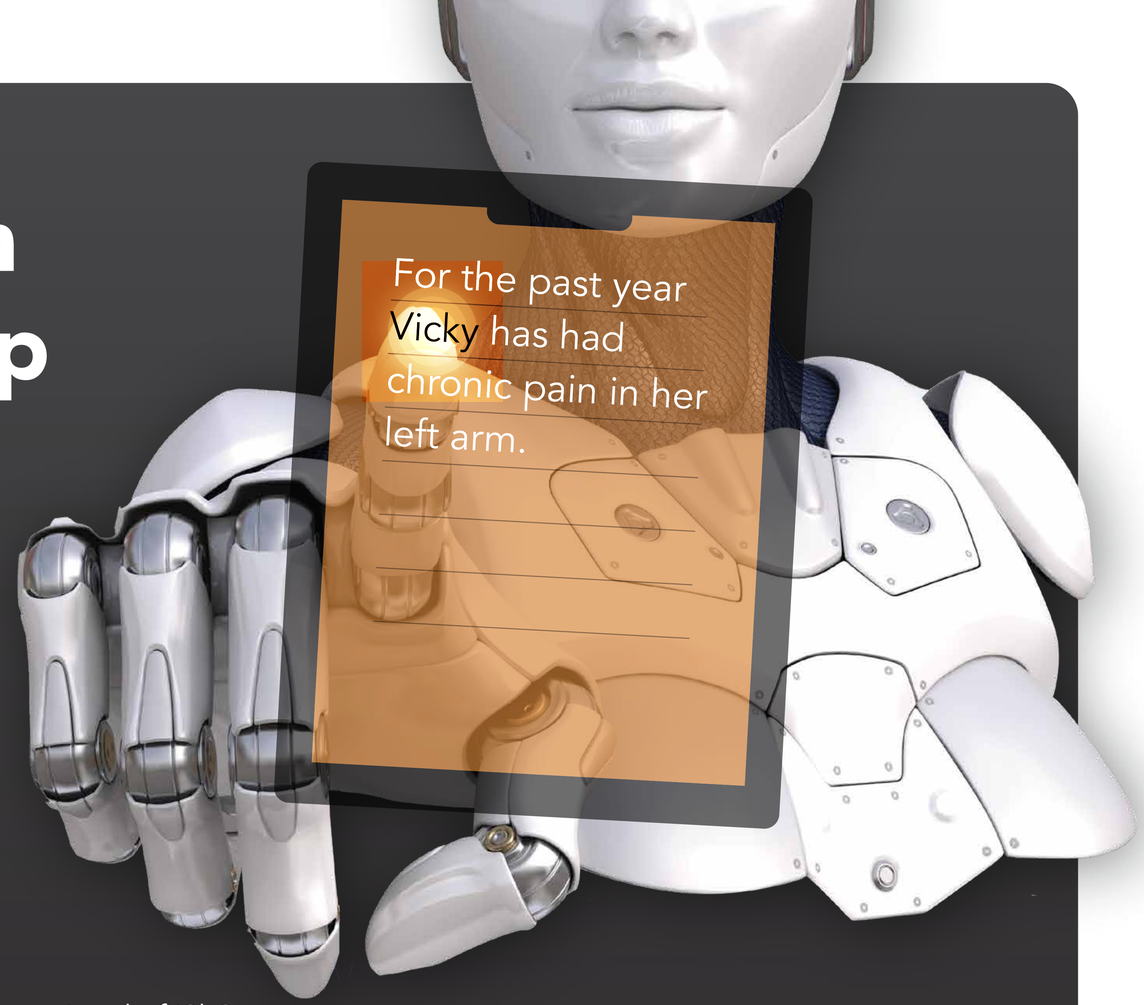
Automated de-identification of case narratives using deep neural networks for the UK Yellow Card scheme

Eva-Lisa Meldau, Shachi Bista, Carlos Melgarejo-González, G Niklas Norén
Uppsala Monitoring Centre



Background

Access to case narratives during signal assessment is crucial to provide a more complete picture of cases[1], however patient confidentiality needs to be considered. Sharing of narratives while preserving privacy requires de-identification. Person names – one of the more common identifiers in case narratives – can lead to (in-)direct identification of patients but are challenging to recognise in free text.

Objective

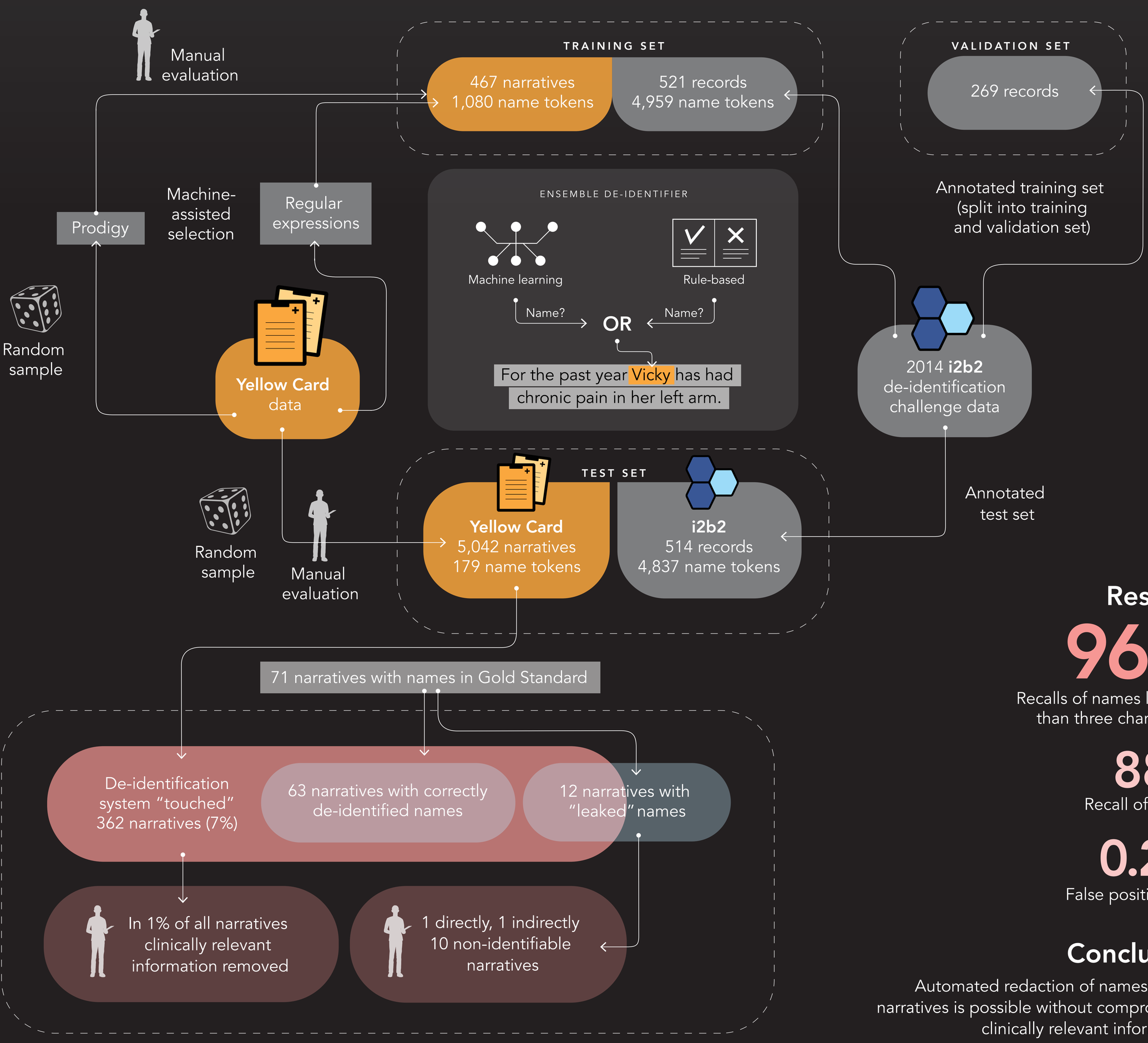
To develop and evaluate a method for automated redaction of names in case narratives.

Data

Training data consisted of i2b2 2014 de-identification challenge data[2] combined with narratives from the Yellow Card scheme[3] provided by the MHRA and annotated using two independent machine-assisted models. Model testing performed on a separate, manually annotated dataset.

Method

An ensemble combining BERT – a transformer-based neural network[4] – with hand-engineered rules for detecting names.



Results

96%
Recalls of names longer than three characters

88%
Recall of names

0.2%
False positive rate

Conclusion

Automated redaction of names in case narratives is possible without compromising clinically relevant information.

References

1. Karimi G, Star K, Lindquist M, Edwards IR. Clinical stories are necessary for drug safety. Clin Med. 2014;14(3):326–7.
2. Stubbs A, Uzuner Ö. Annotating longitudinal clinical narratives for de-identification: The 2014 i2b2/UTHealth Corpus. J Biomed Inform. 2015;58(Suppl):S20–9.
3. Medicines and Healthcare products Regulatory Agency. The Yellow Card scheme: guidance for healthcare professionals, patients and the public [Internet]. [cited 2022 May 12]. Available from: https://www.gov.uk/guidance/the-yellow-card-scheme-guidance-for-healthcare-professionals
4. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:181004805 [cs] [Internet]. 2019 May 24 [cited 2022 Apr 7]; Available from: http://arxiv.org/abs/1810.04805

Uppsala Monitoring Centre (UMC)
Box 1051, SE-751 40 Uppsala, Sweden
+46 18 65 60 60, www.who-umc.org