

“The Future of  
AI-Driven Data”



Jonathan Knight



Data Analyst & IT  
Trainer &  
Learning Coach



## **Session Overview**

- AI – A Definition
- Context of Data Generation
- How does AI transform Data into insights
- The Key Elements of AI
- Co-Pilot demo in excel
- Security and Ethical Considerations
- The Future of AI



“ So, what does an AI-Driven Data world look like?”

“What do we mean by it ?”

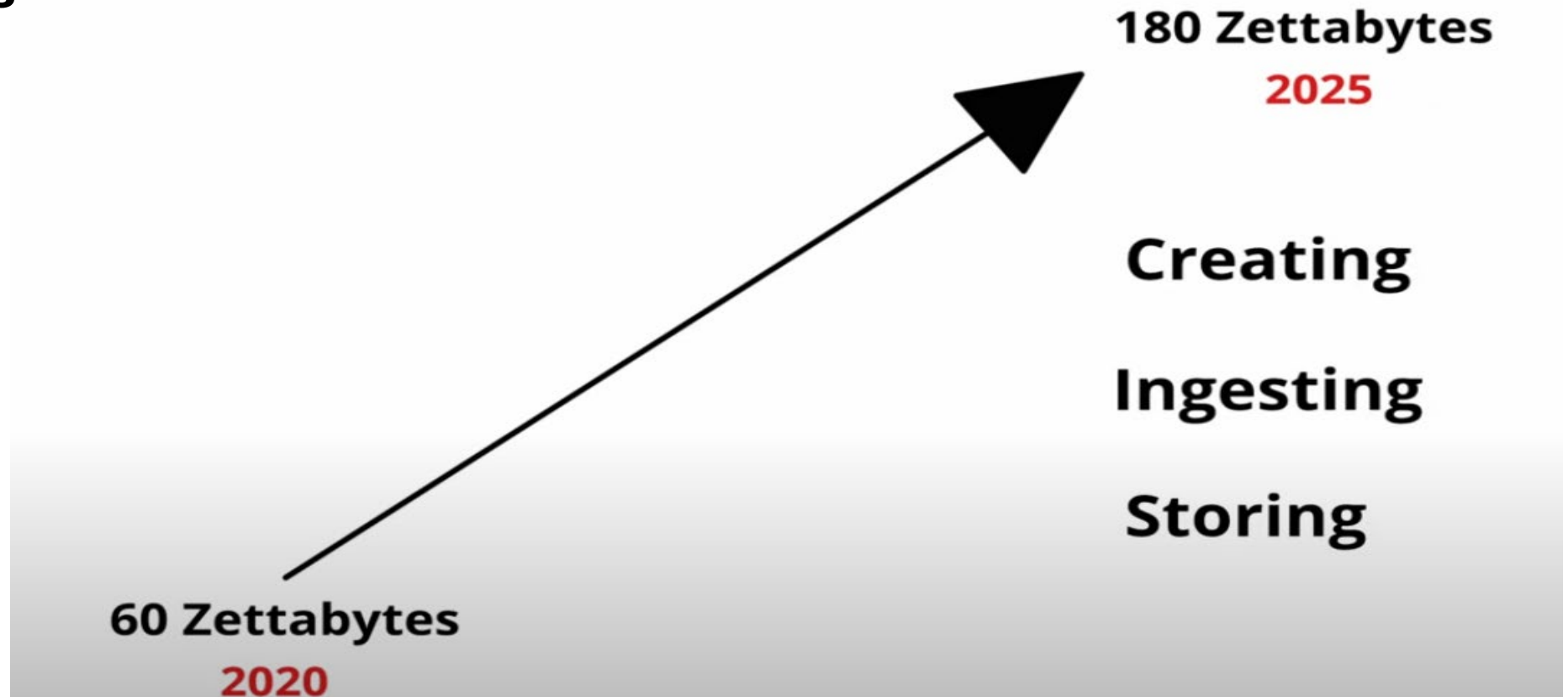


## **Definition:**

**"AI-driven data"** refers to a paradigm where artificial intelligence (AI) and machine learning (ML) technologies are used to manage, process, analyse, and act upon data throughout its entire lifecycle, rather than relying solely on manual, human-driven, or traditional rule-based methods.

It involves using AI to transform vast, often chaotic, and complex data into clean, structured, and actionable insights with high speed and precision.

## Data Generation over the last 5 years





- **1 KB** = a short email without attachments.
- **1 MB** = a high-quality photo.
- **1 GB** = about 250 songs or 1 hour of HD video.
- **1 TB** = roughly 250 full-length HD movies.
- **1 PB** = all the data on Facebook (as of a few years ago).
- **1 EB** = used in massive data centres and global internet traffic.
- **1 ZB = 1 Zettabyte = 1,000,000,000,000 Gigabytes (GB) or  $10^{21}$**

## NoSQL



Gaming



Social



IoT



Web



Mobile



Enterprise



Key/value store



Document database



Column family store

## SQL



Web



Mobile



Enterprise



Data mart



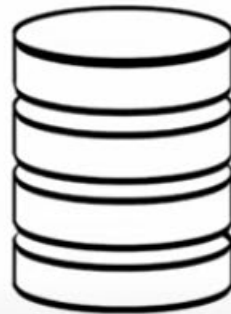
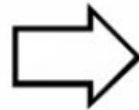
Relational table storage



Relationships use joins

# Structured Data

**Normalized**



Relational Database

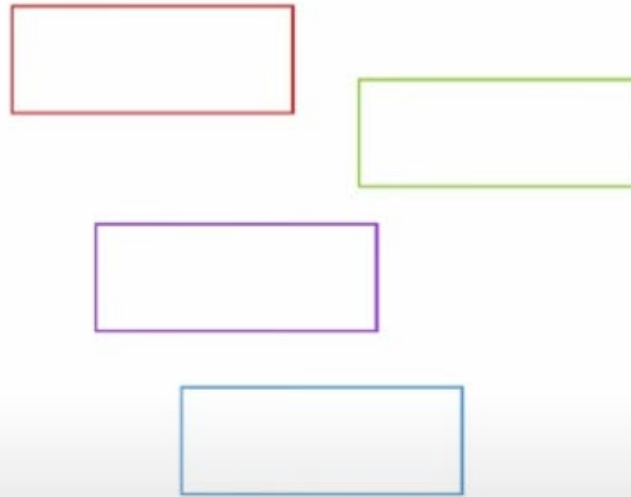


**SQL**

# Unstructured Data



Structured



Variable

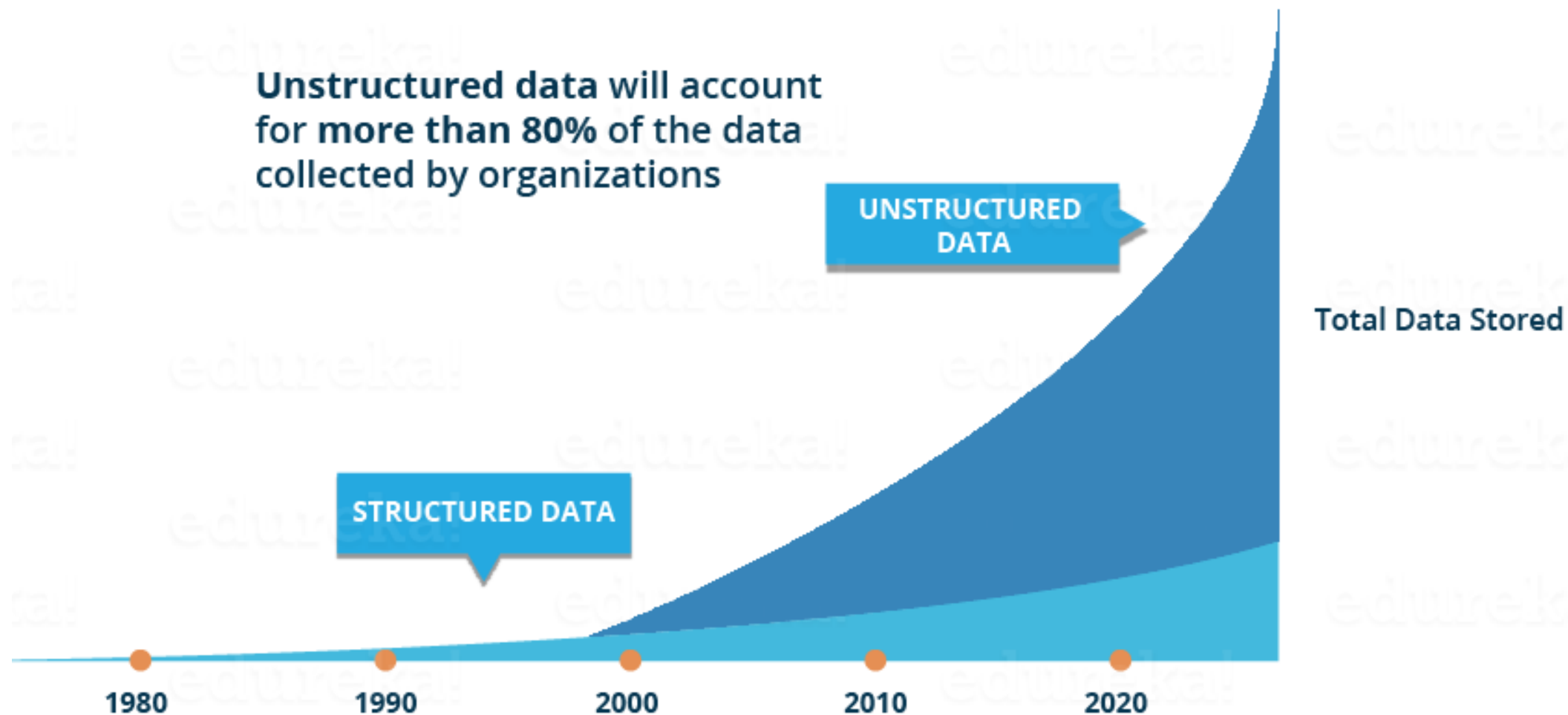
JSON

XML

HTML

Email

**Unstructured data will account for more than 80% of the data collected by organizations**



# The Datalake

Second response to the growth in data

2005

**BIG  
DATA**

2010



Ability to mine and analyze large amount of data

## Machine Learning

Flexibility



Scalability



Cost-Effectiveness



Innovation



Data Quality



Governance



Security



Integration

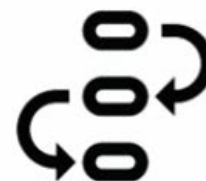


# The Data Lakehouse

Third response to the growth in data



Transactions  
Concurrency control  
Time travel  
Audit history  
Backup and recovery  
Disaster recovery



Data pipelines

Availability  
Reliability



databricks



## Transforming the Data Lifecycle

**Data Collection & Integration:** AI automates the ingestion of data from fragmented sources (Silos), connecting and mapping datasets with minimal human intervention.

**Data Cleansing & Preparation:** AI identifies and corrects errors, fills in missing values, and standardises data formats far faster than humans.



# The Security Perspective

## Key Data Security benefits in AI

AI acts as a "**watchdog**," continuously monitoring network activity, detecting anomalies (like security breaches), and automatically tagging sensitive information for compliance.

Data Discovery: AI scans network devices and data repositories to locate, catalogue, and index new data, uncovering "**shadow data**" that would otherwise go unnoticed.



## Key Data Security Risks in AI

**Data Poisoning & Corruption:** Attackers inject malicious data into the training set to create backdoors or cause the AI to make incorrect, unsafe, or biased predictions.

**Prompt Injection Attacks:** Attackers use specifically crafted prompts to bypass guardrails, causing LLMs to leak confidential information or perform unintended actions.

**Data Leakage & Privacy Inversion:** AI models may inadvertently reveal personal or proprietary data used in their training or allow attackers to infer sensitive information through model outputs.



**Cont...**

**Intellectual Property Theft:** Attackers probe models to replicate their functionality, stealing the proprietary knowledge or data embedded in the model.

**Adversarial Examples:** Subtle, often imperceptible, manipulations of input data that cause an AI to misclassify or misinterpret information.

**Unsecured APIs & Supply Chain Vulnerabilities:** Third-party APIs and libraries used to build AI systems can be compromised, leading to unauthorised access to data.

## Ethical considerations

- **Algorithmic Bias and Discrimination:** AI systems can inherit prejudices from training data, leading to biased, discriminatory, and inaccurate outcomes, especially for vulnerable groups.  
[Artificial Intelligence: examples of ethical dilemmas | UNESCO](#)
- **Privacy and Surveillance :** The immense data requirements for AI, including facial recognition and personal information processing, pose risks to surveillance and violations of personal privacy.
- **Accountability and Liability:** : It is often unclear who is responsible when AI systems fail or cause harm—the developers, users, or the AI itself.

## What are the Core Technologies?

AI-driven data combines several advanced technologies:

- **Machine Learning (ML):** Algorithms that improve automatically through experience, identifying patterns and correlations in data without being explicitly programmed.
- **Deep Learning:** A subset of ML that uses neural networks to analyse complex datasets, such as images or videos.



## Cont.....

- **Generative AI:** Used to create summaries, narratives, or even synthetic datasets to fill in gaps.
- **Natural Language Processing (NLP):** Enables AI to analyse unstructured data like documents, emails, and social media posts, and allows users to interact with data using natural language queries (e.g., "What was the sales revenue in London last quarter?").



## What are the key Capabilities in Data Analytics?

**Real-Time Analytics:** Processes data instantly as it is generated, allowing for immediate responses to market changes or operational issues.

**Predictive and Prescriptive Insights:** Beyond just reporting "what happened," AI forecasts "what will happen" (predictive) and suggests "what should we do" (prescriptive).



**Cont.....**

**Unstructured Data Analysis:** Can analyse non-traditional data, such as images, videos, and free-form text, which are inaccessible to traditional tools.

**Data Democratisation:** Makes data accessible to non-technical users via conversational interfaces, allowing them to gain insights without needing deep data science skills.



## AI Driven Data vs Traditional Data

- **Traditional:** Manual, rule-based, backward-looking (retrospective), and struggles with high volume/complexity.
- **AI-Driven:** Automated, self-learning, proactive, and scales easily with large datasets.

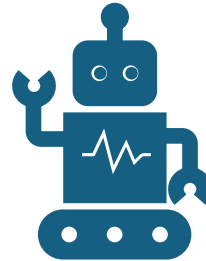


# AI in action – Co-pilot within Excel





Question?



“ Will AI replace my Data role?”



“How will my Data role evolve with AI”

# The Evolution of the Data Professional: Thriving in an AI-Driven World

**Goal:** To help data analysts and technicians understand how their roles are changing and identify the skills needed to stay competitive as AI becomes a standard workplace partner.

The data landscape is shifting from manual, descriptive reporting to AI-augmented predictive and prescriptive action. While AI automates "grunt work" like data cleaning and basic coding, it elevates the human role to one of strategic oversight, ethical stewardship, and complex storytelling.

## From Technical Execution to Strategic Oversight

**The Shift from "Doing"**

**Supporting:** AI handles routine SQL and cleaning, while humans oversee pipelines and validate outputs.

**to "Orchestrating"**

## Projected Impact of AI

	Forbee / McKinsey <b>40%</b> productivity gains	<b>\$2.6T-\$4.4T</b> annual global economic value
	PWC Study <b>\$15.7 trillion</b> global game-changer by 2030	
	Descriptive (Reports)	Prescriptive (What should we do?)

## Human-in-the-Loop Responsibility

Guardians against AI hallucinations.

**Supporting Detail:** Analysts must transition from report builders to "human sign-off" guardians against AI hallucinations.

## The "Stop, Start, Amplify" Framework

 <b>Stop</b> Manual reporting	 <b>Start</b> Bias detection	 <b>Amplify</b> data storytelling and business influence.
-------------------------------------	------------------------------------	---

## The Future-Proof Skill Roadmap

 <b>Adopt an "AI-First" Mindset</b>	 <b>Master AI-Collaboration Skills</b>	 <b>Double Down on Human Context</b>
--	---	---

**Supporting Detail:** Never start from zero; use AI for initial drafts, meeting summaries, and mock datasets.

**Supporting Detail:** Prioritise prompt engineering, AI-assisted workflow design, and critical output auditing.

**Supporting Detail:** Deep domain knowledge and stakeholder influencing remain the most valuable human-only assets.



**Agentic AI and Autonomous Action:** By 2025–2026, AI is shifting from passive chatbots to "agents" that can execute multi-step workflows (e.g., browsing the web, managing calendars, using software) with minimal human supervision.

**Physical World Integration (Humanoid Robots):** The next frontier involves bringing AI into the physical world, with "general-purpose" humanoid robots expected to be deployed in real-world settings (factories, warehouses) by 2030, according to some experts.



**The "Post-Moore" Era:** To handle the massive energy requirements of AI, computing is moving beyond traditional silicon chips toward new architectures, such as neuromorphic computing (mimicking the brain) and optical computing.

**Data Scarcity and Synthetic Data:** As high-quality human data runs out, AI training will increasingly rely on "synthetic data" (AI-generated data) to continue improving.



## **Long-Term Vision (2050 and Beyond)**

Experts suggest that by 2050, the line between human and machine may blur through cybernetic augmentations (e.g., implants to monitor health or assist communication). The long-term, highly speculative vision includes a "post-scarcity economy" where AI-managed agriculture and manufacturing solve basic human needs.



Thank you



**Any Questions?**



