

Data science is multifaceted and can be described as a science, a research paradigm, a research method, a discipline, a workflow, and a profession. In 1998, Hayashi Chikio argued for data science as a new, interdisciplinary concept, with three aspects: data design, collection, and analysis. Turing Award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational, and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge. He describes data science as an applied field growing out of traditional statistics. In 1962, John Tukey described a field he called "data analysis", which resembles modern data science. Vasant Dhar writes that statistics emphasizes quantitative data and description. The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name to computer science. "Data science" became more widely used in the next few years: in 2002, the Committee on Data for Science and Technology launched the Data Science Journal. In summary, data analysis and data science are distinct yet interconnected disciplines within the broader field of data management and analysis. This can involve tasks such as data cleaning, data visualization, and exploratory data analysis to gain insights into the data and develop hypotheses about relationships between variables. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. Davenport and DJ Patil declared "Data Scientist: The Sexiest Job of the 21st Century", a catchphrase that was picked up even by major-city newspapers like the New York Times and the Boston Globe. Others argue that data science is distinct from statistics because it focuses on problems and techniques unique to digital data. "Data science" became more widely used in the next few years: in 2002, the Committee on Data for Science and Technology launched the Data Science Journal. In a 2001 paper, he advocated an expansion of statistics beyond theory into technical areas; because this would significantly change the field, it warranted a new name. Data scientists are responsible for breaking down big data into usable information and creating software and algorithms that help companies and organizations determine optimal operations. F. However, the definition was still in flux. Despite these differences, data science and data analysis are closely related fields and often require similar skill sets. "Data science" became more widely used in the next few years: in 2002, the Committee on Data for Science and Technology launched the Data Science Journal. Data analysis focuses on extracting insights and drawing conclusions from structured data, while data science involves a more comprehensive approach that combines statistical analysis, computational methods, and machine learning to extract insights, build predictive models, and drive data-driven decision-making. In contrast, data science deals with quantitative and qualitative data (e.g., from images, text, sensors, transactions, customer information, etc.) and emphasizes prediction and action. Data science and data analysis are both important disciplines in the field of data management and analysis, but they differ in several key ways. For example, a data analyst might analyze sales data to identify trends in customer behavior and make recommendations for marketing strategies. F.