Finnish AI Development Programme

Preliminary Study commissioned by

VAKE – the State Development Company of Finland

Technology Industries of Finland

UNIVERSITY OF HELSINKI

BUSINESS FINLAND

Background

- The Finnish Government has through various programmes **opened resources and databases produced by Government**-funded activities
- Finland **supports top research** in AI through the Academy of Finland Flagship FCAI and through the development of language resources for research through FIN-CLARIN and the Language Bank of Finland
- Big global companies have indicated that the **Finnish language community is too small** to be commercially interesting
- Finland needs targeted interventions to **ensure the availability of productivity-enhancing AI services in Finnish** in the public as well as in the private sector

Steering Group

- **VAKE** – State Development Company (Pia Erkinheimo, Tuomas Teuri)
- **Business Finland** (Outi Keski-Äijö, Aki Parviainen)
- **Technology Industries Finland** (Alexander Törnroth)
- **Aalto University** – FCAI (Mikko Kurimo)
- **University of Helsinki** – FIN-CLARIN (Krister Lindén)

Collected opinions from and conducted interviews with more than 50 commercial and public organisations in Finland

Goal

- To support wide usability of **Finnish in various AI applications** and to develop services and technology to achieve this
- The main purpose of the study was to **specify interventions** that will enable the use of AI in Finnish to be channelled through a new or existing actor in the field
- To enable the development of **open language resources for Finnish** (e.g. language data and software) **suitable for use in commercial enterprises**

Interventions

1. **Organization** for developing and maintaining Finnish language resources – sustainability, development, training, benchmarking?
2. **Legal considerations** for an organization to be founded or expanding its operations to collect, use and communicate language data
3. Large, balanced, annotated **corpus of everyday speech**
4. Large **multi-purpose text collection**
5. Ensure that other already **publicly available data sets** are pre-processed and available **for commercial purposes**
6. **Workbench** for transliterating and annotating additional data
7. **Pre-trained language models** based on the language resources
8. Open source **language software components** for rapid prototyping

Next step

Technology Industries of Finland

UNIVERSITY OF HELSINKI

BUSINESS FINLAND

To be published on

October 29, 2019