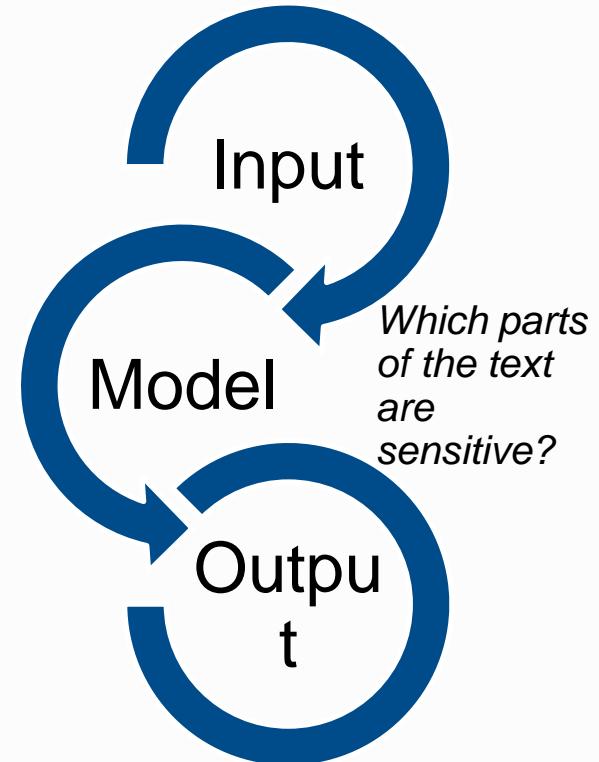
Pseudonymization and Biases

- Deep learning can replicate human-like biases
- This can have an impact on downstream applications
- It can also lead to models generating pseudonyms based on stereotypes

- What we've done so far:
  – We have worked on identifying whether pseudonyms based on perceived ethnicity in a text affect how it is graded (Muñoz Sánchez et al. 2024a, 2024b)

Pseudonymization and Linguistic Research

- The impact on linguistic research data and its usability in future research

- What we've done so far:
  - How is personal information treated in linguistic research?
  - Comparison of research articles on the Swedish language before and after the GDPR (Södergård, In prep.)

Future work

- How general or specific should classes for detection be?
- How reliably can we use LLMs to generate pseudonyms based on context?
- What is the impact of pseudonymization on different types of (Swedish) texts?
- How can pseudonymization affect language assessment?

- A tool for customizable pseudonymization
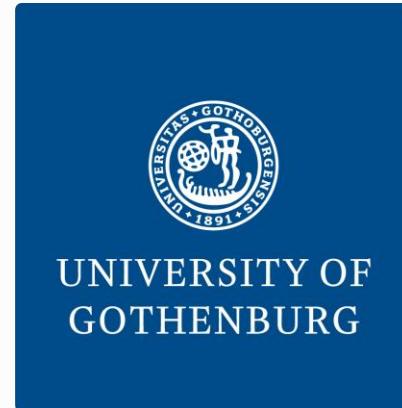
Thank you for your attention!

References

- Ricardo Muñoz Sánchez, Simon Dobnik, Maria Irena Szawerna, Therese Lindström Tiedemann, and Elena Volodina. 2024a. Did the names I used within my essay affect my score? diagnosing name biases in automated essay scoring. In Proceedings of the Workshop on Computational Approaches to Language Data Pseudonymization (CALD-pseudo 2024), pages 81–91, St. Julian's, Malta. Association for Computational Linguistics.
- Ricardo Muñoz Sánchez, Simon Dobnik, Therese Lindström Tiedemann, Maria Irena Szawerna, and Elena Volodina. 2024b. Name Biases in Automated Essay Assessment. In The 28th International Congress of Onomastic Sciences.
- Maria Irena Szawerna, Simon Dobnik, Ricardo Muñoz Sánchez, Therese Lindström Tiedemann, and Elena Volodina. 2024a. Detecting Personal Identifiable Information in Swedish Learner Essays. In *Proceedings of the Workshop on Computational Approaches to Language Data Pseudonymization (CALD-pseudo 2024)*, pages 54–63, St. Julian's, Malta. Association for Computational Linguistics.
- Maria Irena Szawerna, Simon Dobnik, Therese Lindström Tiedemann, Ricardo Muñoz Sánchez, Xuan-Son Vu, and Elena Volodina. 2024b. Pseudonymization Categories across Domain Boundaries. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13303–13314, Torino, Italia. ELRA and ICCL.
- Lisa Södergård. In prep. Behandlingen av personuppgifter i språkvetenskapliga forskningspublikationer [=The treatment of personal information in linguistic research publications]. In Svenskan i Finland 21, 3-4-October 2024, Oulu, Finland.
- Elena Volodina, Simon Dobnik, Therese Lindström Tiedemann, Xuan-Son Vu. 2023. Grandma Karl is 27 years old – research agenda for pseudonymization of research data. In *2023 IEEE Ninth International Conference on Big Data Computing Service and Applications (BigDataService),* Athens, Greece, 2023

Find out more about our research here!