



# Understanding Data

 @ankitrathi



## What is Data?

*Its raw information in the form of numbers, text, images, or symbols*

## Data Formats



*Structured (spreadsheets, databases)*

*Unstructured (emails, videos, social media posts)*

*Semi-structured (JSON, XML)*

## Data Processing Cycle

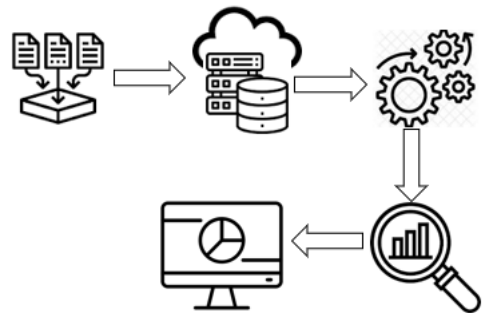
*Collection - Sensors, surveys, transactions*

*Storage - Databases, cloud, servers*

*Processing - Sorting, filtering, analysing*

*Analysis - Trends, patterns, insights*

*Visualization - Graphs, charts, dashboards*



## Data Types & Examples

*Quantitative (Numbers) → Sales figures, temperature*

*Qualitative (Descriptions) → Customer reviews, comments*

*Big Data (Massive sets) → Social media trends, IoT sensor data*

## Importance of Data

*Better Decisions - Business strategies, healthcare, AI*

*Efficiency - Automation, predictive models*

*Innovation - Machine learning, scientific research*



## Data Challenges

*Data Privacy & Security - Hacks, leaks, GDPR*

*Data Overload - Too much data, hard to analyze*

*Bias & Accuracy - Incorrect or misleading data*

## Understanding

# Data Analysis



in @ankitrathi

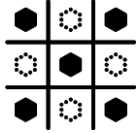


### What is Data Analysis?

*Process of cleaning, transforming, and interpreting data  
To find meaningful patterns, trends, and insights*

*Goal: Convert raw data into useful knowledge for decision-making*

*Like solving a puzzle—each data point is a piece that helps complete the big picture*



### Why is Data Analysis Important?

*Better Decision-Making - Data-driven insights lead to smarter choices*

*Problem-Solving - Identifies inefficiencies, risks, and opportunities*

*Predicting Trends - Helps businesses prepare for future changes*

*Competitive Advantage - Effective data analysis outperform others*

### Types of Data Analysis

*Descriptive Analysis - "What happened?" (sales reports, trend charts)*

*Diagnostic Analysis - "Why did it happen?" (correlation, root cause analysis)*

*Predictive Analysis - "What might happen?" (forecasting, machine learning)*

*Prescriptive Analysis - "What should we do?" (decision-making models)*



### Common Data Analysis Techniques

*Statistical Analysis - Mean, median, variance, hypothesis testing*

*Data Visualization - Charts, graphs, heatmaps for easy understanding*

*Correlation & Regression - Finding relationships between variables*

*Machine Learning Models - AI-driven pattern recognition*

*Text Analysis - Extracting insights from words and language*

### Challenges in Data Analysis

*Dirty Data - Incomplete, inconsistent, or incorrect data*

*Data Overload - Too much data without clear focus*

*Bias & Misinterpretation - Drawing incorrect conclusions*

*Lack of Skills & Tools - Not everyone is trained in data analysis*



# Data Engineering

 @ankitrathi

## What is Data Engineering?

*It is the process of designing, building, and maintaining the systems that collect, store, and process data*

*Goal: Ensure data is accessible, reliable, and ready for analysis & AI*

*Like plumbing for data—moving and cleaning data so it's ready for use*



## Why is Data Engineering Important?



*Reliable Data - Ensures accurate, well-structured data for analysis & AI*

*Scalability - Handles large-scale data efficiently*



*Faster Insights - Automates data flow for real-time analytics*

*Foundation for AI - AI & ML models rely on well-prepared data*

## Key Components of Data Engineering

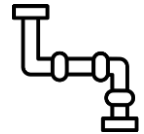
*Data Collection - Extracting data from sources (APIs, databases, logs)*

*Data Storage - Storing data in Data Lakes, Warehouses, or Lakehouses*

*Data Processing - Transforming raw data using ETL (Extract, Transform, Load) / ELT*

*Data Pipelines - Automating data flow using batch & real-time processing*

*Data Quality & Governance - Ensuring accuracy, security, and compliance*



## Tools & Technologies



*Storage: Snowflake, BigQuery, Amazon S3, Delta Lake*

*Processing: Apache Spark, Databricks, dbt, Airflow*

*Pipelines: Kafka, Flink, Fivetran*

*Orchestration: Airflow, Prefect, Dagster*



Amazon S3

## Challenges in Data Engineering

*Data Silos - Breaking barriers between isolated data sources*

*Data Quality - Ensuring clean, consistent data*

*Real-Time Processing - Managing speed & reliability*

*Cost & Complexity - Scaling infrastructure efficiently*



# Data Quality

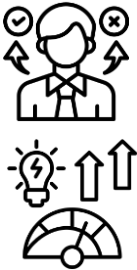
in @ankitrathi



## What is Data Quality?

*Data Quality measures how accurate, reliable, and useful data is for decision-making*

*Goal: Ensure data is fit for use—complete, consistent, and free from errors.  
Like clean water for drinking—bad data leads to bad decisions!*



## Why Does Data Quality Matter?

*Better Decisions - Reliable data leads to accurate insights*

*Fewer Errors - Reduces costly mistakes in business & AI models*

*Compliance & Security - Ensures regulatory compliance (GDPR, HIPAA)*

*Higher Efficiency - Saves time spent fixing bad data*

## 6 Key Dimensions of Data Quality

*Accuracy - Data correctly represents real-world facts*

*Completeness - No missing or incomplete values*

*Consistency - Same data across different systems should match*

*Timeliness - Data is up-to-date and available when needed*

*Validity - Data follows rules & formats (e.g., correct date formats)*

*Uniqueness - No duplicate or redundant records*



## How to Improve Data Quality?

*Data Validation - Check for errors before storing data*

*Deduplication - Remove duplicate records*

*Standardization - Enforce consistent formats and naming conventions*

*Automated Monitoring - Use tools to detect anomalies*

*Data Governance - Clear ownership & accountability for data*

## Challenges in Maintaining Data Quality

*Human Errors - Manual data entry mistakes*

*Data Silos - Inconsistent data across departments*

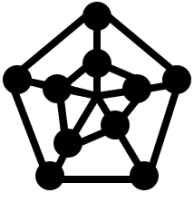
*Outdated Data - Old, irrelevant data reducing accuracy*

*Scaling Issues - Maintaining quality as data volume grows*

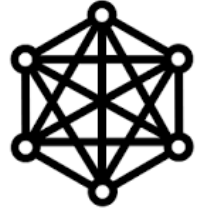


## Understanding

# Data Mesh



 [@ankitrathi](#)



## What is Data Mesh?

*a decentralized approach to data architecture*

*Moves away from centralized data lakes to domain-driven, self-serve data ownership*

*Instead of one giant warehouse, each team has its own organized data store*

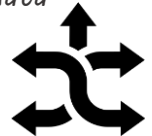
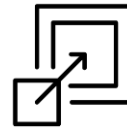
## Why Data Mesh? (Benefits)

*Scalability - No central team bottleneck*

*Faster Insights - Teams access the data they need without delays*

*Ownership & Quality - Teams take responsibility for reliable, high-quality data*

*Flexibility - Works with data lakes, warehouses, and real-time processing*



## Core Principles of Data Mesh

*Domain-Oriented Ownership - Teams own & manage their data as a product*

*Data as a Product - Treat data like a service with defined consumers & quality standards*

*Self-Serve Infra - Empower teams to store, process, & share data independently*

*Federated Governance - Enforce global security, privacy, and standards*

## How Data Mesh Works

*Each business unit (Finance, Marketing, HR, etc.) manages its own data*

*Data is discoverable, shareable, and reusable across teams*

*A common platform ensures security & interoperability without central bottlenecks*

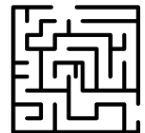


## Challenges of Data Mesh

*Cultural Shift - Teams must take ownership of data*

*Standardization Needed - Common governance rules must be enforced*

*Tech Complexity - Requires the right tools for seamless self-service*







# Understanding

# AI

in @ankitrathi

## What is AI?

*simulation of human intelligence in machines*

*Learning - Adapts from data*

*Reasoning - Makes decisions*

*Self-correction - Improves over time*



## Types of AI

*Narrow AI (Weak AI) → Specialized in one task (Siri, Google Translate)*

*General AI (Strong AI) → Thinks like a human (still theoretical)*

*Super AI → More intelligent than humans (future concept)*

## AI Subfields

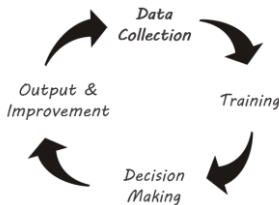
*Machine Learning (ML) - Learns from data (Netflix recommendations)*

*Deep Learning (DL) - AI mimicking the human brain (self-driving cars)*

*Natural Language Processing (NLP) - Understands human language (Chatbots)*

*Computer Vision - Recognizes images (Face recognition)*

## How AI Works



*Data Collection - AI learns from massive datasets*

*Training - Models adjust through experience*

*Decision Making - AI analyzes patterns*

*Output & Improvement - AI refines predictions over time*

## AI in Everyday Life

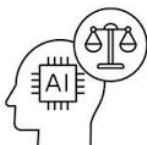
*Voice Assistants (Alexa, Google Assistant)*

*Recommendation Systems (Netflix, YouTube)*

*Healthcare (Disease diagnosis, robotic surgery)*

*Autonomous Vehicles (Self-driving cars)*

*Finance & Security (Fraud detection, stock predictions)*



## AI Challenges & Ethics

*Bias in AI - Unfair outcomes due to biased data*

*Privacy Issues - AI tracking and surveillance concerns*

*Job Automation - AI replacing jobs*

*Ethical AI - Ensuring AI benefits society*



## Understanding

# XAI

in @ankitrathi



### What is Explainable AI (XAI)?

*AI models often behave like black boxes—the 'why' remains missing  
XAI aims to make decisions understandable & interpretable*

### Why Does Explainability Matter?

*Trust - for users to trust AI decisions*

*Fairness - to prevent bias & discrimination in AI models*

*Regulations - to abide by Laws (i.e. GDPR)*

*Debugging - to improve AI performance*

*Safety - in healthcare, finance, autonomous systems*



### How AI Becomes Explainable?

*Feature Importance - data points influencing the decision?*

*Decision Trees - breaking down decision path*

*Local vs. Global Explanations*

*Local: Why was this decision made?*

*Global: How does the model behave in general?*



*SHAP & LIME - Techniques for interpreting black-box AI*

*Model Transparency - Using simpler, more interpretable models*



### Trade-offs: Accuracy vs. Explainability

*Deep Learning Models (Black Box)*

- Highly accurate but hard to interpret
- Used in image recognition, NLP, etc



*Simple Models (Transparent but Less Powerful)*

- Decision trees, linear regression are more interpretable
- Used when explanations are critical (e.g. healthcare, finance)



### Challenges & Future of XAI

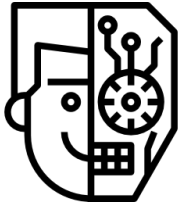
*Trade-off: More explainability can reduce performance*

*Human Interpretation: Even simple explanations can be misunderstood*

*Bias Detection: XAI helps, but bias elimination is tough*

*Future: AI that explains itself in human-like language*





# Understanding GenAI

in @ankitrathi



## What is Generative AI (GenAI)?

*A type of AI that can create new content—text, images, music, code, and more—rather than just analyzing data*

*Like an AI artist, writer, or musician that generates original work based on patterns it has learned.*

## How Generative AI Works?



*Training on Data: AI learns from vast datasets (text, images, code, etc.)*

*Pattern Recognition: Identifies relationships, structures, and styles*

*Content Generation: Uses learned patterns to create new content*

*Refinement & Feedback: Adjusts output based on user input or corrections*

## Popular Generative AI Models

*GPT (Text) - Writes articles, chat responses, and summaries*

*DALL·E (Images) - Creates artwork from text descriptions*

*Codex (Code) - Writes and completes programming code*

*Jukebox (Music) - Generates songs and instrumental music*



## Challenges & Risks of GenAI

*Misinformation - AI can generate fake news & deepfakes*

*Bias & Ethics - AI can reflect biases in its training data*

*Creativity Debate - Is AI-generated content real creativity?*

*Data Privacy - AI models are trained on vast amounts of public data*



## The Future of Generative AI

*More human-like AI assistants*

*Personalized AI-generated content for individuals*

*AI that co-creates with humans in art, music, and writing*

*Ethical guidelines for responsible AI use*

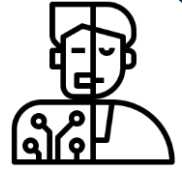




## Understanding



# Agentic AI



in @ankitrathi

## What is Agentic AI?

*AI systems that act autonomously, making decisions, setting goals, and taking actions without constant human intervention*  
*Like a self-driving car that plans its route, adapts to traffic, and makes real-time decisions all by itself*

## Key Features of Agentic AI



*Autonomous Decision-Making - sets its own tasks and goals*

*Planning & Reasoning - doesn't just respond; it strategizes*

*Adaptability & Learning - improves based on feedback*

*Memory & Context Awareness - remembers past interactions*

*Action Execution - takes real-world actions, not just predictions*



## How Agentic AI Works?

*Perception: observes the environment (data, sensors, user input)*

*Decision-Making: determines the best action based on goals*

*Action Execution: performs tasks autonomously*

*Feedback Loop: learns from successes and failures*



## Traditional vs Agentic AI

Aspect	Traditional AI	Agentic AI
Task Execution	Predefined responses	Self-directed decision-making
Adaptability	Limited, follows rules	Learns and adapts
Autonomy	Requires human input	Acts independently
Memory	Short-term	Long-term memory & context



## Challenges & Risks of Agentic AI

*Loss of Control - AI taking actions beyond human oversight*

*Ethical Concerns - Who is responsible for AI decisions?*

*Unintended Consequences - AI optimizing for unintended goals*

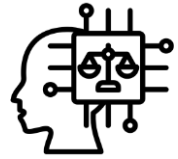
*Safety & Security - Preventing rogue AI behaviour*



## Understanding

# AI Ethics

 @ankitrathi



## What is AI Ethics?

*Study of moral principles that guide the development and use of AI ensuring it is fair, safe, and accountable while respecting human right*

*AI is like a powerful car; without ethical “rules of the road,” it can cause harm*



## Why Does AI Ethics Matter?

*Trust - People must trust AI to use it safely*

*Bias & Fairness - Prevent discrimination in AI decisions*

*Privacy - Protect personal data from misuse*

*Accountability - Who is responsible when AI makes mistakes?*

*Safety & Security - AI should not cause harm or be misused*

## Examples of Ethical AI Challenges

*Hiring Bias - AI in job screening favouring certain groups unfairly*

*Deepfakes - AI-generated fake videos spreading misinformation*

*Facial Recognition - Privacy concerns in surveillance and law enforcement*

*AI in Warfare - Autonomous weapons making life-and-death decisions*



## Solutions for Ethical AI

*Fair AI Training - Diverse, unbiased training datasets*

*Explainable AI (XAI) - Making AI decisions understandable*

*Regulations & Guidelines - Laws ensuring ethical AI use (like GDPR, AI Act)*

*Human Oversight - AI should assist, not replace, human decision-making*

*AI for Good - Using AI in healthcare, climate change, and education*



## The Future of AI Ethics

*Stronger AI regulations worldwide*

*More transparency in AI systems*

*AI designed for social good and fairness*

*Better AI-human collaboration with ethical safeguards*



The

# AI Productivity

in @ankitrathi

Paradox

## The Promise vs. The Reality

### What AI Vendors Claim:

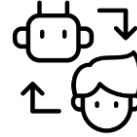
"AI can make work 10x or 100x faster!"

"A task that took 100 days will now take 1!"

"AI will replace entire teams!"

10

100

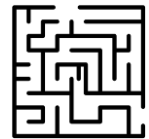


### The Reality:

AI speeds up tasks, but doesn't eliminate human oversight

Quality, debugging, and integration still take time

More automation = more complexity, not always more efficiency



## AI's Hidden Cost: Technical Debt

### AI-Generated Code = Piling Up Problems

Messy & redundant code

Security & compliance risks

Hard to debug & maintain



More automation now → Bigger maintenance headaches later

## Why Executives Fall for AI Hype

### Why do non-tech leaders buy into exaggerated claims?

FOMO - They don't want to be left behind

AI Magic Effect - Demos look impressive

Marketing Spin - Vendors oversell AI's capabilities



### Missing Piece: Understanding AI's Limitations!

## The Need for Tech-Savvy Leadership

### Smart leaders ask the right questions:

What's the real efficiency gain?

How much human oversight is still needed?

What's the long-term cost of AI adoption?



## AI is a Tool, Not a Magic Wand

AI can boost productivity, but it's not a miracle

Used wisely, it's a great assistant

Used blindly, it creates more problems than it solves

Think of AI as a power tool - It's useful,  
but you still need a skilled worker!



The

# Agentic Pipeline

in @ankitrathi

Problem

## Data Pipelines vs. Agentic Pipelines

✂ *Data Pipelines* → Structured, deterministic, and human-supervised

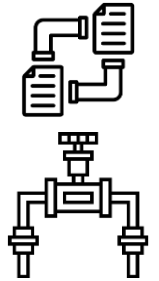
🤖 *Agentic Pipelines* → Autonomous, probabilistic, and harder to debug

### What's Common?

Both rely on multiple hand-offs

Both struggle with data quality & governance

Both suffer when complexity increases



## The Four Big Problems in Agentic Pipelines

### Too Many Complex Handoffs

Agents pass data to other agents without clear oversight

Each step adds uncertainty & potential errors

### Transformations Without Transparency

No clear visibility into what each agent is doing

Difficult to track errors or debug failures

### No Visibility Into Downstream Use

Who uses the data? How is it consumed?

Without human oversight, errors go unnoticed until it's too late

### Ripple Effects - One Error = System-Wide Chaos

A single issue can cascade across all dependent agents

Errors multiply, making debugging a nightmare

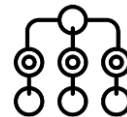
## The Solution: AI Governance & Contracts

### Define clear AI contracts for:

Data inputs & expected format

Prompts & model constraints

Expected outputs & downstream dependencies



*Without guardrails, agentic pipelines will spiral out of control!*

### Final Thought:

*Agentic Pipelines = Data Pipeline Problems, But Worse  
If we don't solve governance now, trust in AI-driven  
systems will collapse!*