

# GenAI for Data Quality Management

# Data and AI Team



**Lukas Dech**

Associate

+49151 17517208

Lukas.dech@pwc.com



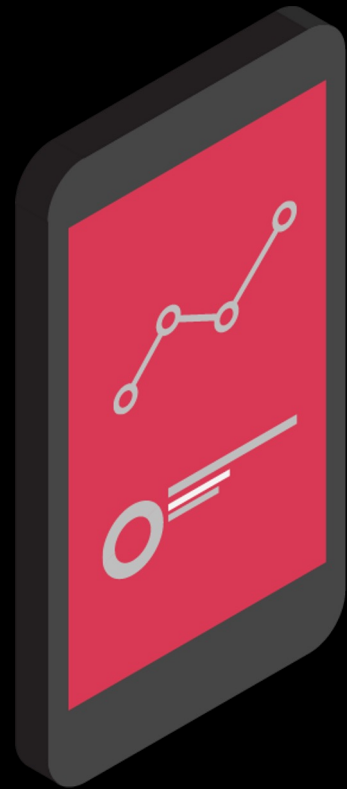
**Philipp Plöhn**

Manager

+49151 29607070

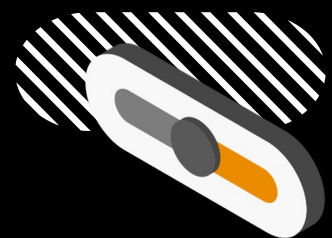
Philipp.ploehn@pwc.com





Imagine receiving one of the first iPhones in 2007...

...but without a charger and you are not able to use it...



So how will you be

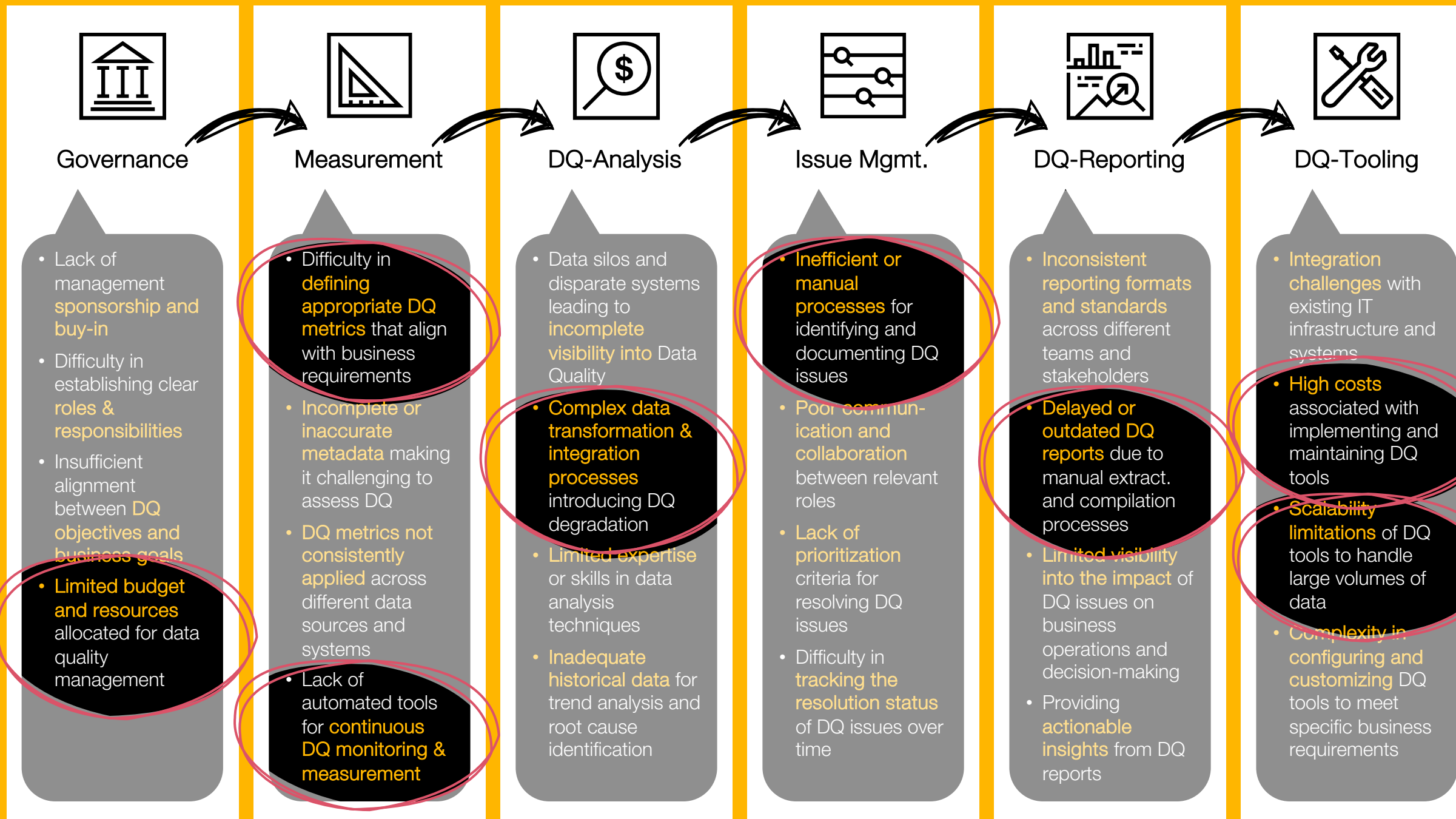
using new technologies & AI

without good Data Quality?



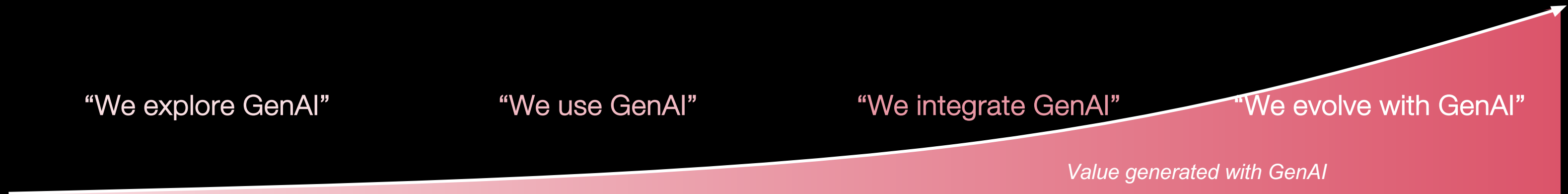
# Data Quality Management poses several challenges

## across all relevant capabilities and processes



- Traditional Data Quality Management is rule-based – which makes it time-consuming, expensive, incomplete and non-scalable
- Commonly, it is the aim of Data Quality Management to comply with regulatory requirements rather than adding value to the company's data
- Data Quality is the cornerstone of any new technology and the usage of Artificial Intelligence – without Data Quality none of it can be used

# The market overview on GenAI maturity shows a progressive approach towards usage



You should ask yourself ...

Where to start?

Developing an initial feeling for GenAI and its possibilities

How to plan for scale?

Setting up the organization and getting everyone on board

How to be compliant, secure and efficient?

Growing your GenAI capabilities with "Compliance by Design"

How to evolve to differentiate?

Staying ahead of the market through adaption and reinvention

How PwC can support you

Selection of a pilot use case that can be implemented in a project as a PoC

GenAI readiness check

Developing a strategy for scaling the current GenAI efforts including use case management

Enabling employees through fostering a bottom up innovation culture as well as upskilling

Integration of GenAI into the whole organization through "Compliance by Design" affecting processes & organization, governance, platform, data management, culture at scale, and use cases

Assessing the impact of GenAI on the business model and the business strategy and the possibilities for business model reinvention and new business opportunities

# Generative AI can help solve numerous Data Quality Use Cases



## Data Culture

Complex concepts around data culture and governance are explained in a simple yet insightful manner that promotes the data skills of employees and fosters the awareness of potentials and challenges.

## DQ Rule Generator

DQ rules are derived by the generative AI from exemplary data or meta data with technical and business descriptions. The relationships between data points are used to identify required consistencies and dependencies. The identified logics can enhance or validate the existing DQ framework or create a completely new one.



## Data Migrator

Similar to a human analyst, the Data Migrator uses generative AI to learn from source and target data structures, descriptions and examples to create mappings for data migration.



## DQ Masterbrain

Bring data quality to life with the AI Data Steward! The AI, as a 24-hour contact for data-related topics, serves as a one stop shop for everything related to data quality.

## DQ Cleaner

Artificial improvement of data quality. The DQ Cleaner detects errors and anomalies and makes correction suggestions that can be adopted either automatically or with the help of experts.



## DQ Reporting

DQ KPIs are measured by the generative AI to monitor data quality in dashboards and establish transparency over trends and gaps. The status of DQ issues is tracked over time to check for improvements.



## DQ Prevention

The business users interact with the generative AI to talk about specific DQ issues and are presented with actions to prevent such issues: issues can be detected earlier, the root cause can be tracked down or some issues can be avoided completely.



# Let us demonstrate how the PwC Data Quality Rule Generator works and excels

## Business Context and Description

50% Productivity increase

GenAI's context processing delivers fast business insights, significantly cutting down analysis time for enhanced productivity

## Specification of data quality rules

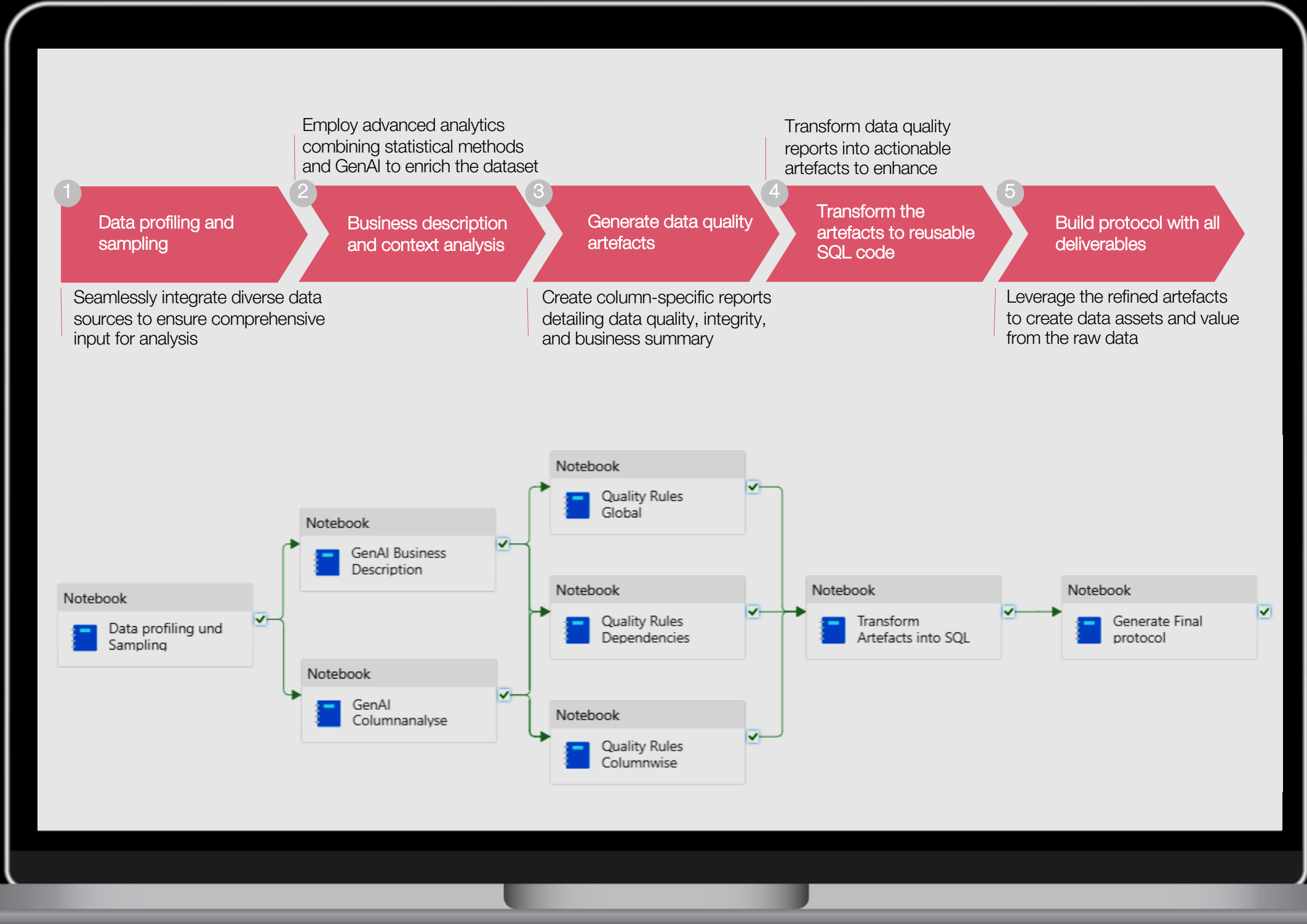
90% Productivity increase

With GenAI, data quality rules are generated with greater speed and accuracy, streamlining workflows and boosting analyst productivity

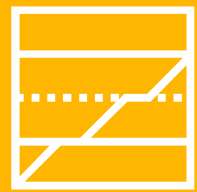
## Implementation of data quality rules

75% Productivity increase

GenAI's quick rule conversion and implementation save valuable time, enhancing data management efficiency and overall productivity



# The PwC Data Quality Rule Generator enables automated quality assessments and data quality gates for data integrity



## Requirements for success

**Data Accessibility:** Access to comprehensive and representative datasets across domains and data types.

**GenAI Training:** Access to state-of-the-art Generative AI models capable of learning complex data patterns and relationships.

**Domain Expertise:** Collaboration between data scientists, domain experts, and data stewards to define relevant data quality dimensions, such as accuracy, completeness, consistency, and timeliness

**Rule Interpretability and Transparency:** AI-generated rules should be interpretable and explainable to facilitate understanding and acceptance by stakeholders.

**Performance Evaluation and Monitoring:** Feedback loop mechanisms to incorporate user feedback, domain knowledge, and changing business requirements into the GenAI training and rule generation process.



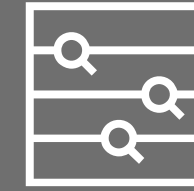
## Our insights & know-how added

**Data Science & ML Expertise:** Proficiency in machine learning techniques, particularly in Large Language Models, and other AI algorithms.

**Data Quality Management Knowledge:** Deep understanding of data quality dimensions, standards, and best practices in data governance, data stewardship, and data quality assurance.

**Domain Expertise:** Domain-specific knowledge in the industry or functional area where the project is implemented, such as finance, healthcare, retail, etc.

**Regulatory Knowhow:** Navigating existing regulations & standards that are supplemented by new focus topics around data privacy and the upcoming EU AI Act



## Deliverables and results

**Automated Business Insights:** Use descriptive statistics and domain knowledge, to craft automated business descriptions and contextual insights. Fill data catalog and advances data literacy within the organization.

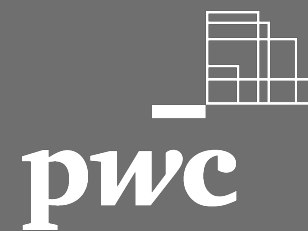
**Scalable Quality Assessments:** Utilizes generative AI for automated, thorough data quality assessments. Identify and resolves data issues, improving data quality throughout the enterprise.

**Data Quality Protocols & Recommendations:** Offer proposals for dataset enhancements and to create detailed protocols for each dataset in your data warehouse, promoting ongoing data quality improvements.



# We will provide best practices & know-how to further develop GenAI DQ Use Cases in **Co-Creation**

Joint Investment. Joint Results. Joint Success.



- Facilitation of **collaborative workshops** to define a shared vision and establish clear project goals
- Assembly of **cross-functional teams** comprising experts from various domains
- Implementation of an **iterative process** of ideation, prototyping, testing, and refinement
- Establishment of **communication channels** and facilitation of regular collaboration sessions
- Ensure close collaboration to identify **opportunities for value co-creation** throughout the project



- Active **engagement in workshops**, providing insights and aligning on objectives
- **Nomination of key stakeholders** to join the project team, ensuring representation from relevant departments
- **Provision of feedback** and participation in testing and validation exercises
- **Engagement in open communication**, sharing insights, challenges, and expectations transparently
- Collaboration with PwC to **prioritize features and functionalities** that deliver maximum value to stakeholders