

Convergence of AI, data & Analytics technologies:

The impact of GenAI on Data & Analytics

Einat Shimoni



The Data Analytics Value Chain

How can AI help?



Measure
Impact

Extract
Insights

Redesign
processes /
decide

Get the data

Make it
accessible and
trusted



01

Get the data



What are the solutions?

01

Get the data

35% of DS time is spent on getting data needed.

“Either you have the data or you don’t” state of mind.

Many high impact projects are defined “unfeasible”

What are the solutions?

01

Get the data

External Data Marketplaces:
AWS Data Exchange,
Snowflake Data Marketplace,
Dawex ...

Solutions:

Option 1: "Buy it"

SNOWFLAKE DATA MARKETPLACE



01

Get the data

Solutions:

Option 2: “Make it”

Synthetic data is artificially generated data that simulates real data. Rooted in the Autonomous Driving industry, It's used as a substitute or supplement when real data is difficult, expensive, or constrained by privacy issues. Result: Labeled (!), categorized, and formatted training datasets

Players: Nvidia, Tonic, Datagen, Mdclone, Hazy...



What's real and what's fake?

Ai vs Real Headshot



Is this image real or created with Ai?

☐ Real Photography

☐ Ai Created



Synthetic data powered by GenAI "feels" pixel-perfect real

Copyright@STKI_2024 Do not remove source or attribution from any slide, graph or portion of graph

01

Get the data

Synthetic data use case examples:



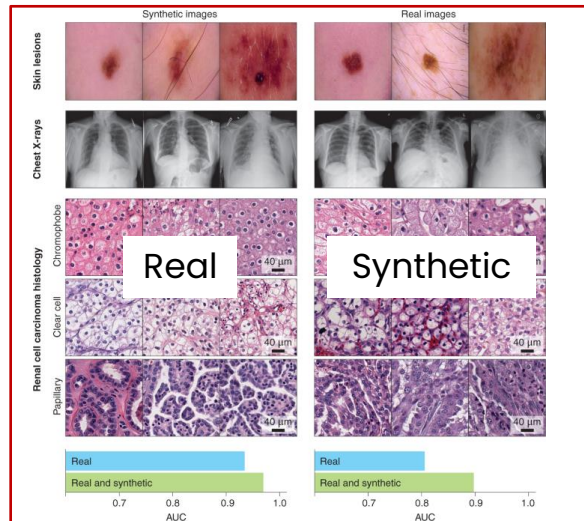
Simulating long-tail driving scenarios to train **AV models**

Additional use cases:

Financial Services: sharing financial data without compromising privacy

Manufacturing: simulate supply chain operations

Construction: simulate safety scenarios



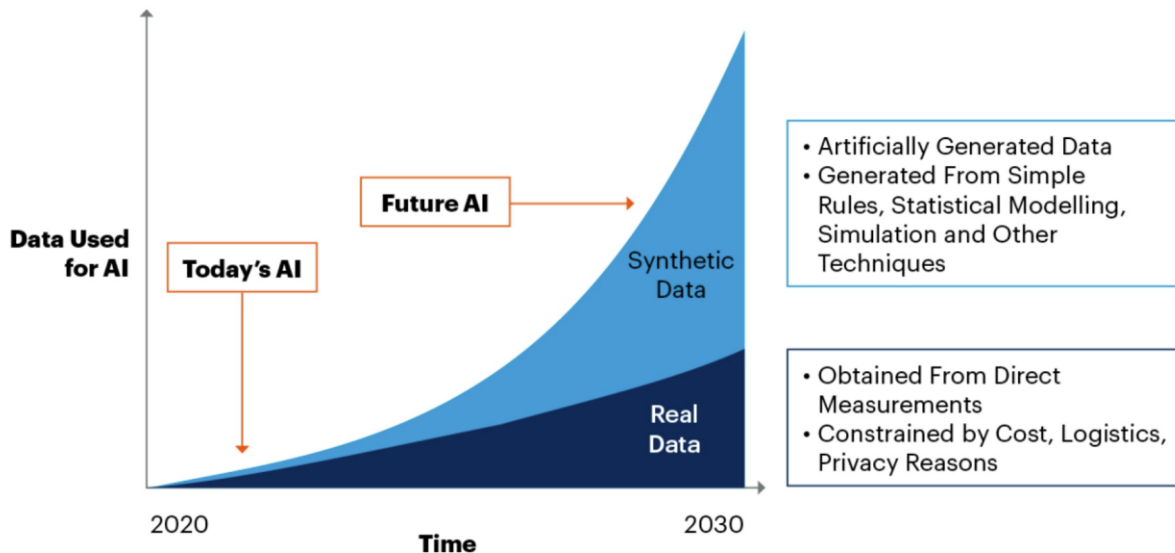
Creating **genomic** and **health** data to improve patient care and research



01

Get the data

By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models



Source: Gartner
750175_C

Gartner



STKI.INFO

Copyright@STKI_2024 Do not remove source or attribution from any slide, graph or portion of graph

01

Get the data

Who is the data curator?

Collects, organizes, manages, and preserves valuable data:

Managing, cleaning, enriching, and cataloging data to make it more usable and valuable for analysis

Do we really need additional roles in data mng?!?

Data Governance mngr

Data Trustee

Data Owner

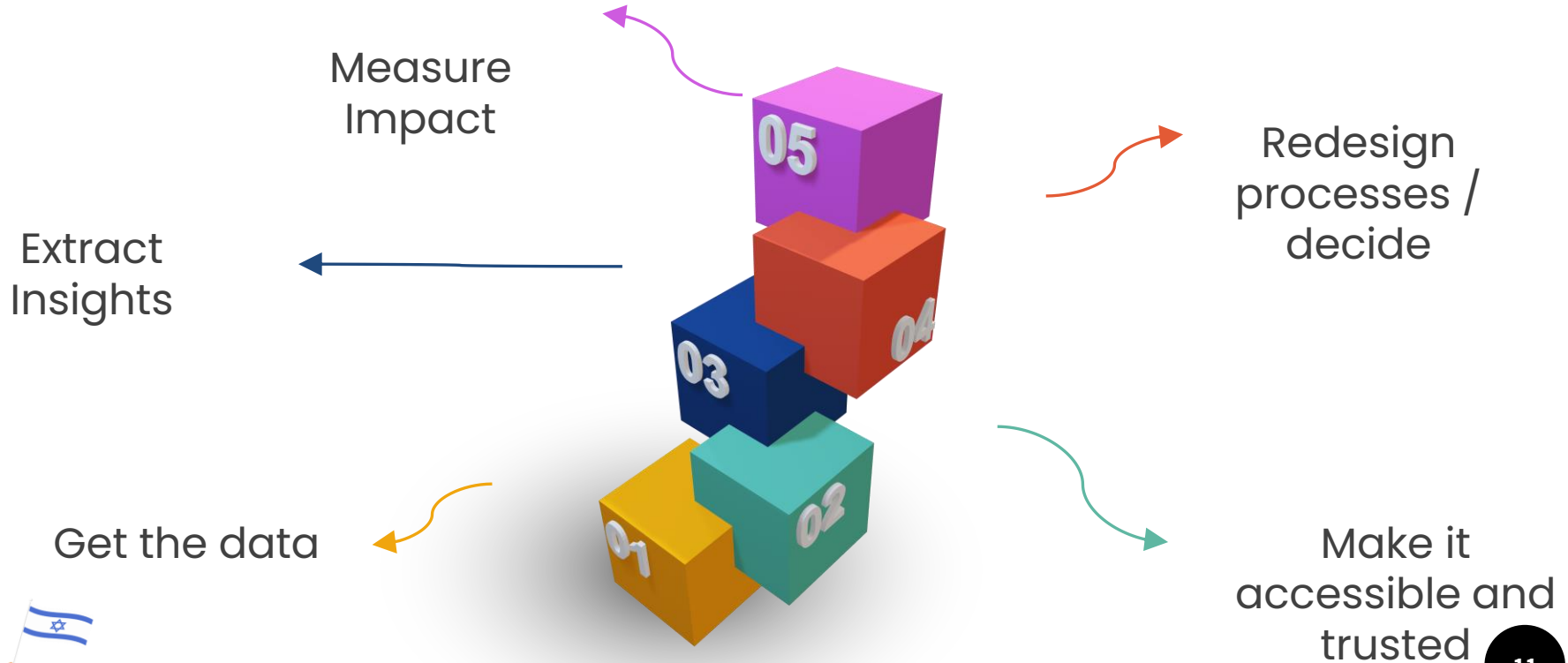
Data Steward

Data Curator

Data Custodian



The Data Analytics Value Chain



02

Make it accessible and Trusted

Cleaning, tagging and managing the data is the most time and effort consuming task.

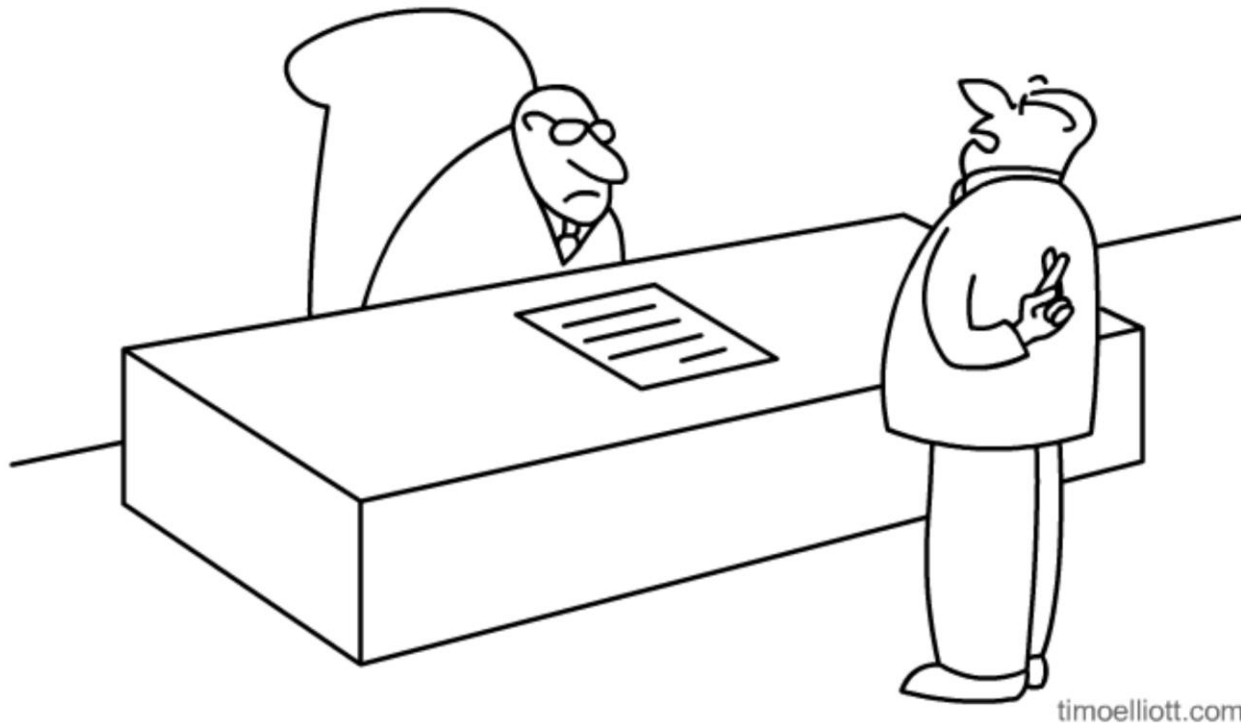
Is it trustworthy?

Is it discoverable?

Is it safe?

Is it not fair and not biased?

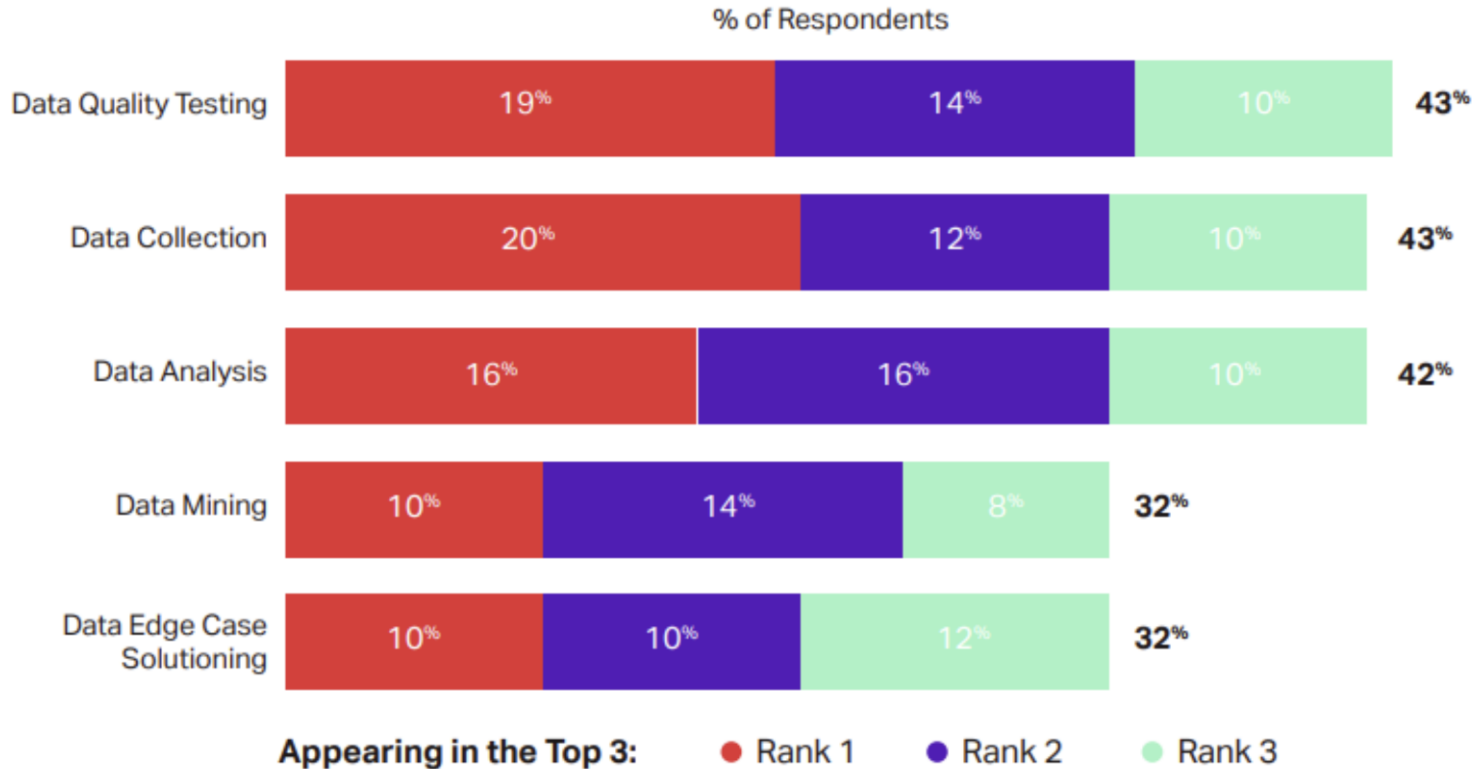




"Yes sir, you can absolutely trust those numbers"

Source: Tim Elliott

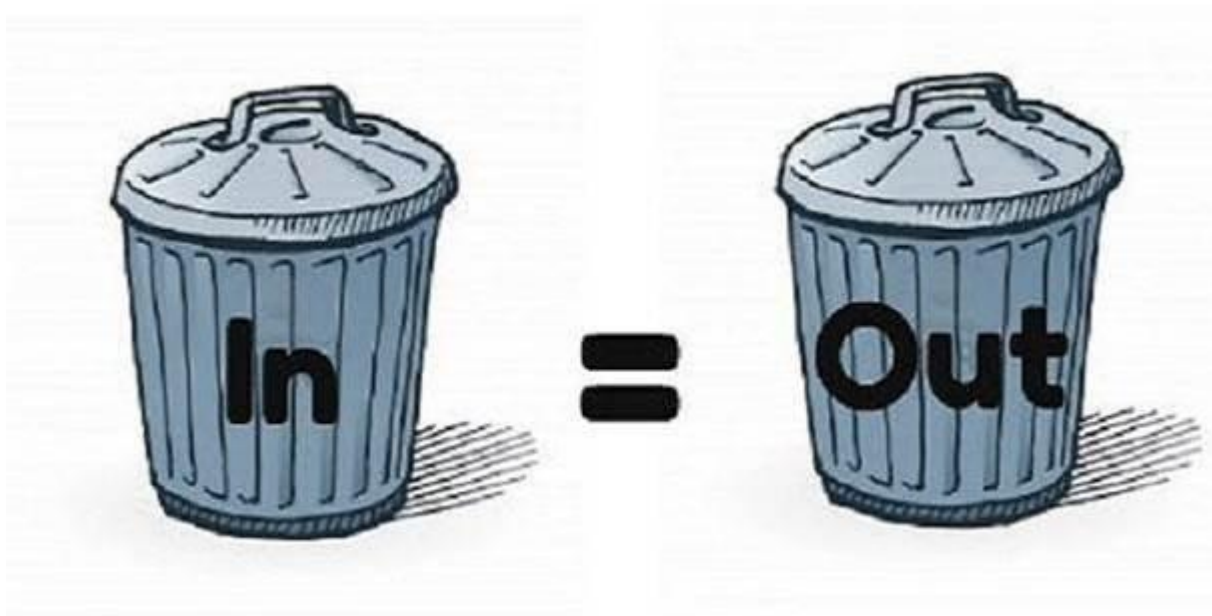
Data Quality remains the top challenge for AI



Source: iMerit and VentureNeat – State of MLOps report 2023

Copyright@STKI_2024 Do not remove source or attribution from any slide, graph or portion of graph

Is your data ready for GenAI?



Data management Is stuck in the 90's



02

**Make it
accessible
& trusted**

Solutions:

- Cloud Data Platforms
- Data Products approach
- Use of AI for manual, time-consuming tasks in data management

"It's not about replacing the human with the robot, it's about taking the robot out of the human."

- Usama Fayyad, Institute of Experiential AI

Copyright@STKI_2024 Do not remove source or attribution from any slide, graph or portion of graph



02

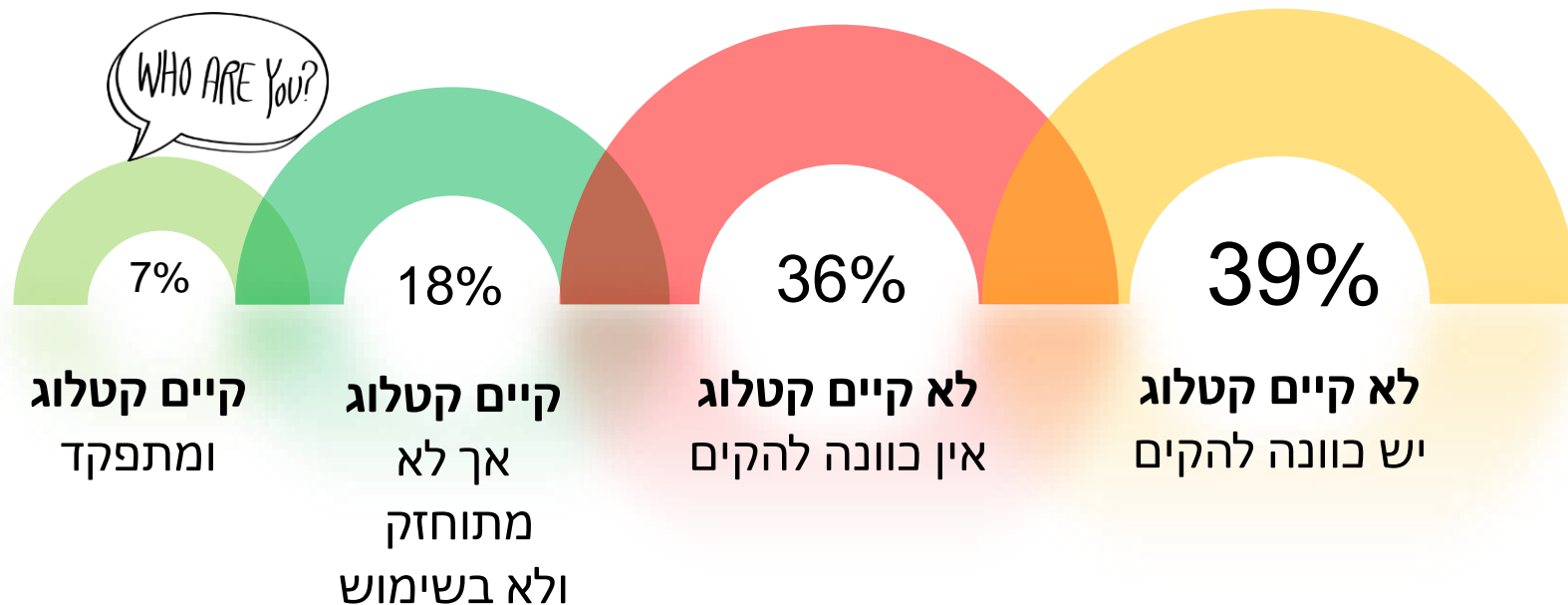
**Make it
accessible
& Trusted**

How can AI help?

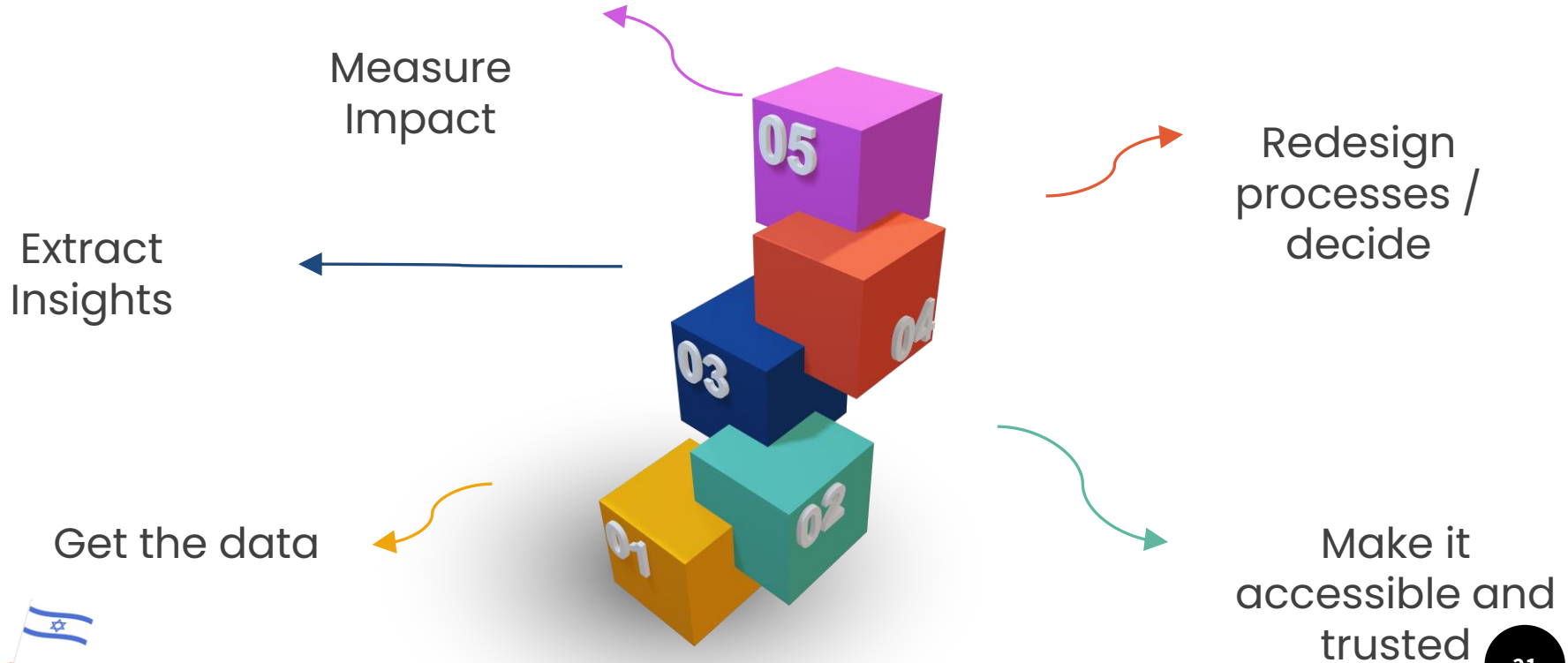
- Data profiling: identify data anomalies, biases, errors, duplicates
- Data cleaning
- Auto-metadata generation based on content & context
- Taxonomy building and knowledge graphs



Reminder: almost ALL data catalogues projects in Israel **FAILED**



The Data Analytics Value Chain



03

Extract Insights

Currently, 30%–50% of employees use BI and discovery tools.

How can organizations empower people to make data-driven decisions?



03

What if...

- 🔍 You could just “talk” to the data?
- 🔍 You could have your own 24/7 “assistant”
- 🔍 Knows you and your preferences?
- 🔍 Explore the data in free language?
- 🔍 Doesn't mind “stupid questions”?



03

Extract Insights

Augmented Analytics



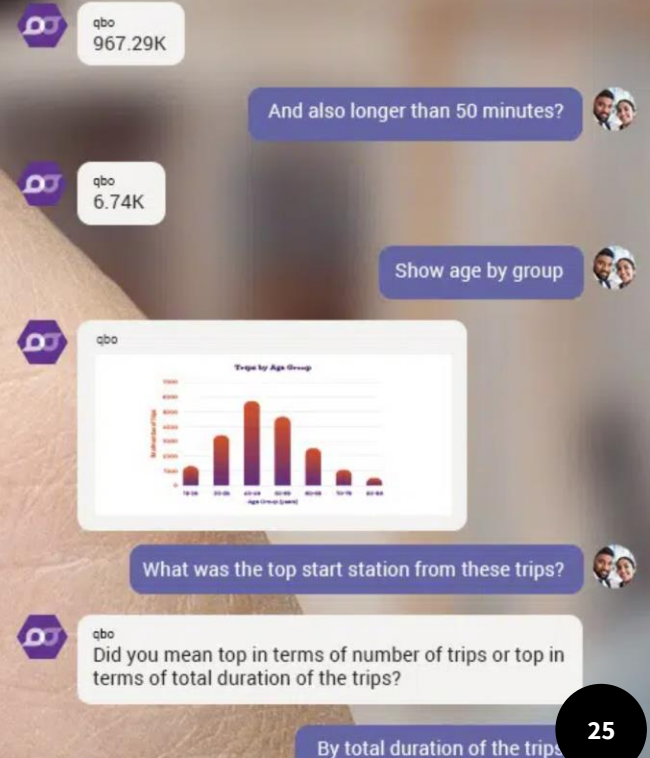
Augmented analytics is the use of enabling technologies such as ML & AI to assist with data preparation, insight generation and explanation to augment how people explore and analyze data (Gartner)



Conversational analytics

Source: Unscramble

- "What does this data talk about"?
- "What are the top 10 insights"?
- Explain them to me in a non-technical way
- Tell me a story (narrative) that explains it
- Use examples from my organization's world
- Show them in a graph
- Change the graph

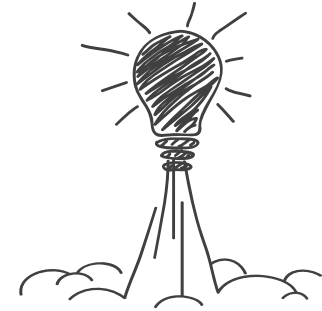




Passive assistant
not personalized



Passive assistant
personalized



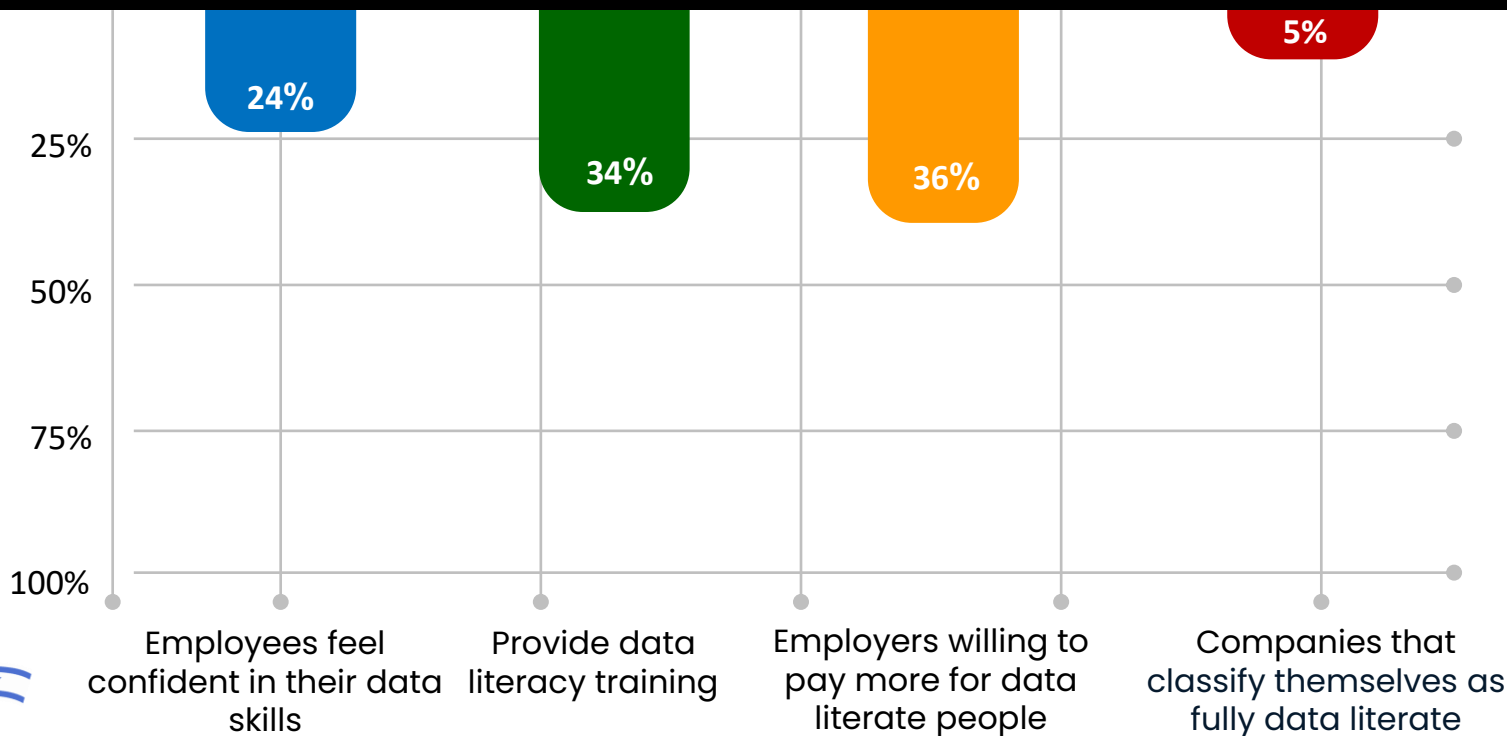
Proactive assistant
Personal "concierge"

Conversational Analytics phases

Copyright@STKI_2024 Do not remove source or attribution from any slide, graph or portion of graph

Wait, does this mean **data literacy** is no longer an “issue”?

Source: Qlik



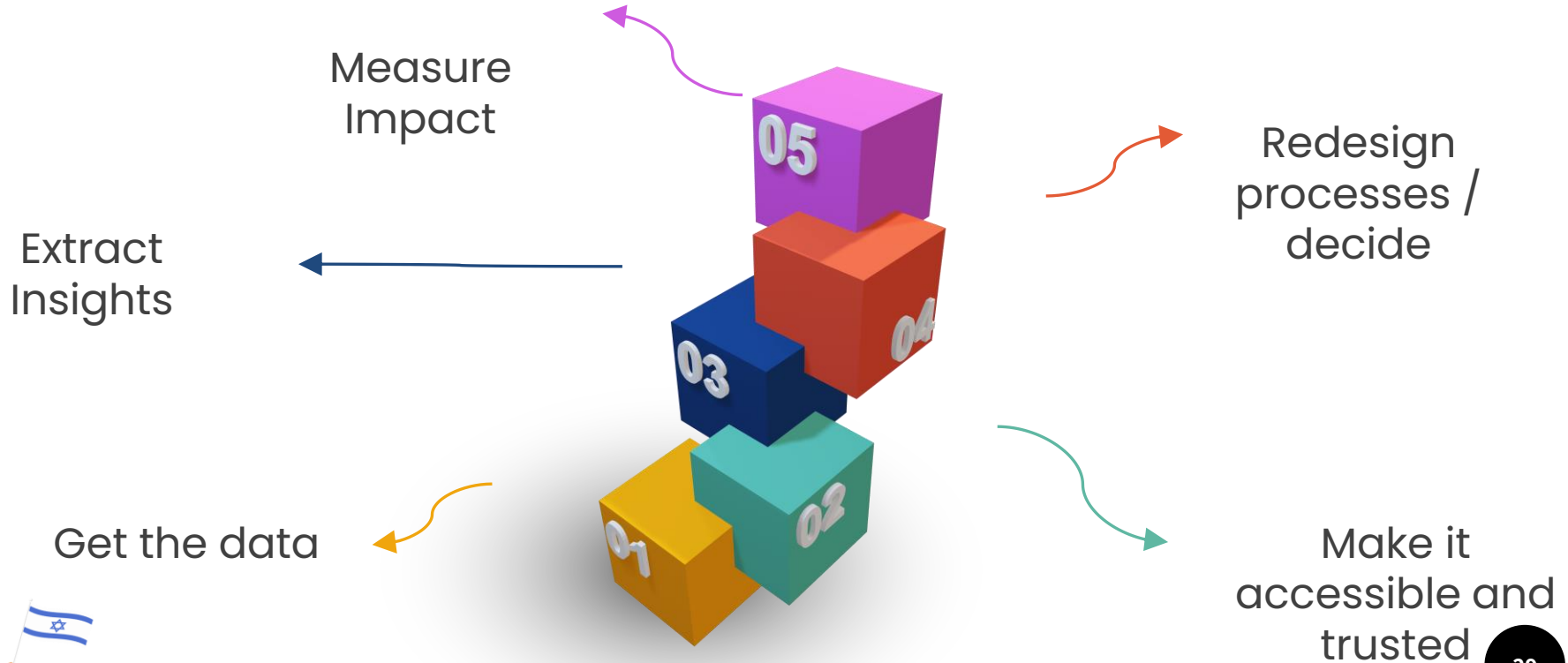
Will AI literacy become the new data literacy?

50% of employees
believe AI skills will
be essential for
their role

Only 13% were
offered AI training
in the last year



The Data Analytics Value Chain



04

Redesign Processes

Analytics are disruptive to existing processes

Resistance to change is the #1 roadblock for analytics



04

Redesign Processes

- Change management
- Product approach that looks at the WHOLE experience



LAST MILE



05

Measure Impact

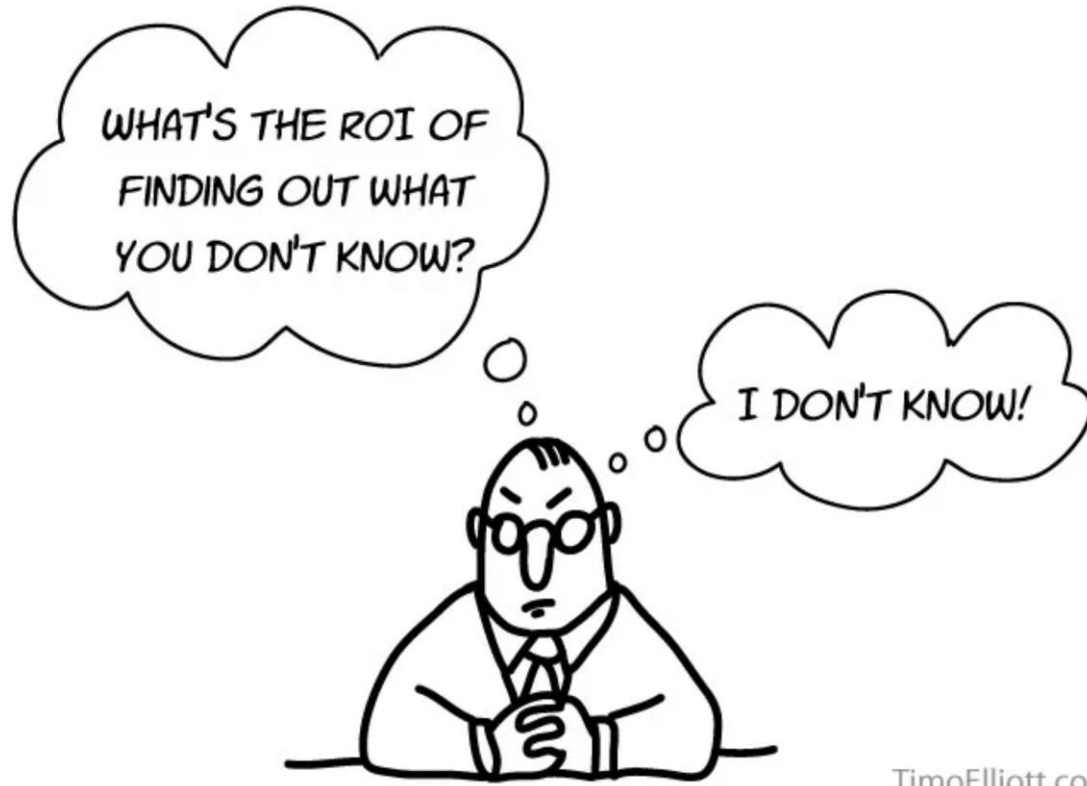
What is the (business) impact?

How is it measured?

How is it communicated to the organization?

Is there a continuous process of discovery and value measurement?





Source: Tim Elliott

Copyright@STKI_2024 Do not remove source or attribution from any slide, graph or portion of graph

“Truths” in the D&A world:



“You either have the data or you don’t”



“Data management is a manual, time-consuming but essential effort ”



“Data literacy is a major obstacle for data transformation”



“Data is the new oil”, the data you own is your differentiator



“There is no substitute for having a clear business case, trustworthy data, and an ongoing process of discovery”