





Network of language technology researchers in Northern Europe; six university research groups (Denmark, Finland, Sweden, Norway); supported by national e-infrastructure providers in Finland and Norway; multiple national clusters and LUMI; discipline-specific software & data; funding from NeIC, EOSC, HEU, in-kind contributions from all partners.

# HPC-Enabled NLP Research — The Big Picture

## Collaboration Infrastructure

- Distributed team of 25 or so (very) part-timers; mostly a self-help initiative;
- cross-border sharing: everyone can get access to the same superclusters;



# HPC-Enabled NLP Research — The Big Picture

## Collaboration Infrastructure

- Distributed team of 25 or so (very) part-timers; mostly a self-help initiative;
- cross-border sharing: everyone can get access to the same superclusters;

## Virtual Laboratory

- Community-maintained repository of discipline-specific software and data;
- modularity, interoperability, uniformity, reproducibility: our 'own' module-s;



# HPC-Enabled NLP Research — The Big Picture

## Collaboration Infrastructure

- Distributed team of 25 or so (very) part-timers; mostly a self-help initiative;
- cross-border sharing: everyone can get access to the same superclusters;

## Virtual Laboratory

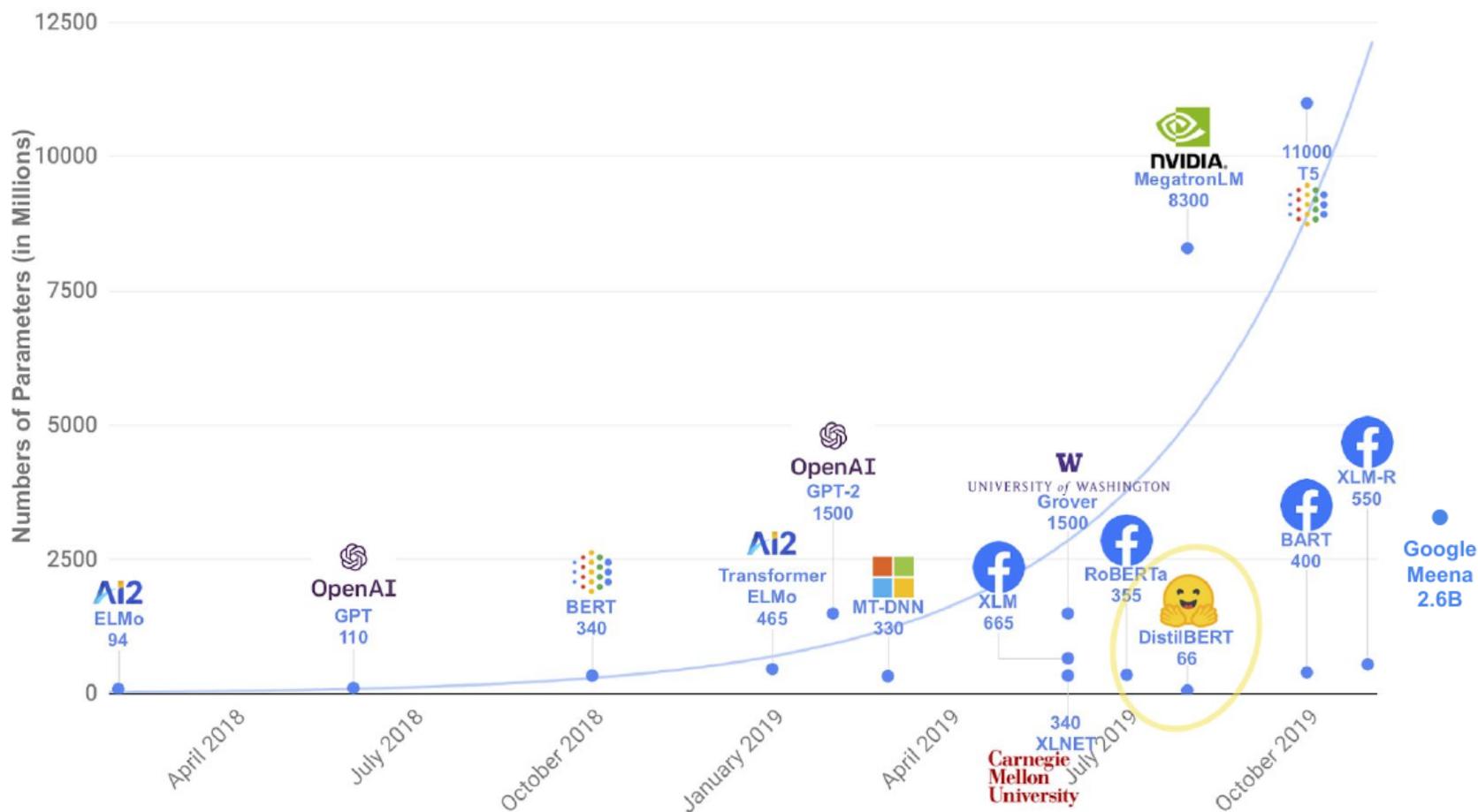
- Community-maintained repository of discipline-specific software and data;
- modularity, interoperability, uniformity, reproducibility: our 'own' module-s;

## Automation & Standardization

- Version deluge: Python × NumPy × Nvidia × PyTorch × own code × ...
- since 2020, automation: repository of EasyBuild recipes; parameterization?
- proprietary third-party dependencies, e.g. CuDNN, NCCL, ROCm, RCCL.



# VLLMs: Representative of Large-Scale Deep Learning



Stephan Oepen

University of Oslo

oe@ifi.uio.no

