



From Models to Data: How Gen AI is Changing the Game for Data Scientists and Data Teams

Presented by: William McKnight

“#1 Global Influencer in Big Data” Thinkers360

President, McKnight Consulting Group

3 X **Inc 5000**

 /in/wmcknight

www.mcknightcg.com
(214) 514-1444





erwin[®]
by Quest

The Modern Data Architectures

Common Problems You May Be Facing



COST



CULTURE



PROCESS



PERFORMANCE



ROI

where are you and where would you rather be?

Using the data
that's available



Small team of
report builders



Executive dashboards and
reports that inform the
organization

Getting data
from many
sources



Knowing that it's
trusted and
available



Having more smart people
help make decisions at
every level

Bad Data = Bad AI

AI Governance

Structured Data → Modern Data Stack

Is your **data** ready for **AI**?

Unleash the full potential of your data.

[Contact Sales](#)



Set up your **AI** initiative for **success**

Avoid costly failures in AI implementation by ensuring your data foundation is strong. From data preparation to governance and quality, maximize your AI investment and achieve positive results with reliable, AI-ready data.

 **\$1T**
In AI spending by 2033

 **85%**
Formalizing AI governance

 **59%**
Don't measure data quality

 **74%**
Struggle with AI adoption

Pillars of **AI** excellence

Enter the AI era with momentum, accuracy and efficiency with Quest solutions.



Blueprints for AI-ready data

De-risk your AI journey and lay the foundation for AI success with data modeling, cataloging, governance, lineage and quality management. With Quest solutions, you can build structured, scalable and trusted data ecosystems that ensure compliance, reliability and seamless AI integration.

[Contact Sales](#)



Maximize the value of your data platform

You can maximize your modern data platform investment with tools that streamline migrations, optimize performance, enable cross-platform governance and speed up data product delivery designed to unlock quick wins and long-term value.

[Contact Sales](#)



Boost AI success with enhanced data quality

Don't allow inconsistent, incomplete or incorrect data to undermine your AI outcomes. Address data quality issues with automated profiling, cleansing and enrichment to guarantee reliable insights and actionable results for AI initiatives.

[Contact Sales](#)

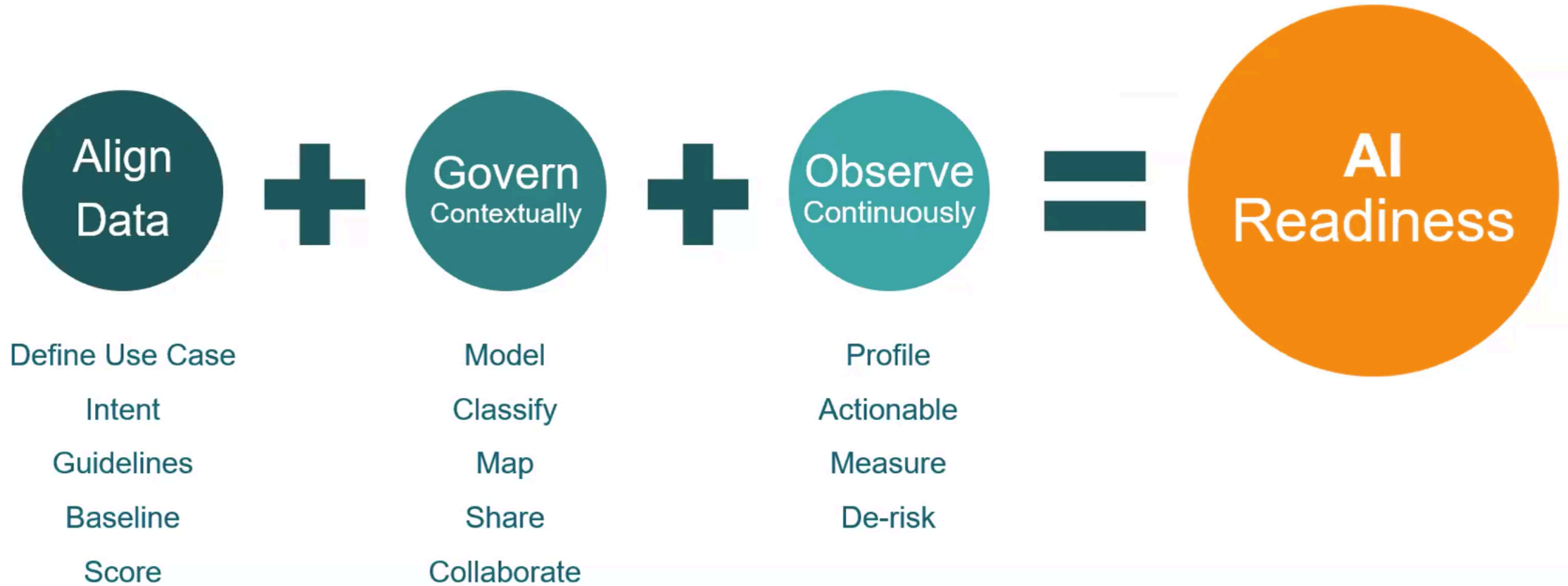


Conquer AI model governance

Effective governance ensures your AI models are compliant, traceable and trustworthy. Implement frameworks that monitor AI usage, align with regulatory requirements and maintain the ethical deployment of machine learning.

[Contact Sales](#)

creating responsible, reliable and smart AI



Model to Marketplace

a framework to help you succeed



Model to Marketplace

a framework to help you succeed



AI model governance

govern input | certify | monitor output

Input Data (erwin by Quest 7 steps)

Model

The blueprint

Curate

Business context
(rules, policies,
definitions)

Observe

data lineage,
data profiling +
quality

Shop

centralized,
governed
availability for
use

Catalog

Identify where
the data resides

Govern

ownership, data
classification +
management

Score

popularity,
reliability,
governance
completeness



erwin Certified

AI models and
data certified as
fit for use

Output Model Observation

- Data Quality – profiling + business rules
- Data Drifts – alerts as data changes drastically
- Sensitive Data – alerts when sensitive data leaks

output – AI model observation

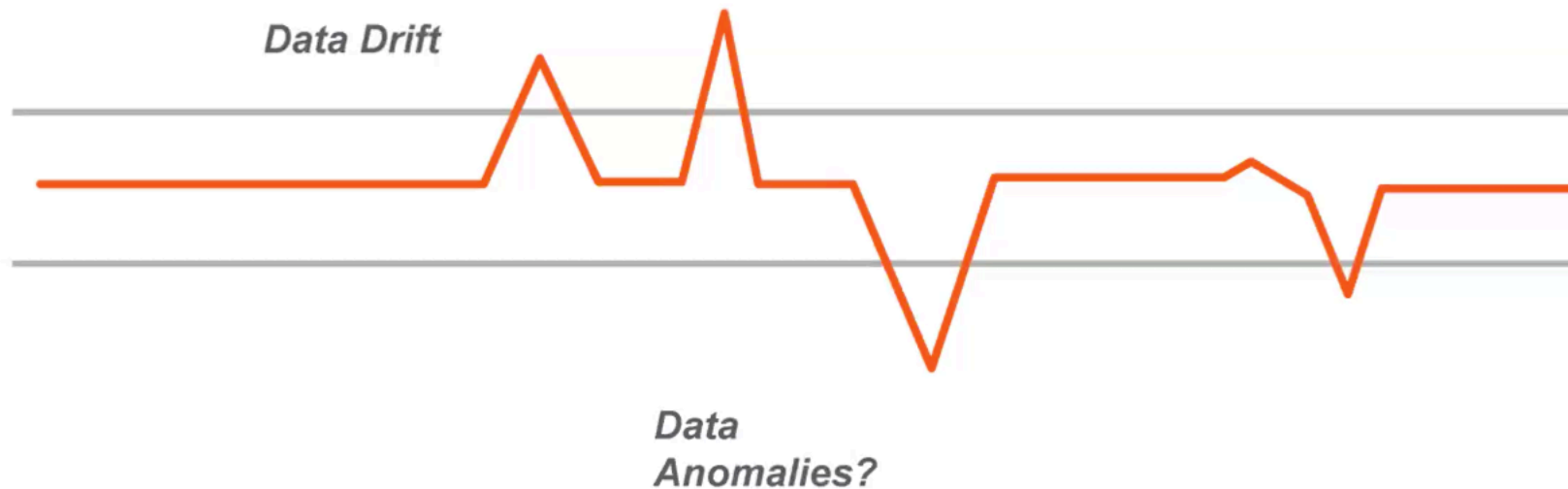
alert | react and tune | collaborate | remediate

Observe Data



Automated data observability alerting via Slack, Teams and email

Reveal Insights



where are you and where would you rather be?

Using the data
that's available



Small team of
report builders



Executive dashboards and
reports that inform the
organization

**Getting data
from many
sources**



**Knowing that it's
trusted and
available**



**Having more smart people
help make decisions at
every level**

where can you start?

1

Adopt best practices for building and submitting data products (clear metadata, governance, quality scores).

2

Collaborate across teams: Share knowledge between IT, data engineers, data scientists, and business users.

3

Start small: Begin with a few high-impact data products, refine the process, and expand.



Download our guide
to further explain our
7-step methodology
for Data and AI
Governance



From Models to Data: How Gen AI is Changing the Game for Data Scientists and Data Teams

Presented by: William McKnight

“#1 Global Influencer in Big Data” Thinkers360

President, McKnight Consulting Group

3 X **Inc 5000**

 /in/wmcknight

www.mcknightcg.com
(214) 514-1444



Partial Client List

CONSUMER PRODUCTS/RETAIL



FINANCIAL



INSURANCE/HEALTHCARE



PUBLISHING



OTHER



GOVERNMENT AND UTILITIES



EDUCATION



PHARMACEUTICAL



TELECOMMUNICATIONS



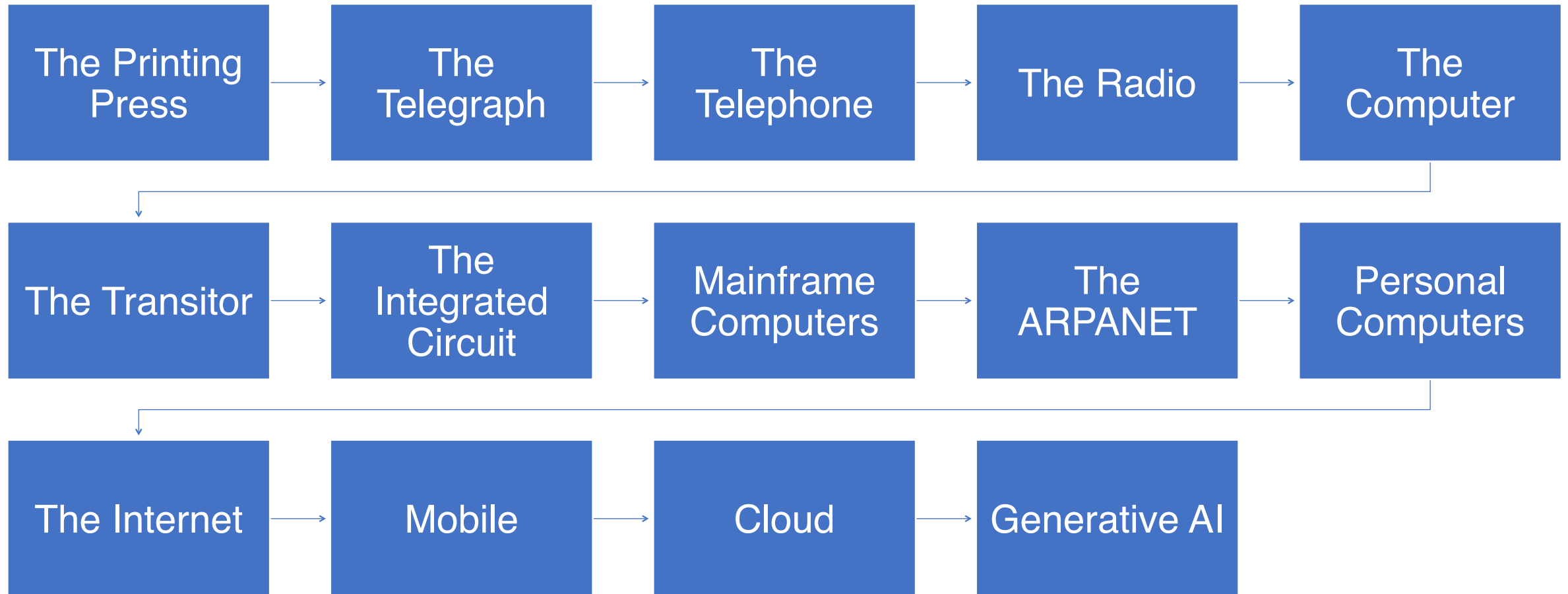
Dataversity Analytics Architecture with William McKnight 2025

1. 2025 trends in Analytic Architecture
2. Standardizing Data Collaboration: The Role of Open Table Formats in Data Architecture
3. From Models to Data: How Gen AI is Changing the Game for Data Scientists and Data Teams
4. The Rise of Automated MDM: How AI and Machine Learning are Revolutionizing Master Data Management
5. The Data Observability Advantage: Unlocking the Secrets to Reliable, High-Quality Big Data
6. Promising AI Use Cases for the Enterprise in 2025
7. Edge Computing Evolved: Introducing the Zero-DBA, Zero-ETL Embedded Database
8. The ROI of AI is (usually) there – Let's Run the Numbers
9. Translytical Databases: A Framework for Evaluation and Use Case Analysis
10. From Pre-Trained to Fine-Tuned: How to Get the Most Out of Vector, RAG, and Small Language Models
11. Designing a Data Platform for the Future: Principles, Patterns, and Best Practices for Data Fabrics and Data Meshes
12. AIOps: Applying DevOps to Competitive Advantage

Why is GenAI Changing the Game

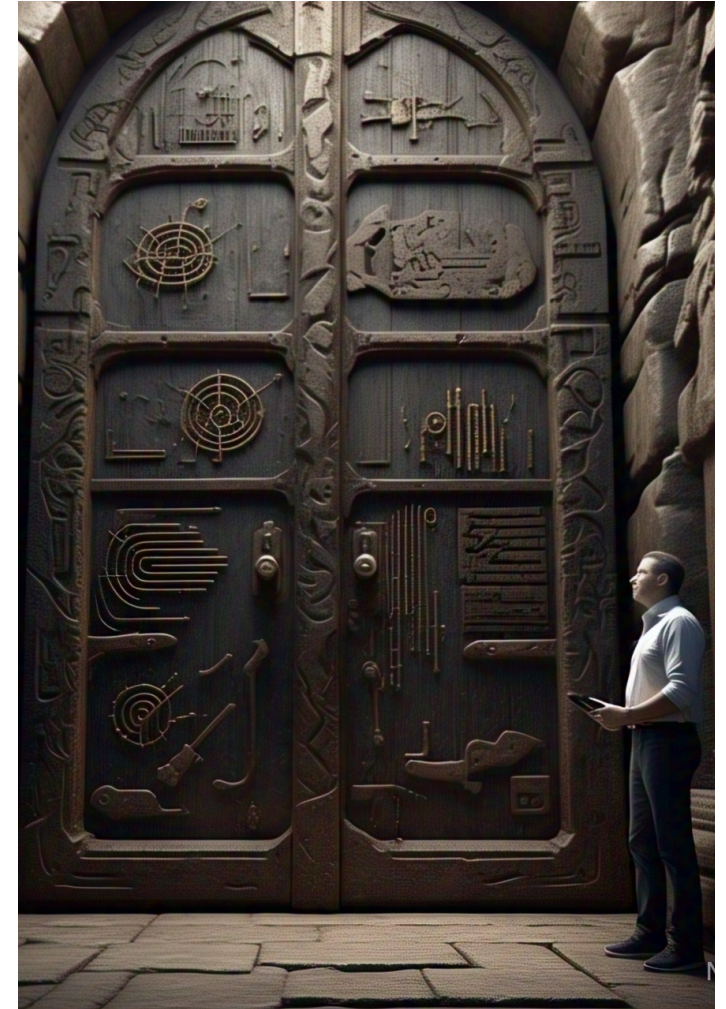
AMERICA · AFRICA
ASIA · AUSTRALASIA

The Big Technology Waves 1440-Present



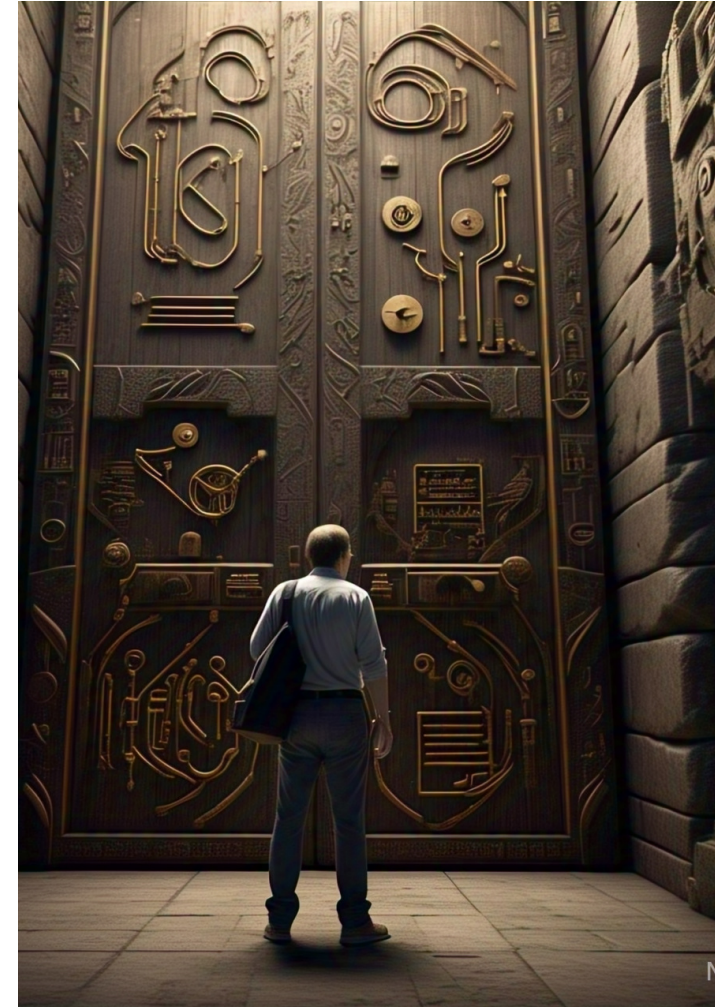
GenAI Simplifies Business

- User needs answer to business questions
- Project or request?
- They get a report, a dashboard or use a tool to interact with data
- Frequently the effort is more than the value, or the effort is misplaced
- GenAI can simplify the use of analytics



How LLMs are Trained

1. **Data Collection:** Gathering a massive dataset of text from various sources (e.g., books, articles, websites).
2. **Preprocessing:** Cleaning and formatting the data (e.g., tokenization, stopword removal).
3. **Model Architecture:** Designing the LLM's architecture (e.g., transformer, recurrent neural network).
4. **Training:** Feeding the preprocessed data into the model, adjusting parameters to minimize errors.



Using GenAI is Straightforward

- Cloud
- License
- API
 - I.e., Meta AI's official API for LLaMA
 - Register for an API Key
 - Configure API Settings
 - Use Python/other to send API requests to LLaMA
 - Receive output data
- Pipelines

```
import requests

api_key = "YOUR_API_KEY"
model_name = "llama"
input_text = "Your input text here"

url = f"https://api.huggingface.co/models/{model_name}"
headers = {"Authorization": f"Bearer {api_key}"}
data = {"inputs": input_text, "parameters": {"temperature": 0.7}}

response = requests.post(url, headers=headers, json=data)
output = response.json()["generated_text"]

print(output)
```

Gen AI and LLM Success



- Organizations will successfully implement GenAI and large language models without facing knockout challenges related to data quality, governance, ethical compliance, and cost management
- Generative AI apps, like chatbots, will become commonplace in daily tasks, with large language models being a powerful and user-friendly form of AI
- Small Language Models Fine-tuned AI models will emerge for specific industries and use cases

Reliability and Trust with RAG

Retrieval-Augmented Generation combines the strengths of LLMs and specific knowledge sources

Enhances accuracy, relevance, and diversity of generated content

- **LLMs and RAG: A Powerful Duo:**
 - LLMs provide context and language understanding
 - RAG leverages external knowledge to augment LLM capabilities
- **Key Benefits:**
 - Improved accuracy and relevance
 - Increased diversity and novelty
 - Enhanced ability to handle nuanced and complex topics
- **Applications:**
 - Content generation (text, chatbots, etc.)
 - Question answering and information retrieval
 - Summarization and knowledge graph construction



Future of GenAI is Bright

- Fact-Checking Itself
- Multi-Modality
- Improve Reasoning Ability
- Bigger Context Windows

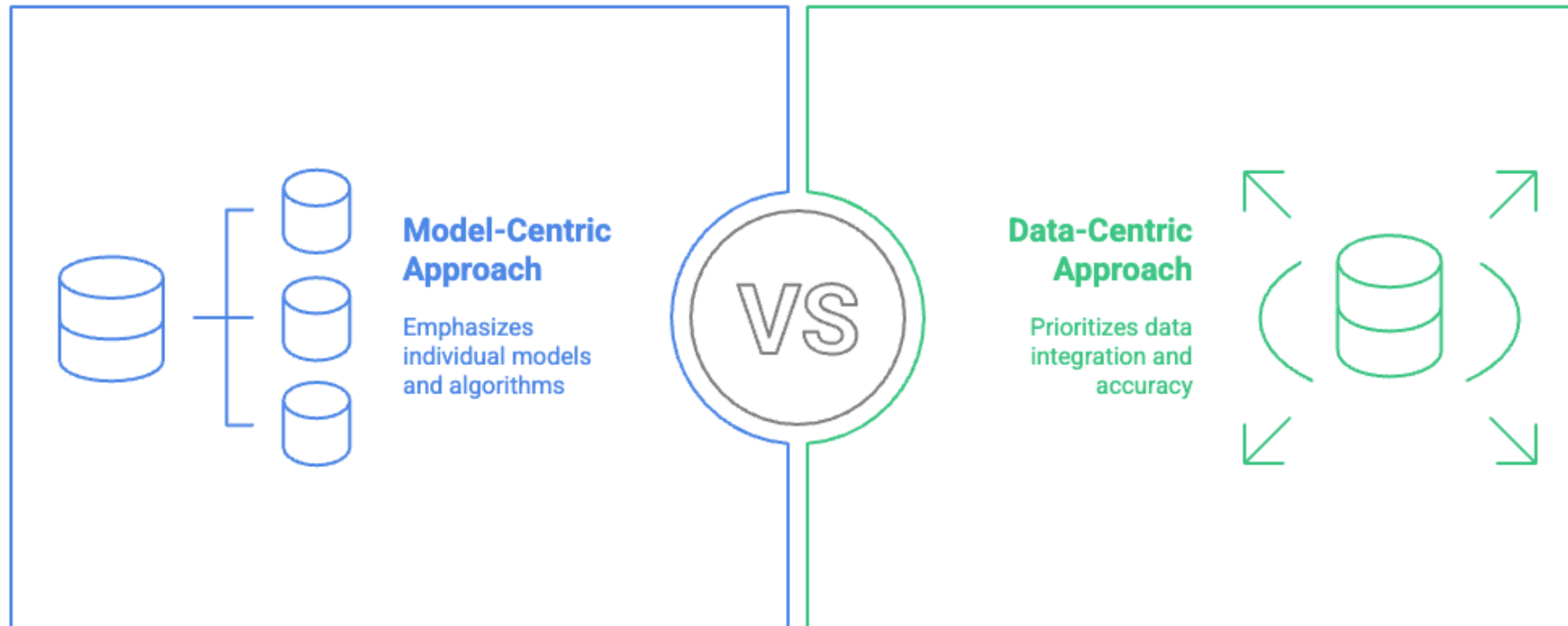


Trends Affecting Data Teams

A photograph of a grand, multi-story library. The space is filled with tall wooden bookshelves reaching up to the ceiling, packed with books. A central staircase with a dark wooden railing leads down to a lower level. In the background, a large, ornate stained-glass window is visible. The lighting is warm and ambient, highlighting the architectural details and the vast collection of books.

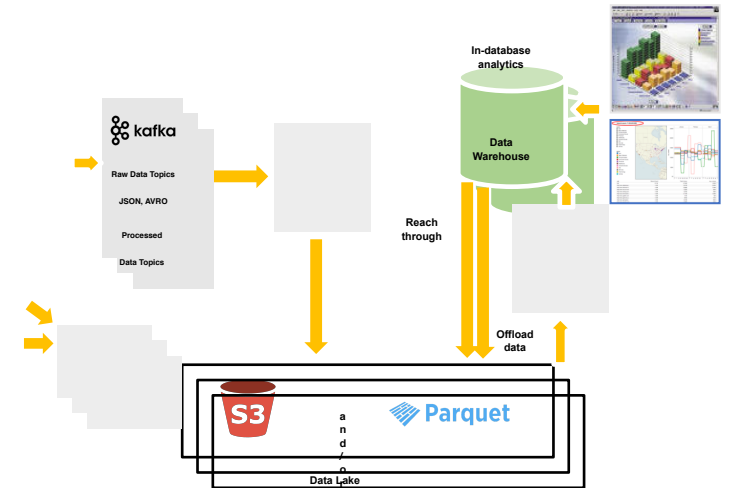
AMERICA · AFRICA
ASIA · AUSTRALASIA

A Focus on Data



Data Mesh Implementations

- Data Mesh emphasizes self-service analytics, domain-driven data ownership, data-as-a-product, and federated governance
- In 2025, data mesh will become a significant trend, focusing on distributed data management
- IT will play a crucial role in creating and distributing **data products** throughout the enterprise
- Data mesh initiatives are being driven from line-of-business units and business leaders, rather than central IT teams



Financial Concerns Drive Efficiency

- Businesses are facing a significant challenge in managing cloud data expenses, which are escalating unsustainably
- By analyzing usage patterns, optimizing storage solutions, and adopting FinOps principles, businesses can balance cost savings and operational efficacy, with a rise in FinOps dashboards in 2025
- To maintain high-quality service and competitive performance, businesses must analyze cloud expenses, identify inefficiencies, and consider cost-effective storage options
- Regular monitoring, forecasting, and implementing financial best practices in cloud management can help organizations balance cost savings and operational efficacy
- In 2025, FinOps dashboards will be increasingly used to manage cloud data charges



The Need to Operationalize ML

- Companies struggle to realize the full potential of machine learning (ML) due to difficulties in putting models into production, inefficient processes, and complexity overload, resulting in untapped value and hindered scalability
- Few businesses have been successful in putting the majority of their ML models into production, leaving a sizable amount of value untapped
- Machine learning operations, also known as MLOps, are a set of standards, tools, and frameworks that are used to scale ML to reach its full potential

More Data Governance and Regulation

- European AI Act affect
- Digital resilience is becoming a crucial aspect of economic output, with regulators enacting strict governance and compliance frameworks to ensure businesses can prepare for adverse events
- Minimum standards for incident response plans, data security, supply chain diversification, risk assessment, disaster recovery plans, cyber hygiene practices, and reporting protocols
- Governments will offer tax breaks or subsidies to companies that invest in building and demonstrating strong resilience capabilities
- Cyber insurance mandates



Organizations and Process

A photograph of a grand, multi-level library. The space is filled with tall wooden bookshelves on both the ground and upper levels, densely packed with books. A central staircase with a dark wooden railing and decorative balusters leads down to a lower level. In the background, a large, ornate stained-glass window is visible. The ceiling is high with exposed wooden beams and several hanging light fixtures. The overall atmosphere is one of a well-maintained, historic institution.

AMERICA · AFRICA
ASIA · AUSTRALASIA

Current Approach

Roles

- Project Manager
- Data Engineer
 - Data Integration+Architecture
- Data Analyst
- Data Modeler
- Data Scientist
- CDO
- Business Analyst
- Data Governance Group
- Data Security Analyst
- Business Intelligence Architect
- System/Cloud Engineer
- DevOps Engineer

Process

- Agile'
- Major Steps
 - Data Preparation
 - Data Modeling
 - Design and Prototyping
 - Development and Testing
 - Deployment and Maintenance
- Data Governance
- Data Security
- User/Project Community

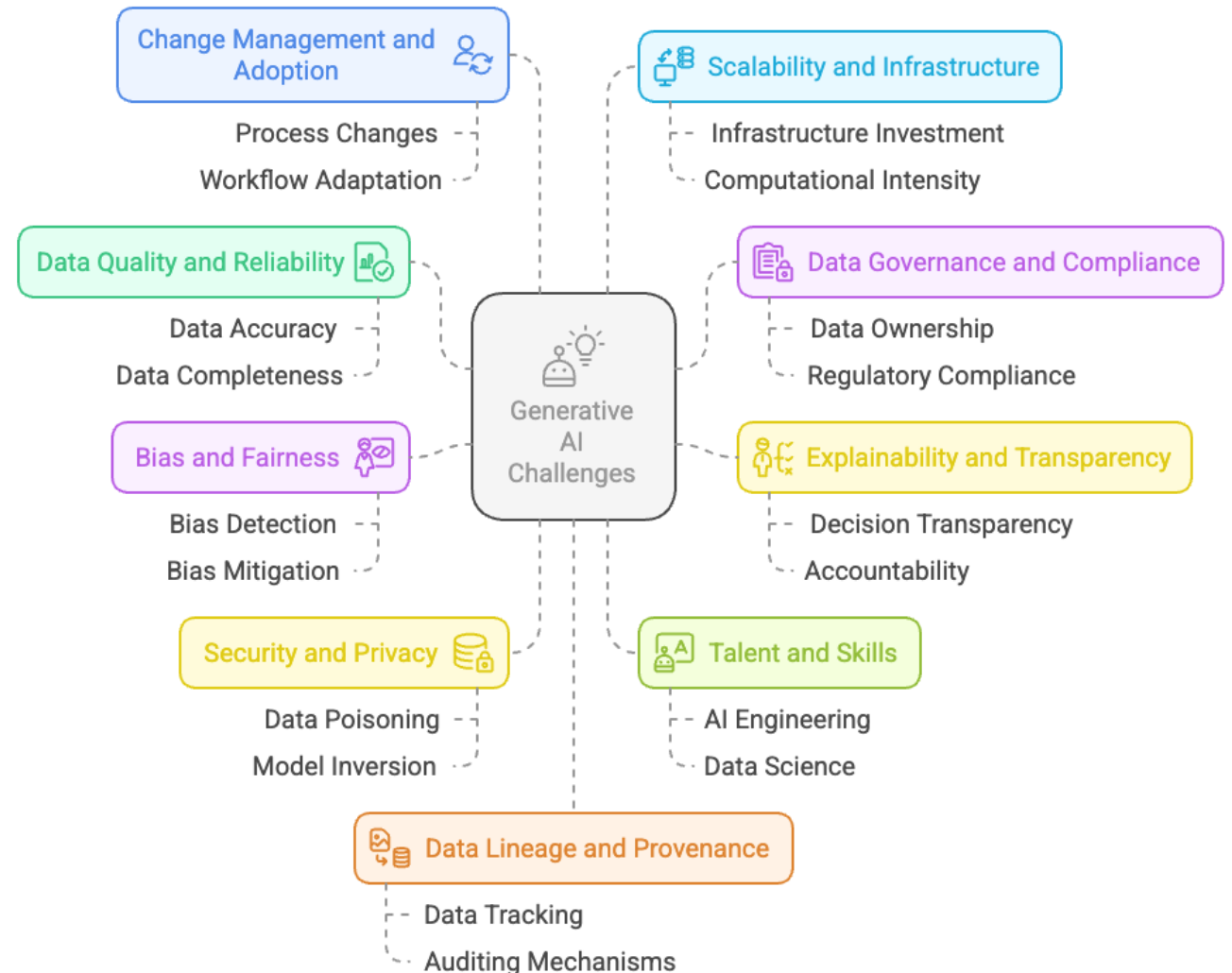
GenAI Organizational and Process Challenges

•Process Challenges

- Align AI with business goals
- Operationalizing ML

•Organizational Readiness

- Internal Resistance
- Emerging Roles
- Skills Gap



GenAI Requires Organization Transformation

- Require much more than “the right data” and “a good database” and “good technology”
- Present great opportunities, but also poses significant implementation risks
- Encounter many risks that are “people” related, which must be managed for successful implementation



GenAI Impacts to Organization and Process

- Cross-Functional Teams
- New Roles and Responsibilities
- Data Ownership and Governance
- AI Center Of Excellence
- Inclusion of Unstructured and Highly Diverse Data
- New Model Development
- Continuous Monitoring and Feedback

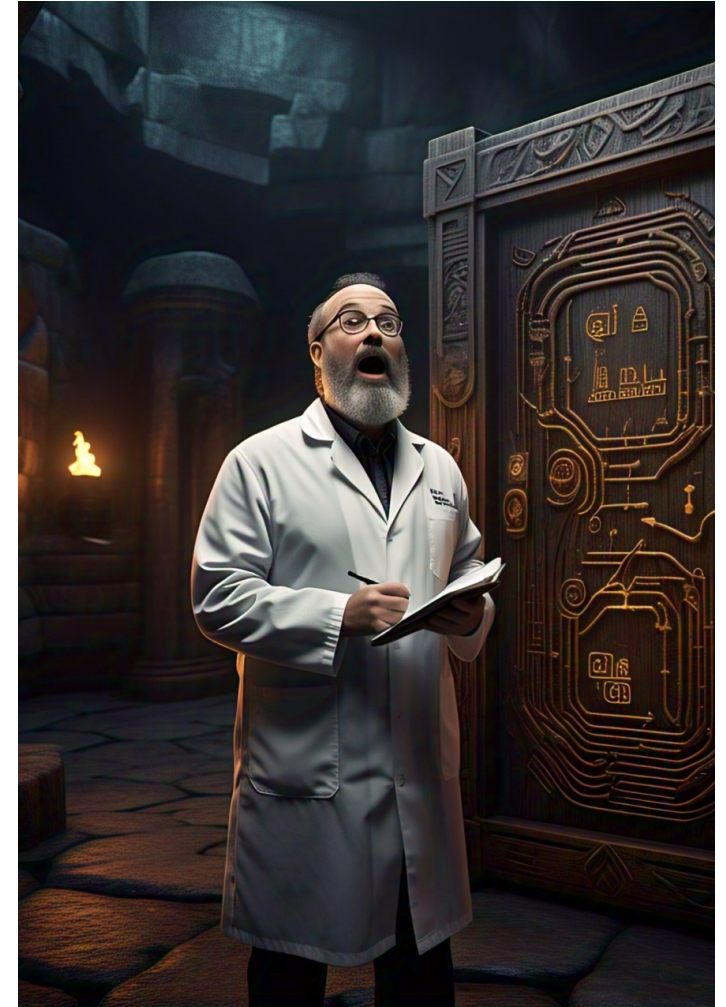
Human-in-the-Loop

- Human-in-the-loop refers to the involvement of human oversight, review, and validation in AI-driven decision-making processes.
- This approach acknowledges that AI systems are not perfect and can make mistakes, ensuring that humans are involved in critical decision-making steps.



New Approach for Data Scientists

- **Data Preparation**
- Model Selection and Hyperparameter Tuning
- **Faster Experimentation and Prototyping**
- Rapid Model Development
- Automated Model Evaluation
- **Increased Focus on Model Evaluation and Validation**
- Model Interpretability and Explainability



Change in Data Scientist Role



- Added focus on data curation
- **Increased Focus on Model Use**
- **More Emphasis on Domain Expertise**
- **Specialization**
- **New Skills:**
 - GenAI Frameworks and Tools
 - Deep Learning and Neural Networks
 - Data Augmentation and Generation
 - Explainability and Interpretability

Emerging Roles in the Enterprise

- Prompt Engineer
- AI Trainer
- AI Auditor
- AI Ethicist
- Machine Manager
- Cybersecurity Analyst
- AI Business Strategist
- Data Broker
- AI Explainer



Data Career Tips in a GenAI World

- Asking "How?" gets you to tactics & implementation. Asking "**Why**?" will get you to strategy and the real objectives.
- Pull the trigger: An imperfect decision can be fixed, but a non-decision will never happen.
- Don't always defend your actions.
- The most self-deceiving are those who've learned something and think they needn't learn anything else.

Summary

- GenAI is revolutionary
- Many trends are trends because of GenAI and they must be considered:
 - A Focus on Data
 - Data Mesh Architecture
 - Driving for Efficiency
- GenAI presents many challenges to the current/pre-GenAI organization and processes
- Data Scientists must become proficient in automation, efficient experimentation, model interpretability, domain expertise and GenAI Frameworks
- There are emerging roles and career strategies to embrace: growth mindset, strategic questions, decisive action, and remaining open to learning and feedback





From Models to Data: How Gen AI is Changing the Game for Data Scientists and Data Teams

Presented by: William McKnight

"#1 Global Influencer in Big Data" Thinkers360

President, McKnight Consulting Group

3 X **Inc 5000**

 /in/wmcknight

www.mcknightcg.com
(214) 514-1444

