
Department of Data Science
Current News on Innovation & Technology

Get AI's mojo back intelligently

As AI emerged, it was met with optimism, but soon concerns over AI safety, explainability, data privacy, and deepfakes overshadowed the excitement. Experts began to caution that AI could threaten humanity's existence. To harness the positive potential of AI, we must rekindle the initial hope, joy, and curiosity associated with technological advancements.

AI entered the world accompanied by optimism. But, soon, the narrative took a darker turn. There were rumblings around AI safety and explainability, data privacy, and deepfakes, and pundits began to warn us about an AI putting humankind's existence at peril. To realise AI's positive potential, we must return to the days of hope, joy and curiosity we all felt when creating new science and tech.



Source: <https://economictimes.indiatimes.com/opinion/etcommentary/get-ais-mojo-back-intelligently/articleshow/111586226.cs>

New York Times Fights OpenAI's 'Unprecedented' Bid for Journalistic Materials



The copyright battle between The New York Times and OpenAI intensified as the AI company pushed for access to journalists' notes and memos in a contentious discovery dispute.

The Times [sued OpenAI in December 2023](#), contending that ChatGPT was trained on its articles without permission, with the chatbot able to generate “near-verbatim excerpts” from its articles.

As the dispute develops, OpenAI is trying to access materials it argues are crucial to its defense as part of the pretrial phase of the lawsuit.

The Times, however, is trying to prevent OpenAI from accessing what it considers to be an “overbroad and unduly burdensome” amount of material “not relevant to any party’s claims or defenses.”

OpenAI is demanding copies of the related reporter’s notes, interview memos and records of materials and articles claimed to have been copied by ChatGPT.

The AI developer argued in a [July 1 memo](#) to U.S. District Judge Sidney H. Stein that its discovery demands were relevant to both the Times’ infringement claims and OpenAI’s defenses, such as [fair use](#).

OpenAI argued that simply providing the actual works and not the wider related materials that went into creating them would be “insufficient” to allow them to test the assertions that the articles are the Times’s “original works of authorship” entitled to copyright protection.

Related: [The New York Times sues OpenAI, Microsoft](#)

Source: <https://aibusiness.com/nlp/new-york-times-fights-openai-s-unprecedented-bid-for-journalistic-materials>

TIET ties up with Nvidia for AI, data science school

Nvidia will assist in training the faculty of Thapar School of Advanced AI & Data Science, offering guidance on curriculum enhancements. Additionally, it will provide access to its accelerated computing platforms for research purposes and support qualifying AI startups incubated at TIET through its inception programme.

Thapar Institute of Engineering & Technology (TIET) Friday said it has tied up with Nvidia, one of the world's biggest producers of graphics processing units, to set up a school of advanced AI and data science.

Nvidia will provide technical support to the school from the Nvidia AI University programme.

TIET will also launch a Bachelor of Technology in artificial intelligence and machine learning from this year.

"We have already received AICTE approval for the same. This year, we are looking at starting with a batch size of 120 students," said Padmakumar Nair, director of Thapar Institute of Engineering & Technology. "This also establishes our vision of providing education that creates future-ready students."

TIET is also launching a master's programme in AI, machine learning and data science.

Source: <https://www.firstpost.com/tag/data-science/>

Author Article on- The Importance of Data Analytics



Why Data Analytics is important

Data Analytics is a good choice for those who enjoy working with numbers, and solving puzzles, writes **Rohini R Rao**

We are living in the era of 'datafication,' technology is recording every aspect of our lives as data. People and their devices are constantly connected to the internet and an unprecedented amount of data is being generated. Data recording the daily lives of humans can be integrated, analysed, and put to social use. However, data analysis is not a trivial task, we are in a 'data-rich' but 'information-poor' situation. Although a massive amount of data is being recorded in detail, the data repositories are not being used effectively. Large amounts of data, in various formats, are distributed across repositories. Decision-making is mostly based on intuition rather than information. There is a need to integrate this data, prevent information overload, and use the processed information to make data-driven decisions. There is a need for making sense of the data.

Story building

Data Analytics is the science of analysing the raw data to summarise and visualise the data. This insight into the data should help the decision-maker interpret it and make 'data-driven' decisions. The first step of the Data Analytics project is to specify the problem and the objective of the project. The objective of the analysis could be to simply summarise the data for an informed deci-

sion. For instance, the analysis could help identify customers for the next marketing campaign. The data can also be used to build predictive models or make personalised recommendations.

The team involved in Data Analytics facilitates the collection of relevant data and its exploration. Sources of data are identified or collected, data is cleaned, converted, integrated, and prepared for analysis. Tools such as Python, R, and MS Excel, various techniques are deployed for the analysis and visualisation. Data Analysts transform the data to interpret patterns and trends, which help the team translate the patterns into actionable items. Data Analysts must have exceptional interpersonal skills, in addition to technical

Essential skills

To understand the problem, and measure the outcome is crucial. While building predictive models, it is important to declare the level of accuracy of the model. Data from medical records, and patient history could be analysed to predict and quantify the risk of a particular disease. The consequences of incorrect predictions need to be quantified for effective decisions. Programming skills in Python, R, or Matlab are essential. The ability to do data visualisation using matplotlib, seaborn, PowerBI or Tableau is important.

skills like programming, data management, data analysis, and data visualisation tools. Expertise in story building, statistical analysis, data visualisation, machine learning, and a thorough understanding of cloud platforms, is expected. Data Analysts may choose to specialise in a particular domain like business, finance, or healthcare.

Make a strong CV

Students must hone all the required technical skills. However, the technical tools and programming languages used for the project do not matter. Data Analytics is all about analytical thinking, finding patterns in data and how to use it in decision making. Students can acquire this skill by dabbling in projects and in-

Data intuition is the most important skill that a data analyst must have to excel

ternships. The focus should be on putting the theory learnt into action. Students can dabble in the numerous datasets hosted online on websites like Kaggle. As freshers, students should pay attention to data pre-processing tasks like dealing with missing or noisy data and encoding and integrating data. Students can participate in playground training competitions that are beginner-friendly. Usually, the objective of the challenge is defined, and the student can work on finding the solution. They can learn from the online notebooks of experts who build industry-grade data analysis solutions. From there on the student could work on data sets wherein they can identify the problem statement, and objectives. The student can dabble in story building using data summaries, and visuals. They can also learn to make recommender systems or prediction models. The student can share their data analytics portfolio online to showcase their abilities.

(The author is programme coordinator, BTech Data Science & Engineering, Department of Data Science and Computer Application, Manipal Institute of Technology, Manipal Academy of Higher Education)



How Do You Make a Data Scientist? Not Easily.

Having a facility with numbers helps. But that's only a small part of the equation.

BY DEBORAH GAGE

THE PATH TO becoming a data scientist is not a clear one. And that's by design.

Consider the data-science team at **Alpine Data**, a San Francisco software startup that helps companies analyze their data to make predictions about their businesses. It includes a former marketing manager, a former physicist, a former operations researcher and a former business consultant. Helping the team as well is a former mathematician who was hired as a software engineer.

"We strongly believe that having people from different backgrounds collaborating around a problem is more im-

One test: Can a job candidate pick out the most interesting data?

portant than selecting some fancy algorithms," says Alpine co-founder Steven Hillion.

In other words, despite its name, data science isn't just about being skilled with numbers. Rather, an effective data scientist also has an ability to see how particular subsets of data may be more useful than others, and what conclusions can be drawn from them.

The term data science didn't even emerge until about 2008, when it was becoming clear that the volume of data being accumulated was beyond the capacity of humans to analyze or comprehend without a machine's help. The ability to analyze billions of rows of data with hundreds of thousands of variables opened new frontiers in environmental science,

medicine, politics, history and dozens of other fields.

Human factors

But as the oceans of data have grown, so has the need for people who can understand statistics and machine learning, work with complex data sets and software, and explain it all to customers.

Mr. Hillion, who has a Ph.D. in mathematics, says he saw the need for data scientists at his previous company, Greenplum, now part of EMC Corp., and had to develop techniques for creating them because there weren't enough people who could do the job. He uses the same methods at Alpine.

One test, he says, is whether a job candidate, given a choice of data sets, can pick out and work with the most interesting one. The test he designed at Alpine, for example, includes a set of New York Police Department data on motor-vehicle collisions in New York City that can be subdivided in several ways—by number and types of vehicles and drivers, number and types of injuries and deaths, contributing causes and several types of locations. Alpine Lead Data Scientist T.J. Bay, who made the data part of the test, says it stood out to him because of the number of interesting fields that could be used to help visualize and predict accidents.

It's a particularly good test for a data scientist, Mr. Hillion says, because "it was literally, in a tech sense, multidimensional. You can break it down by geography, time, vehicle type, accident type, driver characteristics and so on. And there's no one aspect of it that is obviously a path you should go down." Also, he says, the results are something everybody's interested in—how to avoid accidents.

Every Friday, team members explain their projects and give



Alpine Data co-founder Steven Hillion (left, with T.J. Bay), values data scientists' diverse skills.

one another feedback. Given all the skills a data scientist needs, Mr. Hillion says, "you can't have all that in one person."

Studying accidents

Once Alpine participated in a challenge to analyze several years of U.S. traffic accident data to better understand trends in and causes of serious accidents. Lead Data Scientist Emilie de Longueau pored over the data and isolated variables to analyze, visualize and ultimately predict the severity of injuries and accidents.

But Mr. Hillion thought her visualizations were "a little dull," he says, so he asked the other members of the team to look again at the data. While one engineer focused on Ms. de Longueau's visualizations, another asked her to explain in detail why she chose certain data and how she made her predictions.

He suggested using a new algorithm he'd devised. When the analysis was complete, two product managers—whose focus is to translate technical into business concepts—then used it to create an easy-to-use Web application. The app gives users a forecasting tool for estimating rates and sever-

ity of traffic accidents based on variables such as rates of drunken driving or speeding.

The finished product, Mr. Hillion says, "created a way to take the machine learning [that was first applied to the data] and make it usable by the average person."

To achieve such insights, Mr. Hillion says he hires people who can design algorithms, people who can write code to make the algorithms work on different computer systems, and people who can apply those algorithms to customers' data and then explain what they've done. That last set of skills Mr. Hillion refers to as "the human layer."

Another essential skill: knowledge of an industry. Alpine Senior Data Scientist Anshuman Mishra, who is researching how a financial-services company can detect money laundering and fraud, is a former derivatives trader, while Ms. de Longueau, who has a master's in engineering in operations research, is working on supply-chain and workforce management.

Wall Street science

Big companies with deep pockets—and lots of data—of-

tees," Mr. McMillan says. "Really what I'm focused on is connecting the science with the practice."

Universities get involved

Helping aspiring data scientists forge their own career paths, more universities are offering programs in data science or analytics.

The University of California, Berkeley, is in its second year of a program to make data-science classes available to all undergraduates. So far, about 1,200 students from 60 majors have enrolled.

So-called connector courses are available that help them apply data-science techniques to specific areas, such as environmental engineering. An ethics class is taught as well, so students can think about "the boundaries that could be crossed if data is not used responsibly," says Cathryn Carson, an associate professor of history.

Berkeley is trying to bring students in other fields, and from underrepresented and underprivileged groups, into data science because they bring diverse perspectives, Dr. Carson says. An anthropology major, for instance, "will think deeply about the social contexts and human contexts that gave rise to the data," she says. "What kinds of questions were prompted to generate this data? Were those good questions or biased questions?"

Interest from public-health students at Berkeley drove a project to study data on child mortality in different countries. Some social-psychology students, meanwhile, want to study how humans react to mobile data collected about their health.

Dr. Carson says of such students, "They're also appreciating the social good that can be done by working with examples of human welfare, rather than just data about Twitter."

Ms. Gage is a writer in San Jose, Calif. She can be reached at reports@wsj.com.

Source: <https://x.com/EricTopol/status/841291466836074497>