

# Hardware and software challenges for massive-scale AI

*Tuesday 10 May 2022 14:55 (30 minutes)*

OpenAI's GPT-3 language model has triggered a new generation of Machine Learning models. Leveraging transformers architectures at billion-size parameters trained on massive unlabeled datasets, these language models achieve new capabilities such as text generation, question answering, or even zero-shot learning - tasks the model has not been explicitly trained for. However, training these models represent massive computing tasks, sometimes performed on dedicated supercomputers. Scaling up these models will require new hardware and optimized training algorithms.

At LightOn - a spinoff of university research -, we develop a set of hardware and software technologies to address such massive-scale computing challenges. The Optical Processing Unit (OPU) technology makes some matrix-vector multiplications in a massively parallel fashion, at record-low power consumption. Now accessible on-premises or through the cloud, the OPU technology has been used by engineers and researchers worldwide in a variety of applications, for Machine Learning and scientific computing. We also train in an efficient manner large language models that can be used for various research and business applications.

**Presenter:** DAUDET, Laurent (CTO and co-founder at LightOn, Professor (on leave) of physics at Université de Paris. )

**Session Classification:** Industry Session