



Artificial Intelligence & Big Data Glossary

Algorithms, fuzzy logic, zettabytes—understanding the language involved in artificial intelligence (AI) and big data can be daunting when you aren't a data scientist. But the potential to leverage data-driven insights to inform strategies, support growth and manage risk means getting familiar with the lingo.

This glossary aims to demystify the process by defining and explaining some of the key terms used by data scientists. In the process, we demonstrate how Nexis® Data as a Service supports companies to take advantage of the opportunities offered by AI and big data.

A

Aggregation - Identifying and collecting data, often combining different datasets.

Nexis® Solutions aggregates more than 40 years of news data from leading sources around the world to provide the broadest insights across brands and markets.

Artificial Intelligence (AI) - Intelligent machines and software that can perceive their environment and act on it, often learning from those actions. AI can be applied in a wide range of fields, including risk and fraud detection, purchase and investment predictions, logistics and supplier management, news and entertainment creation, and online customer support interactions using chatbots.

Analytics - The discovery of insights based on data. There are three types:

- Descriptive analytics summarises data to create an overall narrative.
- Predictive analytics analyses historical and current data to predict future behaviour based on probabilities. For example, it can use trends in consumer preferences or in the stock market to inform buy-sell decisions.
- Prescriptive analytics builds on predictive analytics by analysing its outcomes to decide the best action to take. It is the next evolution in deep learning to support decision-making without human interaction.

Algorithm - A mathematical formula that performs an analysis on a set of data, often embedded in technology.

Application - A software programme that performs a specific function for the person (or application) using it.

Application Programming Interface (API) - An API provides a way to deploy the features of a specific application or service, which lets two applications interact with each other. For example, an API may specify how to retrieve data from an application.

- Bulk API is used to process large amounts of data in batches.
- RESTful API can handle multiple types of data in different formats.

Nexis Data as a Service lets companies use flexible and easy-to-integrate APIs to tap into our unrivalled content universe to support predictive analytics, risk screening and other data-driven use cases.

Archive - Usually a store of historical data that is no longer actively used. Data archives should be indexed for easy location and retrieval of files.

B

Batch Processing - Using computers to efficiently process high volumes of data over a period of time.

Big Data - Very large data sets that can be analysed by computing technologies to reveal patterns and trends. Big data is the fuel for a wide range of AI applications.

C

Classification - Categorising a data point based on traits it has in common with other data points. This allows the user to extract important and relevant information from a big dataset more quickly and easily.

Our award-winning automated classification system, SmartIndexing Technology™, analyses our news data and applies metadata related to more than

7,000 subjects and industries. This enables users to cut through the noise and discover the data needed for predictive analytics and other big data applications.

Computational Notebook - An open-source web application that allows researchers to combine software code and can be used with a number of different programming languages such as Python. The Jupyter Notebook is particularly popular, dubbed “data scientists’ computational notebook of choice” by Nature.

Leveraging the best-in-class Jupyter Notebook environment, Nexis® Data Lab for Academic enables users to search, refine and analyze our extensive collection of enriched licensed data.

Correlation Analysis - Analysing data to determine a positive or negative relationship between different variables.

Our comprehensive news data allows users to identify correlations between, for example, a company’s actions and its reputation.

Customer Relationship Management (CRM)

- A system or strategy used by a company to manage its sales and business processes, which can be informed by big data.

Integrate negative news, company, or legal data into a CRM system to provide additional context that empowers Sales.

D

Data as a Service - Providing data to users over a network on-demand. This allows users to acquire and use external datasets, often in combination with their own data. Data as a Service that uses big data is growing rapidly, and Gartner predicted its market value would nearly quadruple from 2019-2025.

Nexis Data as a Service (DaaS) offers APIs and on applications for delivering highly relevant, archival, and current datasets to power an organisation’s big data projects.

Data Analyst - An employee with the data and statistical skills to interpret and analyse data for insights. This job role is in high and growing demand from companies.

Data Cleansing - Reviewing data to see if it is still valid, as well as correcting errors, eliminating duplicates, and standardizing data formats for greater consistency.

Data Engineering - The behind-the-scenes work to build systems that allow data scientists to do their analysis more quickly and efficiently.

We have decades of experience in managing and engineering data for optimal use by data scientists and other executives within companies.

Data Feed - A stream of data, for example an RSS feed or a social media feed.

Data Governance Framework - The set of rules and processes for how data is organised, aggregated and managed.

Data Journalism - Using data to tell stories and identify patterns and trends. Data journalists have gained prominence with analysis of topics ranging from the impact of political ads in the media to spread of global COVID-19 pandemic and effectiveness of various responses.

Media companies can and do analyse our extensive news data to find trends and stories.

Data Lake - A way of storing a vast amount

of raw data, whether structured, semi-structured or unstructured. This data can be stored within an organisation's data centre or using cloud services.

Data Visualisation - Communicating data visually, often using infographics, colour-coded graphs, or data dashboards.

Data Wrangling - Taking raw data and formatting and restructuring it to make it useful. Data scientists often spend more than half of their time on data wrangling.

Nexis Data as a Service lets users move more easily from data wrangling to data analysis and interpretation. This frees up highly skilled data analysts from doing the more mundane work of cleaning and tagging data to focus on providing new insights.

Deep Learning - Using very large neural networks to solve complex problems, such as facial recognition.

E

Enrichment - The process of making a raw dataset more useful and insightful by normalizing the format and apply tags that make it easier to search and use.

Nexis Data as a Service complements its comprehensive content coverage with a data fabrication, classification, and enrichment process unmatched in the industry.

F

Fuzzy Logic - An approach to logic that is widely used in Artificial Intelligence. Rather than judging whether a statement is true or not, it judges how close to the truth it is.

H

High-Performance Computing - Using supercomputers to solve very complex, advanced computing problems.

I

Internet of Things - Interrelated computing devices, machines and physical objects that exchange or transfer data with each other over the internet. The term is commonly used to describe 'smart homes' in which thermostats, lighting and security cameras can be controlled by connected devices like smartphones.

M

Machine Learning - An application of AI that means computer systems that are able to learn, adapt and improve through experience and without following express instructions. These systems use algorithms and statistical models to analyse patterns of data and draw insights.

Nexis Data as a Service empowers companies to leverage relevant datasets for machine learning, predictive analytics, and other big data applications.

Metadata - Data that describes and gives information about other data - known as "data about data". By summarising basic information, it makes it easier to find and use the data.

Nexis Data as a Service's enrichments cover 125 descriptive metadata applied to our news content including headline, topic, index time, publisher, country language, editorial source rank, source topic and news category.

N

Natural Language Processing - A type of AI concerned with the interactions between computers and the human language, particularly how to programme computers to process and analyse large volumes of natural (ordinary) language data. The technology can 'understand' text documents, including nuances in the language, and accurately extract information and insights from them. NLP is an example of machine learning.

We have been fine-tuning our NLP for many years to improve search relevance across our platforms and Data as a Service APIs.

Neural Networks - A system of connected nodes like neural connections in the brain that are used as a method of Machine Learning. Connections between the layers lead to outputs and a prediction.

Normalisation - The process of reorganising data in different databases to make comparisons between the data easier and more meaningful.

P

Pattern Recognition - Identifying patterns in data, usually via algorithms, which allows predictions to be made when similar data is encountered.

Q

Quantitative Analysis - Using algorithms to find insights from large amounts of quantitative data. This is particularly useful in the financial sector, where trading decisions are often made by quantitative analysis of high volumes of numerical, financial data.

R

Robotic Process Automation - Software that is programmed to do repetitive and often mundane tasks. RPA deploys robots to improve efficiency and free human resources for more high value tasks. It can have a dramatic impact on productivity, efficiency, and accuracy within business processes, such as fraud detection and risk mitigation.

Our targeted datasets are designed to fuel Robotic Process Automation which can optimise the efficiency and effectiveness of supplier and risk management processes.

S

Smartindexing - A classification technology that helps researchers to find relevant information from large volumes of data by tagging documents. This is particularly useful for research.

Nexis Solutions uses SmartIndexing Technology on our content universe, which allows searches based on topics to surface the most relevant results.

T

Text Analytics - Deriving insight or meaning from text-based sources. This can be done by applying linguistics, machine learning and statistical techniques.

Training and Testing - This is a key part of the process of Machine Learning. A predictive model uses a set of training data to build understanding, then it uses what it has learned to predict outcomes based on similar data.

U

Unstructured Data - Data that has not been organised in a pre-defined manner. It is often full of text, dates, numbers, and facts, and requires a lot of effort to make it useful.

Semi-Structured Data - Data that does not have a structured format and cannot be contained in a database of rows and columns, but a hierarchy has been established using tags or other markers.

Nexis Data as a Service allows for standard and flexible integration of a semi-structured XML data feed into any database or application.

V

Veracity - The extent to which data is (or isn't) accurate and correct, which in turn determines how effective analysis performed on the data is.

X

XML - A way of tagging data to describe it.

Z

Zettabytes - A measurement for an enormous amount of data - bigger than an Exabyte and a Terabyte, but smaller than a Yottabyte. It was estimated that the digital universe comprised 44 zettabytes by the end of 2020. Data is being created at an exponential rate - 90% of the world's data has been generated in the last two years alone.

Conclusion

If this glossary has demystified the process behind Big Data and Artificial Intelligence, why not talk to us about how Nexis Solutions can support your business to find new insights and manage risk?

Accelerate your time to insight with flexible APIs that connect you to an unrivalled collection of enriched news, company, legal, regulatory, and patents data from Nexis Data as a Service.

- Bulk APIs to access high-volume feeds of news, company, or patents data for historical analysis and predictive analytics.
- Search & Retrieve APIs to access data on demand with enhanced functions—search, retrieve and saved searches—that work within your own user interface.

About Nexis Data as a Service

With an unmatched global content collection enhanced through normalization and metadata enrichments, Nexis® Data as a Service (DaaS) delivers the highly relevant, archival and current data organizations require for a broad range of artificial intelligence applications.



“

Information is the oil of the 21st century, and analytics is the combustion engine.

PETER SONDERGAARD

Former Executive Vice President, Gartner

For more information, contact your Nexis Solutions representative or reach out to us directly:

 internationalsales.lexisnexis.com

 +31 20 485 3456

 information@lexisnexis.com



LexisNexis, Nexis and the Knowledge Burst logo are registered trademarks and Nexis Diligence is a trademark of RELX Inc. Other products or services may be trademarks or registered trademarks of their respective companies. © 2021 LexisNexis. All rights reserved. SEG-DaaS-BigDataAIGlossary 0221