

Module 1 :- Assignment 1.

1. Report on the Evolution of Data Science :-

Trace how data science has evolved over the years.

* Early Stages (Before 1980s)

- Originated from statistics and Mathematics
- Early computers enabled basic data processing

* Growth Phase (1980s-2000s)

- Development of databases and data warehouses
- The rise of the Internet led to more data generation.

* Modern Data Science (2000s - 2010s)

- 2001: Data Science recognized as a separate discipline.
- 2008: Data Scientist became a popular job title.

- Use of tools like Python, R, Hadoop and Spark.

* Current Era (2010s - Present)

- Integration with AI and machine learning
- Cloud computing enables large-scale data analysis.
- Focus on real-time analytics.

3. Essay :- How Generative AI is Transforming DataScience - Explain the impact of AI-generated data.

- GenAI, uses sophisticated algorithms to organize large, complex data sets into meaningful clusters of information in order to create new content, including text, images and audio, in response to a query or prompt.

GenAI typically does two things:-

First, it encodes a collection of existing information into a form that maps data points based on the strength of their correlations (dependencies).

Second, when prompted, it then generates (decodes) new content by finding the correct within the existing dependencies in the vector space.

Positive Impacts of AI-generated Data:

1. Enhanced Training for AI Models

- AI can generate synthetic data to train other AI systems, especially when real data is limited, expensive or sensitive.

For ex - in medical imaging, synthetic X-rays can help train diagnostic models without using real patient data.

2. Cost and Time Efficiency

- Creating data using AI is often faster and cheaper than collecting real-world data.
- This benefits startups or researchers with limited resources.

3. Creative Content Generation:

AI can create realistic text, images, audio and videos - used in marketing, entertainment, game design and education.

- Negative Impacts of AI-Generated Data.

1 Misinformation and Deepfakes: AI can generate fake news, videos or images that are hard to distinguish from real ones, leading to confusion and manipulation.

2 Ethical and legal issues:

Ownership and copyright concerns arise when AI generates data.

Who owns the content - the user, the tool or the original data source?

3 Job Displacement:

AI-generated content can replace tasks done by writers, designers, or journalists raising concerns about employment.