

Big data – Introduction

Many people stream YouTube videos, Twitch game videos, Netflix series, and music on Spotify as well as post photos on Instagram, chat on WhatsApp, shop on Amazon, and search with Google. When they participate in these activities, they of course call up a lot of data. At the same time, they also generate a lot of data – day after day, hour after hour, with each click of the mouse. These data streams remain invisible.

For instance, sensors and tools on smartphones and computers record volumes of data. But weather station measuring instruments also collect information and forward it as data to a digital storage location. Moreover, data are gathered via cameras or networked home devices, for example. The sum total of such data is referred to as “big data”.

Big data thus means gathering a lot of data and storing it in a structured manner, for example, in databases. Digitalized data must be processed using computers and complex software. This also includes making targeted queries in databases and linking data systematically. Insights are gained and assumptions confirmed or disproved in this way.

Three examples of big-data applications

Weather reports on television or in weather apps, for example, are based on enormous volumes of measured data from around the world. For decades, weather services have recorded information such as air pressure, temperature, wind, precipitation, and other weather data. As a result, they can recognize typical weather patterns and make forecasts, as well as develop models for climate change.

The advertisements that individual users see posted on social networks such as Instagram and Twitter are also related to big data: Very specific sponsored posts are displayed to each user. These posts contain information that the operators of the respective platform think might interest exactly that person. The advertisement is selected based on gathered and analyzed data – and not only using data of the respective person to whom the advertisement is displayed, but also of users with similar characteristics (age, gender, location, etc.).

The government, such as intelligence agencies or the police, also collects and analyzes data. Its objective for doing so is to solve crimes or prevent them. The combination of data from various sources provides it with new insights, for example, into the areas of a city where crime is increasing.

That raises many questions

Through the years, numerous companies, but also scientists and the government, have collected huge volumes of data via sensors and programs. This practice raises numerous questions that should be looked at more closely and discussed, such as:

- Who is collecting data? What type of data are being collected?
- What insights can be gained from these data?
- How can big data help?
- What risks are involved?
- What about privacy, thus the protection of data?
- What can each individual do to protect his or her own personal data?