IDA

INSTITUTE FOR DEFENSE ANALYSES

DATAWorks 2024: Operational T&E of AI-supported Data Integration, Fusion, and Analysis Systems

Breeana G. Anderson, Project Leader

Adam M. Miller
Logan K Ausman
John T Haman
Keyla Pagan-Rivera
Sarah A Shaffer
Brian D Vickers

March 2024

IDA Product ID 3001848

IDA

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

Rigorous Analysis │ Trusted Expertise │ Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

DATAWorks 2024: Operational T&E of AI-supported Data Integration, Fusion, and Analysis Systems

Breeana G. Anderson, Project Leader

Adam M. Miller
Logan K Ausman
John T Haman
Keyla Pagan-Rivera
Sarah A Shaffer
Brian D Vickers

Executive Summary

DoD strategies include an important role for artificial intelligence (AI) in future military systems. However, relatively few DoD systems currently implement AI, and the test and evaluation community has limited experience working with AI-supported systems. This has resulted in significant questions about how to test these systems, especially in operational settings.

This briefing discusses an approach for the operational test and evaluation (OT&E) of AI-supported data integration, fusion, and analysis systems. These systems combine data from various sources into a useable database through a series of user-guided and AI-supported data transformations. Data integration systems exist across DoD and are especially relevant to intelligence processing.

DoD must identify and mitigate new OT&E challenges posed by AI in order to support evaluations of system effectiveness and suitability. These challenges include tradeoffs in selecting appropriate data to ingest during testing, characterizing system performance over the operational envelope, and evaluating the AI's contribution to the system output.

Overcoming these challenges will require a combination of new and existing methods, including:

- using confusion matrices and regression error to quantify how well system output matches ground truth,

- using interrater agreement when ground truth is unknown, and

- augmenting traditional human-systems integration metrics with an assessment of trust calibration.

This briefing includes a notional test concept demonstrating how to apply these methods to an AI-supported data integration system. The test concept provides examples of how to characterize human-systems interaction throughout the data pipeline using a combination of surveys, focus groups, and behavioral observations; it discusses how to evaluate AI model performance in terms of the quality of the data products they support; and it presents a method for quantifying the specific AI contributions to system performance by comparing the system under test against a comparison (e.g., legacy) system. Lastly, operational factors relevant to a design of experiments for these types of systems

are discussed, including which characteristics of the system
and the test influence the sample sizes needed

DATAWorks 2024: Operational T&E of AI-supported Data Integration, Fusion, and Analysis Systems

Adam M. Miller

John T. Haman

Logan K. Ausman

Keyla Pagán-Rivera,

Sarah A. Shaffer

Brian Vickers

April 2024

Institute for Defense Analyses

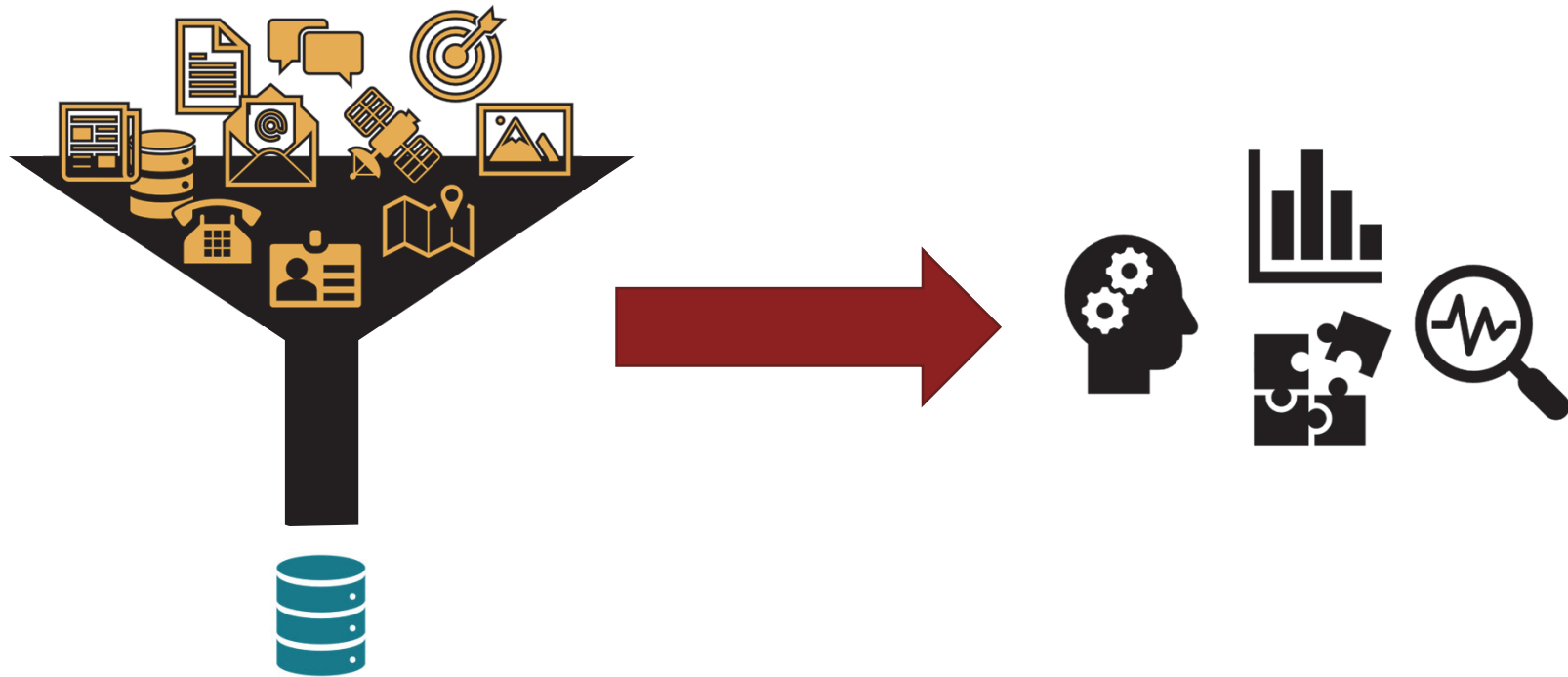730 East Glebe Road ● Alexandria, Virginia 22305

Outline

- Introduction to AI-supported data integration

- Operational T&E challenges posed by AI-supported data integration

- Strategies for operational T&E of AI-supported data integration systems
  - Test data
  - Methods
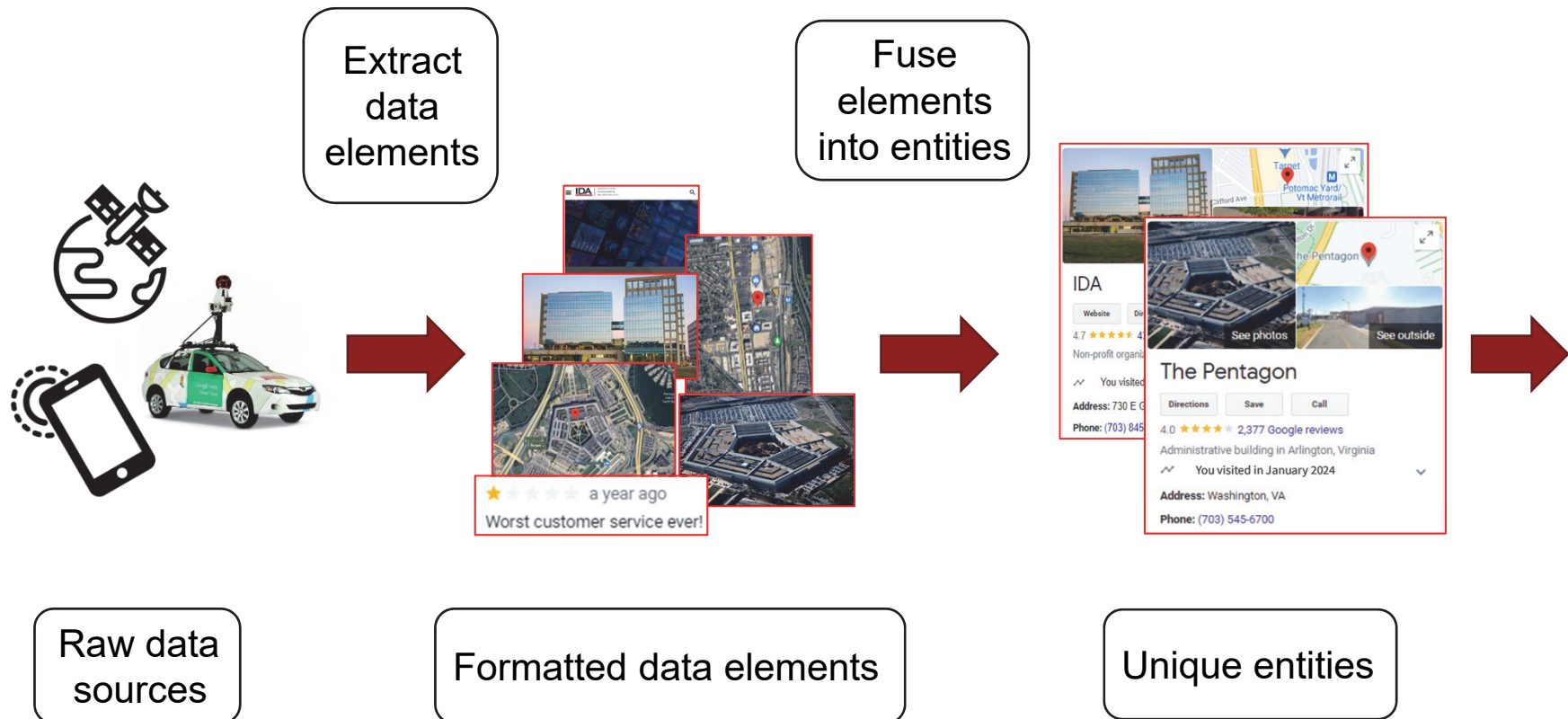 - Evaluation
  - Test design

- Summary

AI – Artificial Intelligence; T&E – Test & Evaluation

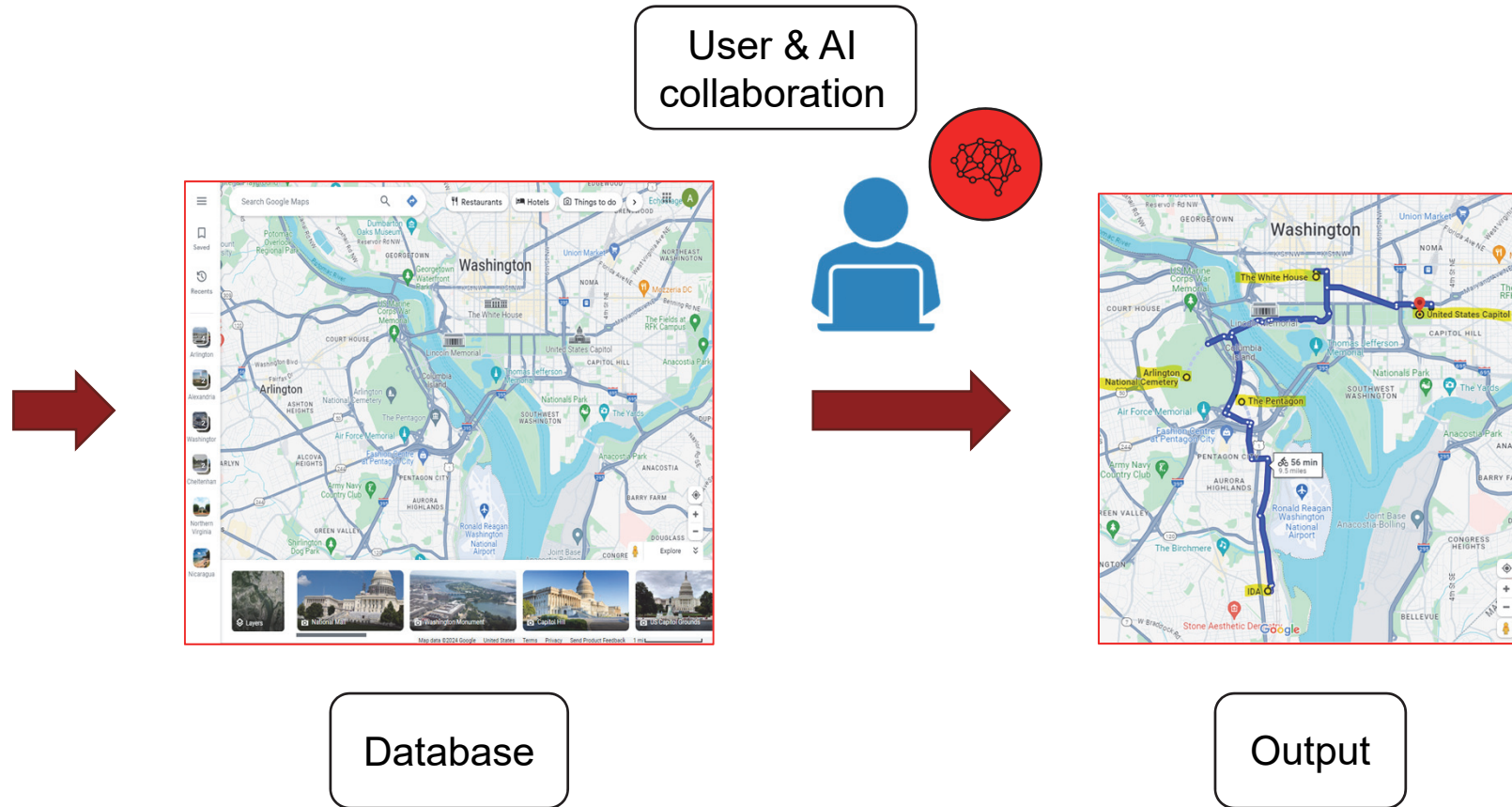# AI Capabilities for Data Integration

AI − Artificial Intelligence

Data Integration is the gathering, formatting, fusing, and refining of data from multiple sources to facilitate analysis
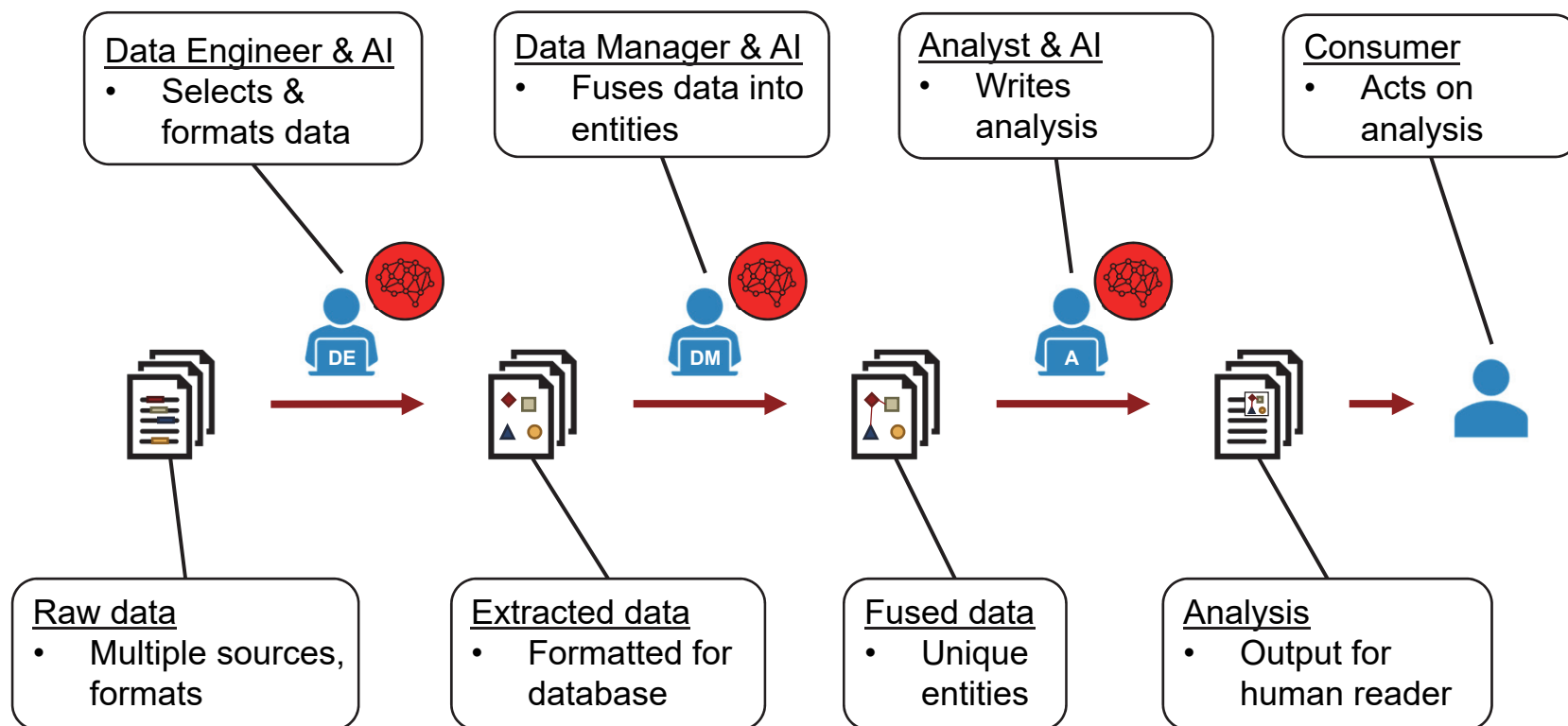
Google Maps as an example data integration system (1 of 2)

Extract data elements

Fuse elements into entities

Raw data sources

Formatted data elements

Unique entities

Google Maps as an example data integration system (2 of 2)

User & AI collaboration

Database

Output

AI – Artificial Intelligence

Notional AI-supported data integration pipeline

Data Engineer & AI
- Selects & formats data

Data Manager & AI
- Fuses data into entities

Analyst & AI
- Writes analysis

Consumer
- Acts on analysis

Raw data
- Multiple sources, formats

Extracted data
- Formatted for database

Fused data
- Unique entities

Analysis
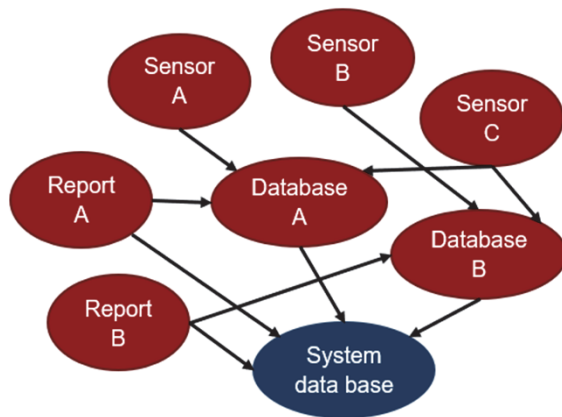- Output for human reader

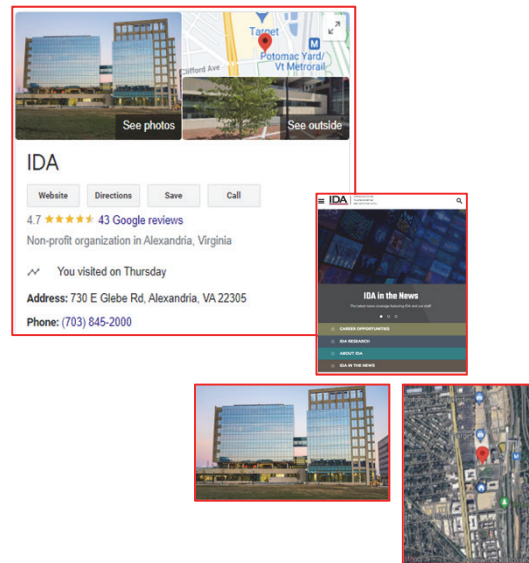A – Analyst; AI – Artificial Intelligence; DE – Data Engineer; DM – Data Manager

Operational T&E Challenges Posed by AI-supported data integration
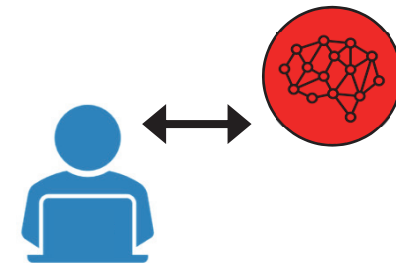
Test challenges for operational T&E of AI

Operationally representative **test data**

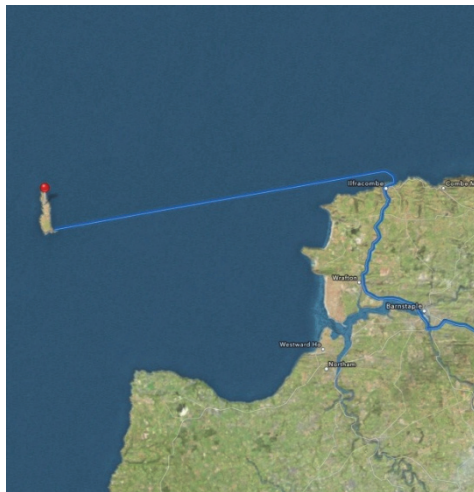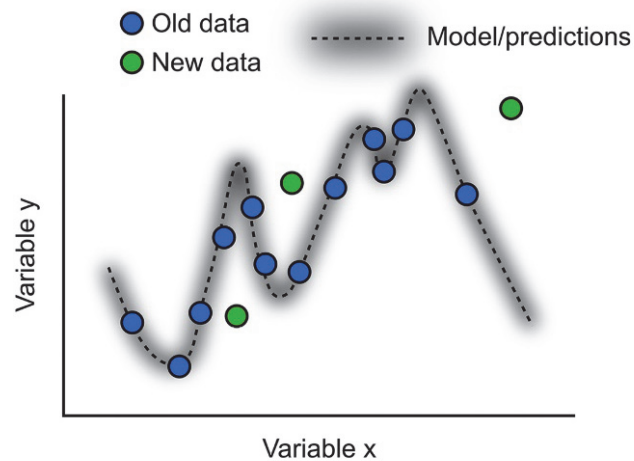Grading the **accuracy** of responses

What was the **AI contribution**?

AI – Artificial Intelligence; T&E – Test & Evaluation

Characterizing performance over the operational envelope

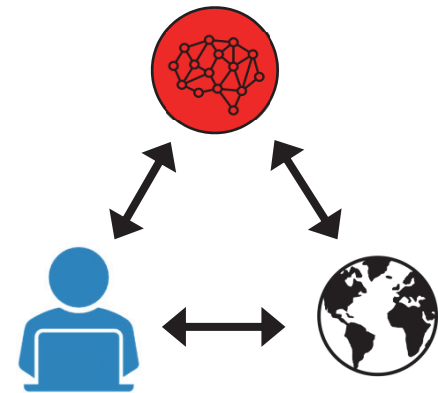### Predicting **performance** with rare inputs

Theamazingios6maps.tublr.com

Generalizability of the model solution

Richards & Frankland (2017) The Persistence and Transience of Memory
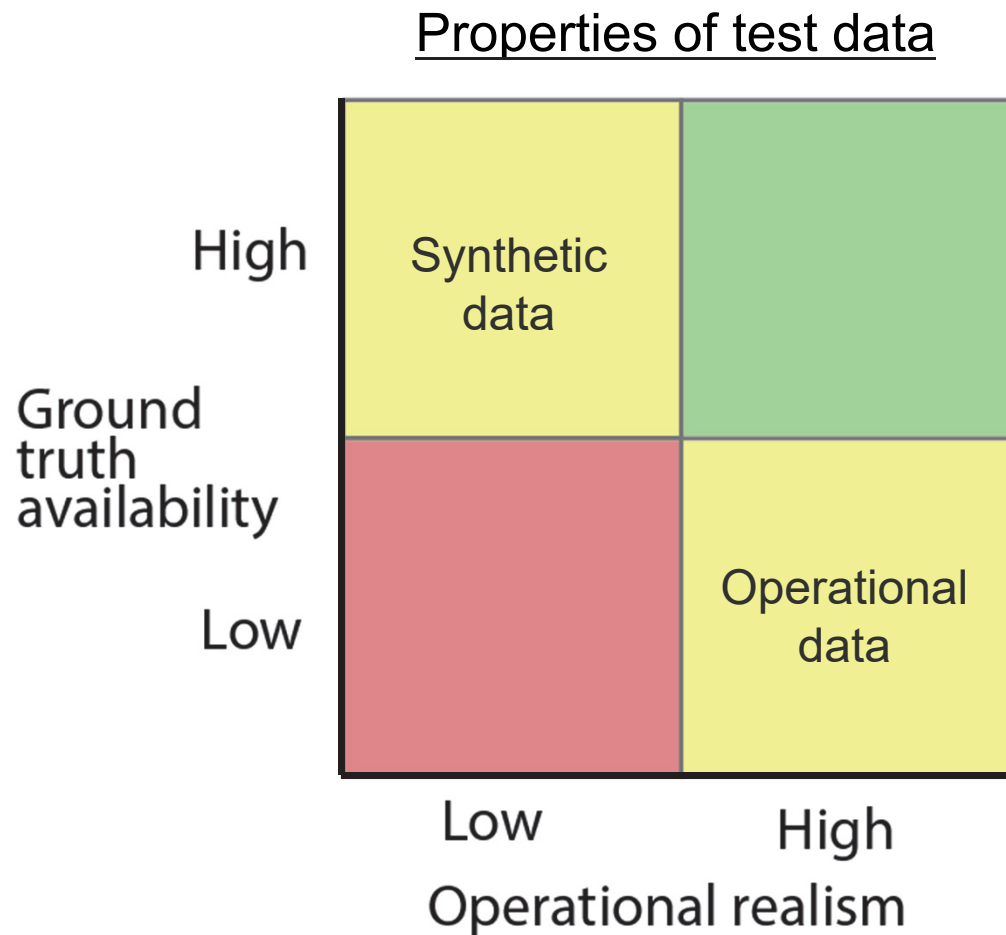
Model may **change** during testing

IDA   9

Strategies for T&E of AI-supported data integration system: <u>Test data</u>

AI – Artificial Intelligence; T&E – Test & Evaluation

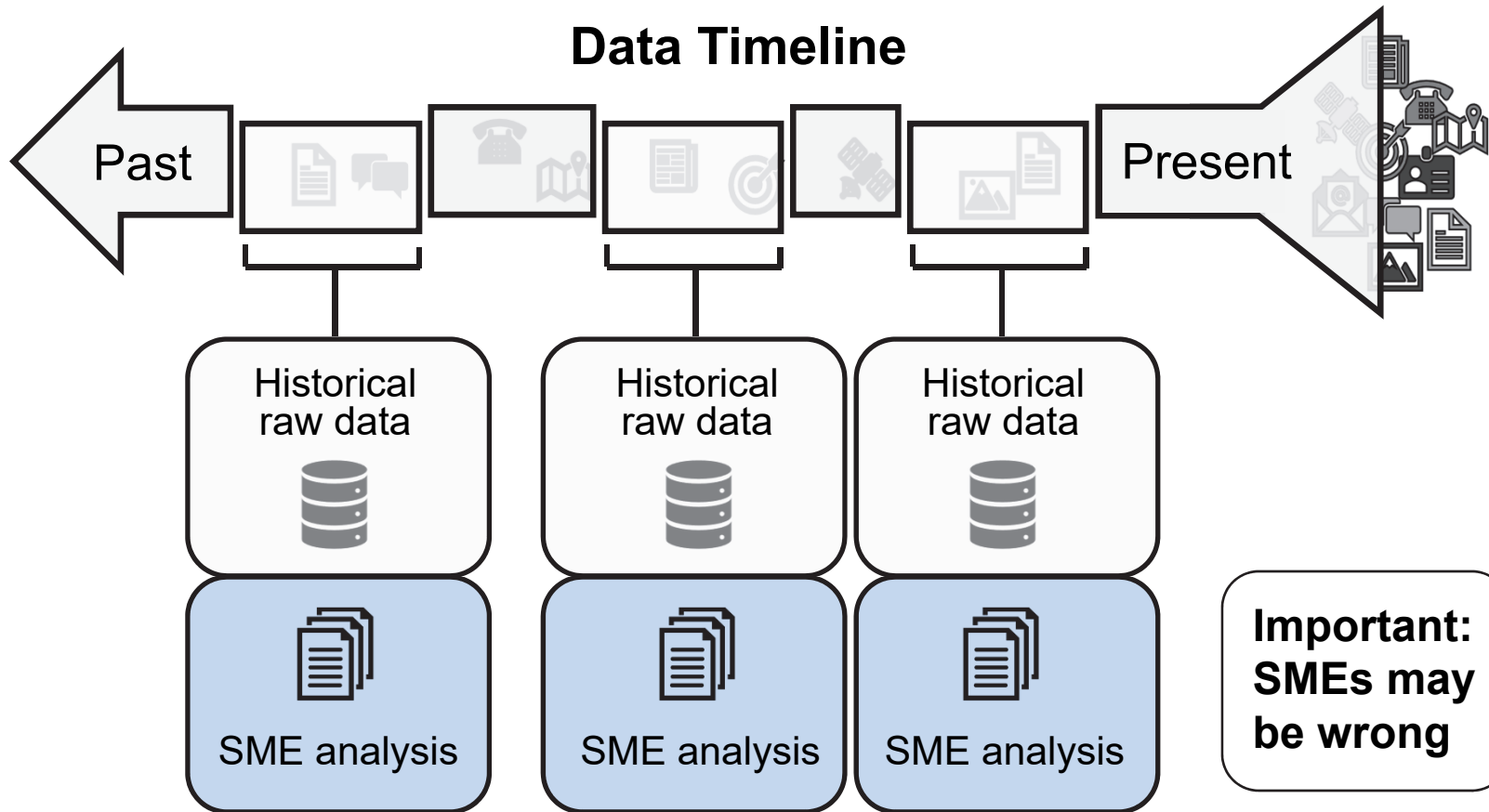# Operationally representative test data

Properties of test data

How to improve?
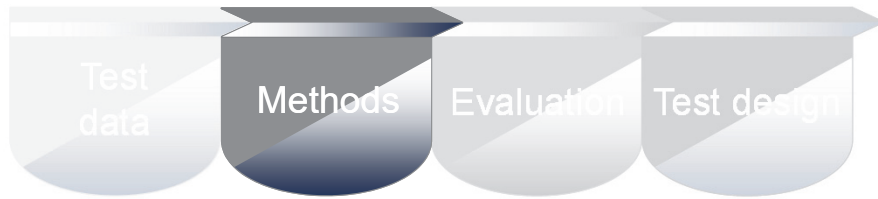- **Enhance** operational realism of synthetic data
- **Approximate** truth for operational data
- **Combine** synthetic and operational data

Approximating ground truth for operational data

Real historical data with SME retrospective analyses

Strategies for T&E of AI-supported data integration system: <u>Methods</u>

AI − Artificial Intelligence; T&E − Test & Evaluation

IDA | 13

Evaluating the accuracy of AI responses

For when you have ground truth

Confusion matrix

		Truth	
		Rating A	Rating B
AI Model	Rating A	Agree A	Disagree
	Rating B	Disagree	Agree B

Regression error

AI Model prediction

Error Distance

Truth variable y

Truth variable x
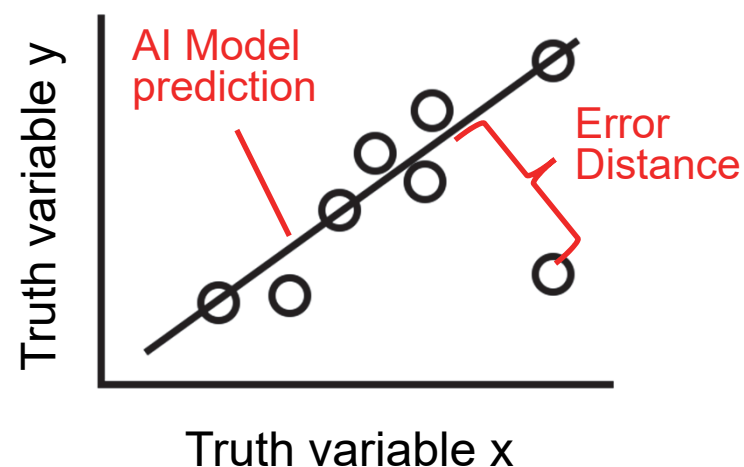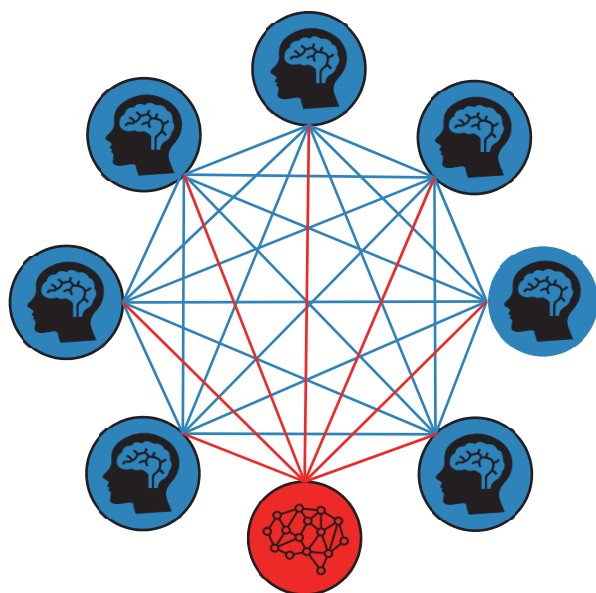
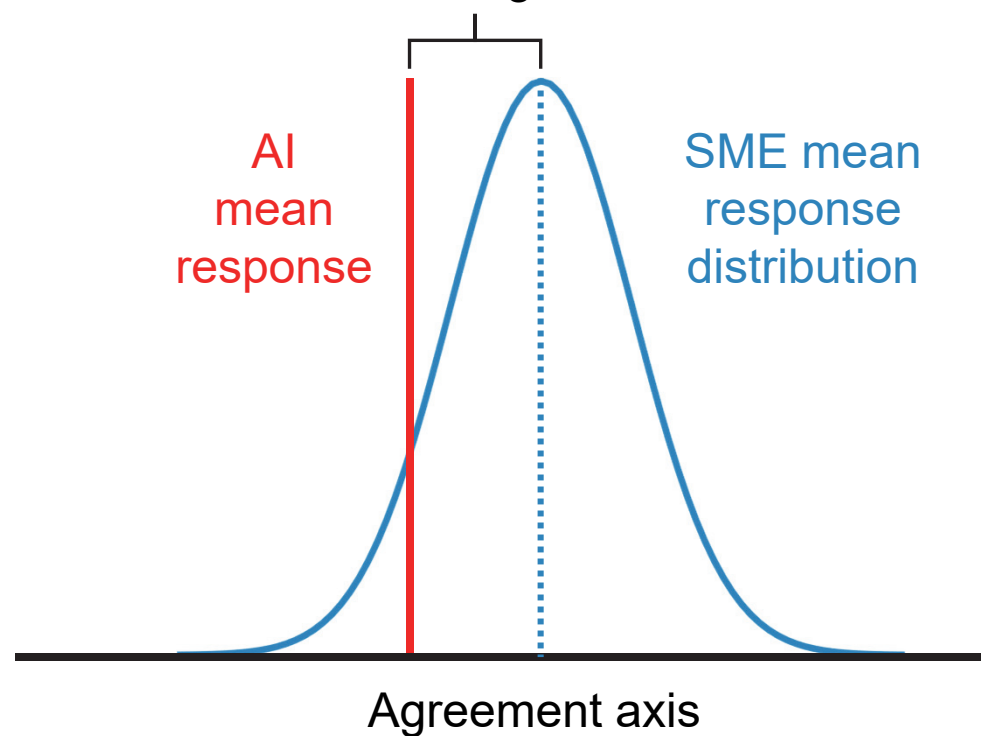# Compare AI responses to SME responses

For characterizing the similarity of AI and SME responses

Interrater agreement

AI-SME Disagreement

AI mean response

SME mean response distribution

Agreement axis

AI − Artificial Intelligence; SME − Subject Matter Expert

IDA | 15

Characterize user trust and utilization

In addition to standard HSI measures

Calibrated trust

Human trust (vertical axis), **AI accuracy** (horizontal axis)
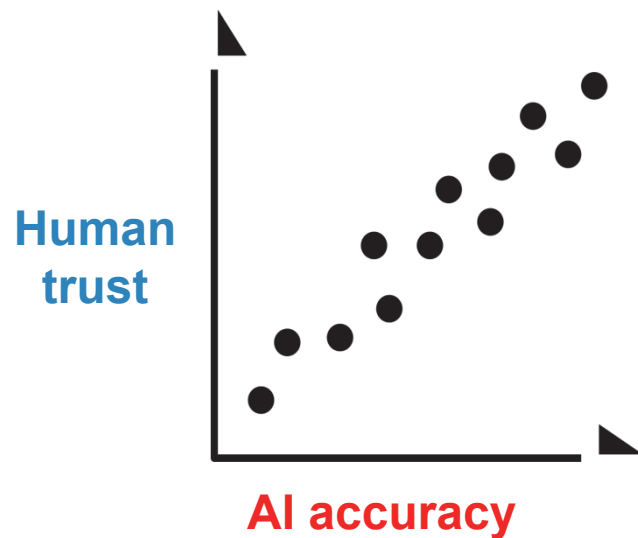
User utilization confusion matrix

		Human Behavior	
		Intervention	No Intervention
AI Accuracy	Unsatisfactory	Hit	Miss
	Satisfactory	False Alarm	Correct Rejection

Strategies for T&E of AI-supported data integration system: <u>Evaluation</u>

Evaluate ability of users to generate data analyses

Curated historical scenario (raw data) → Engineer & AI (DE) → Extracted data → Manager & AI (DM) → Correlated data → Analyst & AI (A) → Analysis → Consumer
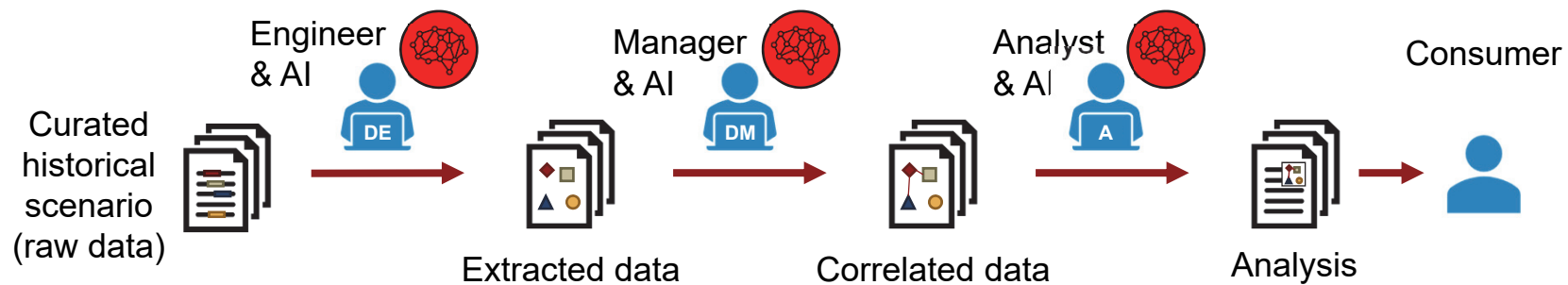
A – Analyst; AI – Artificial Intelligence; DE – Data Engineer; DM – Data Manager

Evaluate ability of users to generate data analyses

HSI response variables (all users)
- Usability
- Workload
- Calibrated trust
- Utilization
- Timeliness

Consumer approval

Engineer & AI — DE

Manager & AI — DM

Analyst & AI — A

Consumer

Curated historical scenario (raw data)

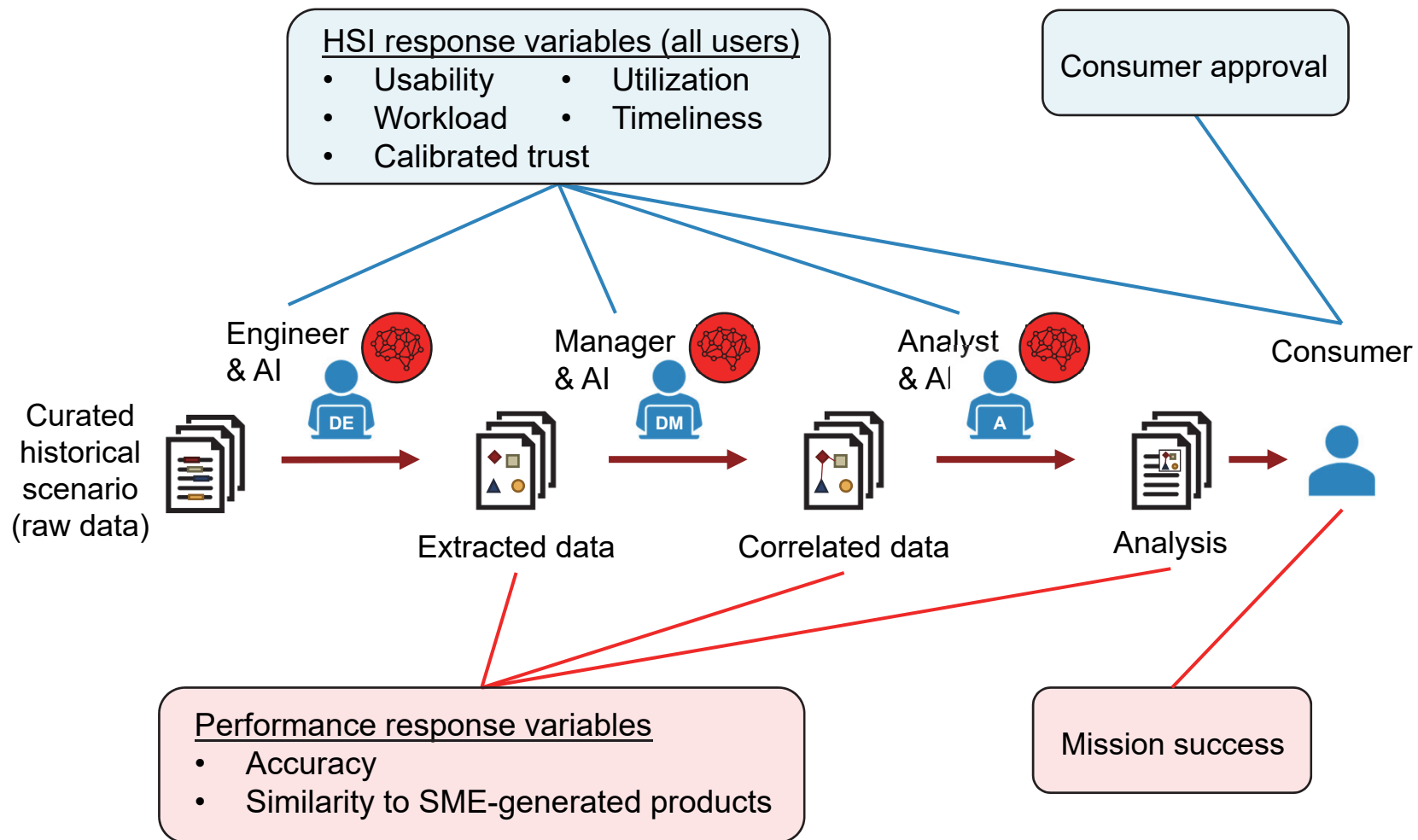Extracted data

Correlated data

Analysis

A – Analyst; AI – Artificial Intelligence; DE – Data Engineer; DM – Data Manager; HSI – Human-Systems Interaction

Evaluate ability of users to generate data analyses

HSI response variables (all users)
- Usability   • Utilization
- Workload   • Timeliness
- Calibrated trust

Consumer approval

Engineer & AI

Manager & AI

Analyst & AI

Consumer

Curated historical scenario (raw data)

DE

DM

A

Extracted data

Correlated data

Analysis

Performance response variables
- Accuracy
- Similarity to SME-generated products

Mission success

A – Analyst; AI – Artificial Intelligence; DE – Data Engineer; DM – Data Manager; HSI – Human-Systems Interaction; SME – Subject Matter Expert

IDA | 20

Evaluate ability of users to generate data analyses

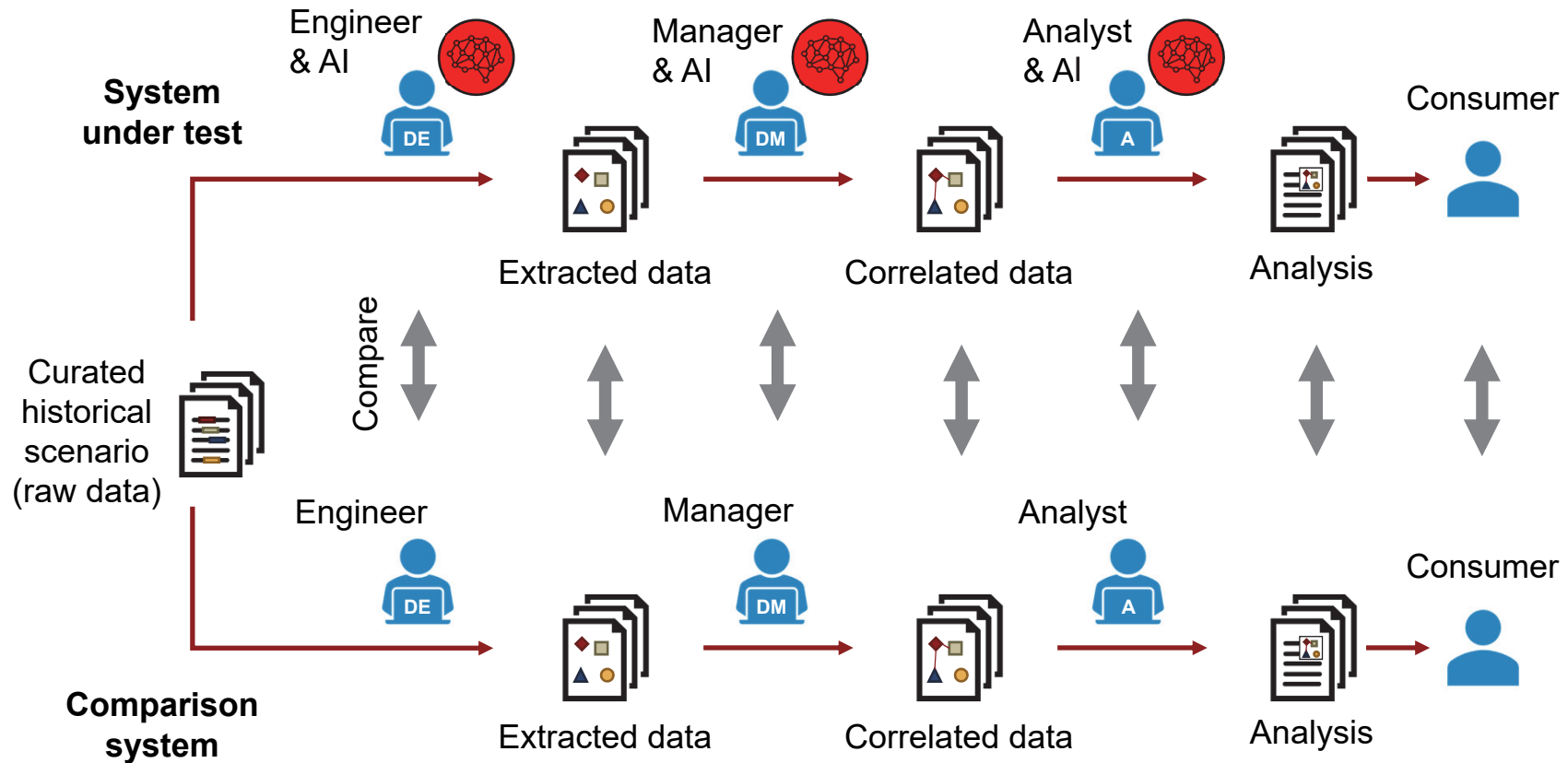A – Analyst; AI – Artificial Intelligence; DE – Data Engineer; DM – Data Manager

Strategies for T&E of AI:
Test Design

Candidate operational factors

Focus should be on operational factors that affect the <u>input</u> to the data integration system

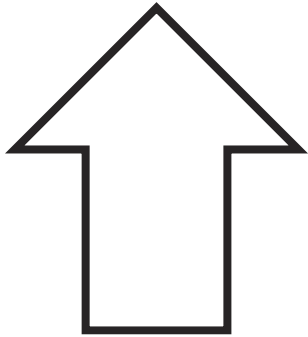Data Sources

Geographical regions

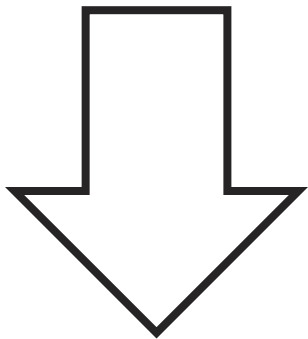Entity Types

Network Domain

Duplicate records

■ ■ ■

Sizing adequate tests

Factors that increase the number of runs needed:

- **Novelty** of system/ mission/ capability/ AI task
- **Magnitude** of AI contribution to mission
- Task **Complexity** (input and decision spaces)

Factors that decrease the number of runs needed:

- **Quality** of training data (relevance & quantity)
- Ground **truth** data availability
- Prior **testing** (relevance & recency)

AI – Artificial Intelligence

Summary

- AI data integration evaluations may become common

- Data classification and interrater agreement metrics are useful for assessing AI performance

- Standard HSI measures should be augmented with behavioral measures of calibrated trust and AI utilization

- Use between-systems comparisons to characterize AI contributions to system performance

- Characterizing the operational space requires identifying operational factors governing system input

AI – Artificial Intelligence; HSI – Human-Systems Interaction

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*

2. REPORT TYPE

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			
					19b. TELEPHONE NUMBER *(Include area code)*