



European
Commission

Large language models pre-training with a twist

14 June 2022



Welcome & introduction

Tom Vanallemeersch (CrossLang)
SMART 2019/0183 Project representative

Context: ELRC action

European Language
Resource
Coordination



- Minimise language barriers across the EU
- Support continued development of the eTranslation tool (DG Translation)
- Support wider deployment of DG CNECT's services in terms of language resources/tools (sector Multilingualism)
- Service contracts, e.g. SMART 2019/1083 (2020-2022)

<http://lr-coordination.eu>

Activities in ELRC

- Collecting/cataloguing language resources, tools (ELRC-SHARE)
- Helpdesk for data providers
- Assessment of techniques and tools for language processing
- Awareness raising, capacity building: country workshops, **technical workshops**, conferences, social media campaigns

Technical workshops

- Overview of recent research, technology solutions
- Use cases in administrations, industry
- Audience: staff of EU / Member State administrations, ...

Topics:

1. Named-entity recognition in the context of machine translation (June 2020)
2. Neural networks for automatic speech recognition (November 2020)
3. Simplify language – capture audience (November 2021)
4. **Large language models: pre-training with a twist**