

He describes data science as an applied field growing out of traditional statistics. While data analysis focuses on extracting insights from existing data, data science goes beyond that by incorporating the development and implementation of predictive models to make informed decisions. Vasant Dhar writes that statistics emphasizes quantitative data and description. During the 1990s, popular terms for the process of finding patterns in datasets (which were increasingly large) included "knowledge discovery" and "data mining". Andrew Gelman of Columbia University has described statistics as a non-essential part of data science. Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured, and unstructured data. Vasant Dhar writes that statistics emphasizes quantitative data and description. Jeff Wu used the term "data science" for the first time as an alternative name for statistics. In 1962, John Tukey described a field he called "data analysis", which resembles modern data science. 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. Both fields require a solid foundation in statistics, programming, and data visualization, as well as the ability to communicate findings effectively to both technical and non-technical audiences. As such, it incorporates skills from computer science, statistics, information science, mathematics, data visualization, information visualization, data sonification, data integration, graphic design, complex systems, communication and business. They work at the intersection of mathematics, computer science, and domain expertise to solve complex problems and uncover hidden patterns in large datasets. The modern conception of data science as an independent discipline is sometimes attributed to William S. 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. Data science is multifaceted and can be described as a science, a research paradigm, a research method, a discipline, a workflow, and a profession. The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name to computer science. Stanford professor David Donoho writes that data science is not distinguished from statistics by the size of datasets or use of computing and that many graduate programs misleadingly advertise their analytics and statistics training as the essence of a data-science program. Later, attendees at a 1992 statistics symposium at the University of Montpellier II acknowledged the emergence of a new discipline focused on data of various origins and forms, combining established concepts and principles of statistics and data analysis with computing. Later, attendees at a 1992 statistics symposium at the University of Montpellier II acknowledged the emergence of a new discipline focused on data of various origins and forms, combining established concepts and principles of statistics and data analysis with computing. Data science is a "concept to unify statistics, data analysis, informatics, and their related methods" to "understand and analyze actual phenomena" with data. After the 1985 lecture at the Chinese Academy of Sciences in Beijing, in 1997 C. Cleveland. Data science also integrates domain knowledge from the underlying application domain (e.g., natural sciences, information technology, and medicine). Jeff Wu used the term "data science" for the first time as an alternative name for statistics.