Press release

Technische Universität München
Julia Rinner
08/12/2024
http://idw-online.de/en/news838060

Transfer of Science or Research
Information technology, Medicine
transregional, national

Interview with Prof. Reinhard Heckel: “Data are the crucial component for generative AI”

These days our data are collected everywhere in the internet and are also used to train large language models like ChatGPT. But how do we train artificial intelligence (AI), how do we avoid distortions – known as bias – in the models and how do we ensure that data are protected? Reinhard Heckel, a professor of machine learning at the Technical University of Munich (TUM), takes the time to answer these questions. Prof. Heckel conducts research on large language models and medical imaging applications.

What role do data play in the training of AI systems?

AI systems use data as training examples. Large language models like ChatGPT can only answer questions on topics in which they have been trained.
Most of the information used by general language models for training purposes consists of open-access data from the internet. The more training data that is available for a question, the better the answers will be. For example, if many good texts can be found for an AI system intended to describe mathematical ideas, the training data will be good as well. However, the selection of data is highly filtered at present. Among the vast quantities available, only the high-quality material is collected and used for training.

When selecting data, how are AI systems prevented from generating biases in the form of racist or sexist stereotypes, for example?

It is very hard to develop a method that does not resort to conventional stereotypes and operates fairly and impartially. For example, preventing biased results with regard to skin color is fairly easy. But if we include gender along with skin color, situations may arise where it is no longer possible for the model to be completely unbiased with regard to both skin color and gender.
Consequently, most language models try to provide a balanced answer to political questions and look at several viewpoints. When training AI systems with media content, preference is given to media that meet journalistic quality standards. In addition, the filtering process ensures that texts containing certain words, for example with racist or sexist intent, are not included.

There is a lot more internet content available in some languages than in others. How does that affect the quality of results?

Most internet content is in English. As a result, LLMs in English work best. But there are also huge amounts of content in German. For lesser known languages, however, for which not so many texts exist, training data are not as plentiful. As a result, the models do not work as well.
But it is easy to determine how well language models can be used in certain languages because the models follow so-called scaling laws. This involves testing whether a language model is capable of predicting the next word. The more

training data it has, the better the model will perform. And the performance not only improves over time – it improves in a predictable way. This scaling law is expressed quite well in a mathematical equation.

How accurate does an AI system have to be in practice?

That depends on the field of application. When AI is used to process photos, for example, it is not necessary for every hair to be in place. In many cases we only want the final image to look good. With large language models it is important to have good answers, while details or inaccuracies are not always crucial. But along with language models I also conduct research in the field of medical imaging. Here it is very important for every detail of a generated image to be correct. If I use AI for diagnoses in this area, it has to be absolutely accurate.

In connection with AI there is also much debate about the lack of data protection. How can we ensure that personal data are protected, especially in the medical context?

In most medical applications patient data are used in anonymized form. The actual danger is that there are situations where personal details can be determined using the data. For example, the age or gender of patients can be roughly determined on the basis of MRI or CT scans. Consequently, the data actually contain some of the anonymized information. In such cases it is important for patients to be properly informed.

What other difficulties arise when training AI systems in the medical context?

A big difficulty relates to the collection of data that reflect many different situations and scenarios. AI works best when we apply it to data that are similar to the training data. But the data will differ from one hospital to another with regard to the patient profiles or the equipment that generates the data. There are two ways of solving this problem: either we improve the algorithms or we have to optimize our data to the point where it can be better applied to other situations.

About Reinhard Heckel:

Prof. Reinhard Heckel conducts research in the field of machine learning. He works on the development of algorithms and the theoretical basis of deep learning, with a focus on medical imaging. In addition, he develops DNA data storage and is studying the use of DNA as a digital information technology.

Prof. Heckel is also a member of the Munich Data Science Institute and the Munich Center for Machine Learning.

contact for scientific information:

Prof. Dr. Reinhard Heckel
Technical University of Munich
Professorship for Machine Learning
reinhard.heckel@tum.de

URL for press release: https://www.tum.de/en/news-and-events/all-news/press-releases/details/data-are-the-crucial-component-for-generative-ai

Reinhard Heckel, Professor for Machine Learning
Astrid Eckert / TUM
© Astrid Eckert, Muenchen