

Fitting Machine Translation into Clients

(Keynote talk)

Kenneth Heafield 

Institute for Language, Cognition and Computation, University of Edinburgh,
IF 4.21, 10 Crichton Street, Edinburgh, EH8 9AB, Scotland, European Union

kheafiel@inf.ed.ac.uk

Abstract. The Bergamot project is making neural machine translation efficient enough to run with high quality on a desktop, preserving privacy compared to online services. Doing so requires us to compress the model to fit in reasonable memory and run fast on a wide range of CPUs.

Keywords: neural machine translation, efficiency, privacy preservation, model compression.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825303.

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In: Proceedings of the 1st Masters Symposium on Advances in Data Mining, Machine Learning, and Computer Vision (MS-AMLV 2019), Lviv, Ukraine, November 15-16, 2019, p. 1