de.NBI in the context of artificial intelligence methods

Abstract: *The German Network for Bioinformatics Infrastructure (de.NBI) supports the German government's AI offensive. The collection, evaluation and skilful use of training data is key to the development of new artificial intelligence methods and their transfer to new application areas such as biomedicine. de.NBI is ideally positioned to make contributions in these areas.*

After decades of relative silence around the topic of artificial intelligence (AI), breakthroughs in deep learning have helped AI methods back into the public awareness. What is less perceived is that the spectacular progress in AI is only made possible by large amounts of available data for learning. Especially in international comparison, it is striking that successful providers of AI-based services deliberately collect large amounts of data in order to gain competitive advantages. On the other hand, the many positive achievements and new products and services must be weighed against data protection. When is the collection of data ethically appropriate, when is it questionable? How can data be used for the general public without harming its owners?

How sensitive data is is also a major topic in related research projects and infrastructures, such as the BMBF-funded Medical Informatics Initiative (MII). The informational self-determination of citizens and patients must be carefully balanced with the benefits for the general public, and a reliable and transparent legal-ethical framework for data use must be created. Likewise, the choice of technical infrastructure (keyword: cloud) must be considered in light of this legal-ethical framework and the necessary security requirements.

In health research and biotechnology, the necessary large data sets must first be collected in experiments and studies. They are then processed using bioinformatic methods and algorithms in order to train AI models. Further data is extracted by experts "by hand" from scientific publications. Both the algorithms and the data maintenance require special expertise to be successful.

Deep learning and AI need large amounts of data, bioinformatics tools, and application domain experts to achieve breakthroughs in the life sciences.

This expertise is provided by de.NBI. In a recent survey of centres within the network, almost all centres mentioned either projects already underway or referred to concrete plans. From de.NBI, there were diverse project proposals within the BMBF's new AI-centred *Computational Life Sciences* funding measure. The methods used are diverse. Both classical methods such as hidden Markov models and support vector machines, but also newer deep learning approaches are used.

The areas of application are equally broad. Deep learning can classify images from microscopy (e.g. sections through cancerous tissue) comparably well as experts, AI can detect the presence of blood poisoning from molecular data much earlier than before, the detection of rare diseases from genome data is possible more quickly, or the microbial composition of faeces can be used for the early detection of colon cancer.

The currency of the AI age is high-quality data. Research infrastructures like de.NBI make these available in a transparent, democratically controlled and directly usable form (FAIR principles). Many petabytes of such data are held at the sites. In addition, de.NBI is developing the tools and infrastructure (e.g. the de.NBI cloud) to bring AI applications to the data.

We need a combination of AI and application expertise to make AI applicable. All of the above examples need a combination of both expertise. We believe it is important and necessary to support both basic research on AI in future funding programmes, as well as not losing sight of **practical application and necessary data sources**. In our view, progress here will only be achieved through **interdisciplinary cooperation**. Many of the planned advances in AI and the life sciences can therefore benefit from collaboration with de.NBI and its European partners (ELIXIR) - and already do in numerous projects. Examples include the *Million European Genomes Alliance* (MEGA+), the *European Lab for Learning and Intelligent Systems* (ELLIS) or the national centres of excellence for AI research (KI-Kompetenzzentren, BMBF).

AI will transform research in the life sciences. This will be facilitated by specialised infrastructures such as de.NBI. At the same time, AI also simplifies the provision of classic services as well as, and in particular, the curation of scientific data.

As an infrastructure in which data from many sources are processed and subject-specific expertise is gathered, de.NBI can make decisive contributions to enabling AI in the life sciences in Germany at an internationally competitive level.

Adopted 21.01.2020