Paradata

What is it and Why do we care?

Patricia C. Franks, PhD, CA, CRM, IGP, CIGO, FAI

Professor Emerita, San José State University

San José, California, USA

Overarching Question

If business is no longer to be transacted only by human beings, but also by AI agents, or some combination of the two, what will evidence of those transactions look like, what will the record be?”

~Jenny Bunn, PHD

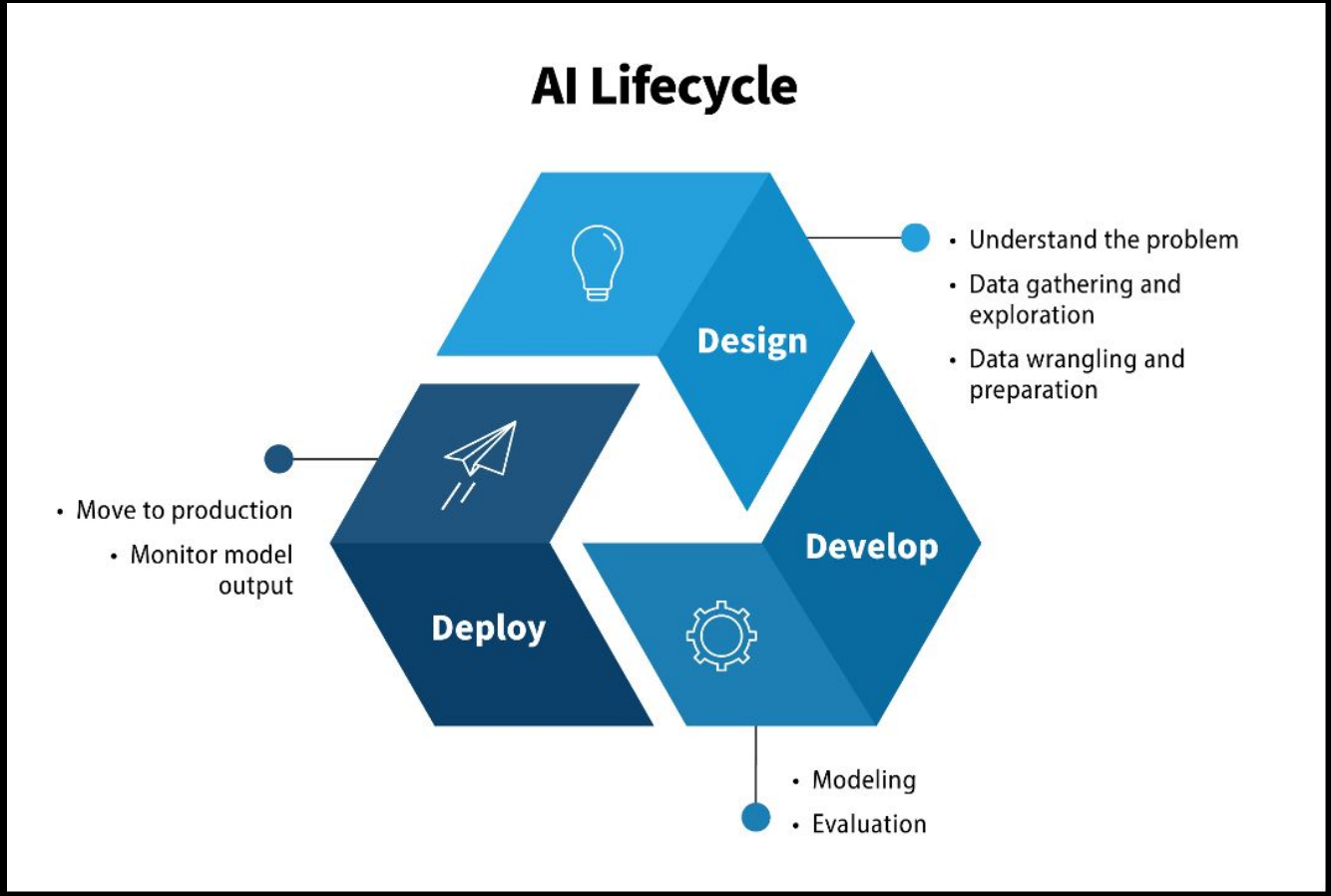
InterPARES Researcher from The National Archives of the United Kingdom

What is it?

Paradata: Document the AI Process

Paradata is **the information about the procedure(s) and tools** used to create and process information resources, along with **information about the persons** carrying out those procedures.

~ITrustAI working definition



AI Lifecycle. Source: *AI Guide for Government: A Living and Evolving Guide to the Application of Artificial Intelligence for the U.S. Federal Government*, GSA, Centers of Excellence.
https://coe.gsa.gov/coe/ai-guide-for-government/understanding-managing-ai-lifecycle/index.html

Paradata as AI processual documentation

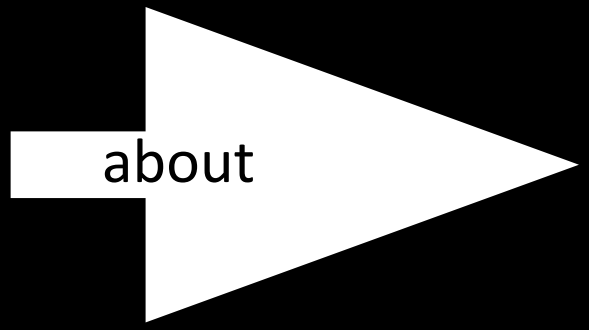
Paradata must document the full scope of application and context of use – not just the algorithm itself.

- XAI: why did a given tool produce a given output from a given set of inputs?
- Paradata: why, how, and to what effect was a given tool used in a particular context?

The National Archives (UK): “Building explainable AI is not just an algorithmic matter, but needs to consider the individuals and the environment in which it will operate” (Jaillant et al., 2020)

Metadata & Paradata -- relationships & purpose

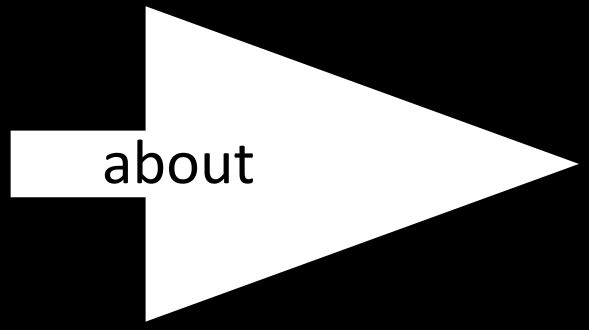
Metadata



The Information Resource

For the purposes of documenting, describing, preserving or managing that resource.

Paradata



The AI Process

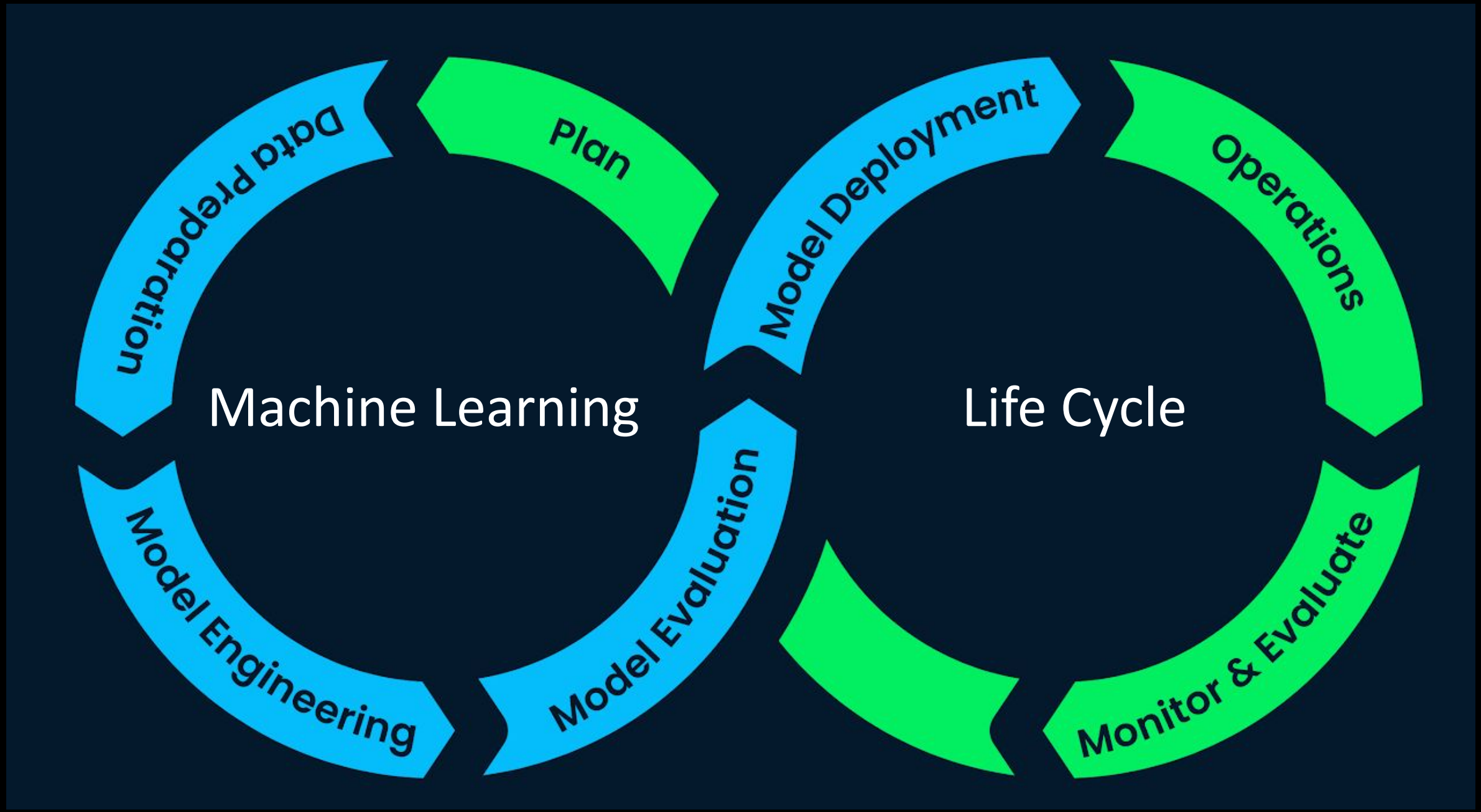
Enables processual insight, transparency, accountability.

Principles supporting an Archival Perspective

- the sanctity of evidence;
- *respect des fonds*, provenance, and original order;
- the life cycle of records;
- the organic nature of records; and
- relationships between records and their descriptions.

History in the true sense depends on the unvarnished evidence, considering not only what happened, but why it happened, what succeeded, what went wrong.
-Burke (1997)

The integrity of the evidential value of materials is ensured by demonstrating an unbroken chain of custody--*Provenance*


Machine Learning
Life Cycle
Plan
Data Preparation
Model Engineering
Model Evaluation
Model Deployment
Operations
Monitor & Evaluate

Examples of Paradata

Technical Paradata

- AI Model (tested & selected)
- Evaluation & performance metrics
- Logs generated
- Model training data set
- Training parameters for model
- Vendor documentation
- Versioning information

Organizational Paradata

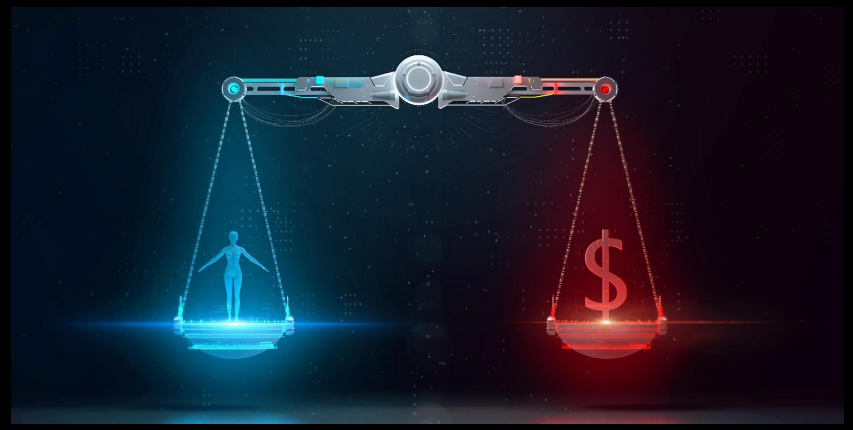
- AI policy
- Design plans
- Employee training
- Ethical consideration
- Impact assessments
- Implementing process
- Regulatory requirements

Why do we care?

AI is here!

AI must be Governed!

AI Governance
The EU AI Act



EU countries give crucial nod to first-of-a-kind Artificial Intelligence law

By Luca Bertuzzi | EURACTIV **Est. 6min** Feb 2, 2024

Next steps

The European Parliament's Internal Market and Civil Liberties Committees will adopt the AI rulebook on 13 February, followed by a plenary vote provisionally scheduled for 10-11 April. The formal adoption will then be complete with endorsement at the ministerial level.

The AI Act will enter into force 20 days after publication in the official journal. The bans on the prohibited practices will start applying after six months, whereas the obligations on AI models will start after one year.

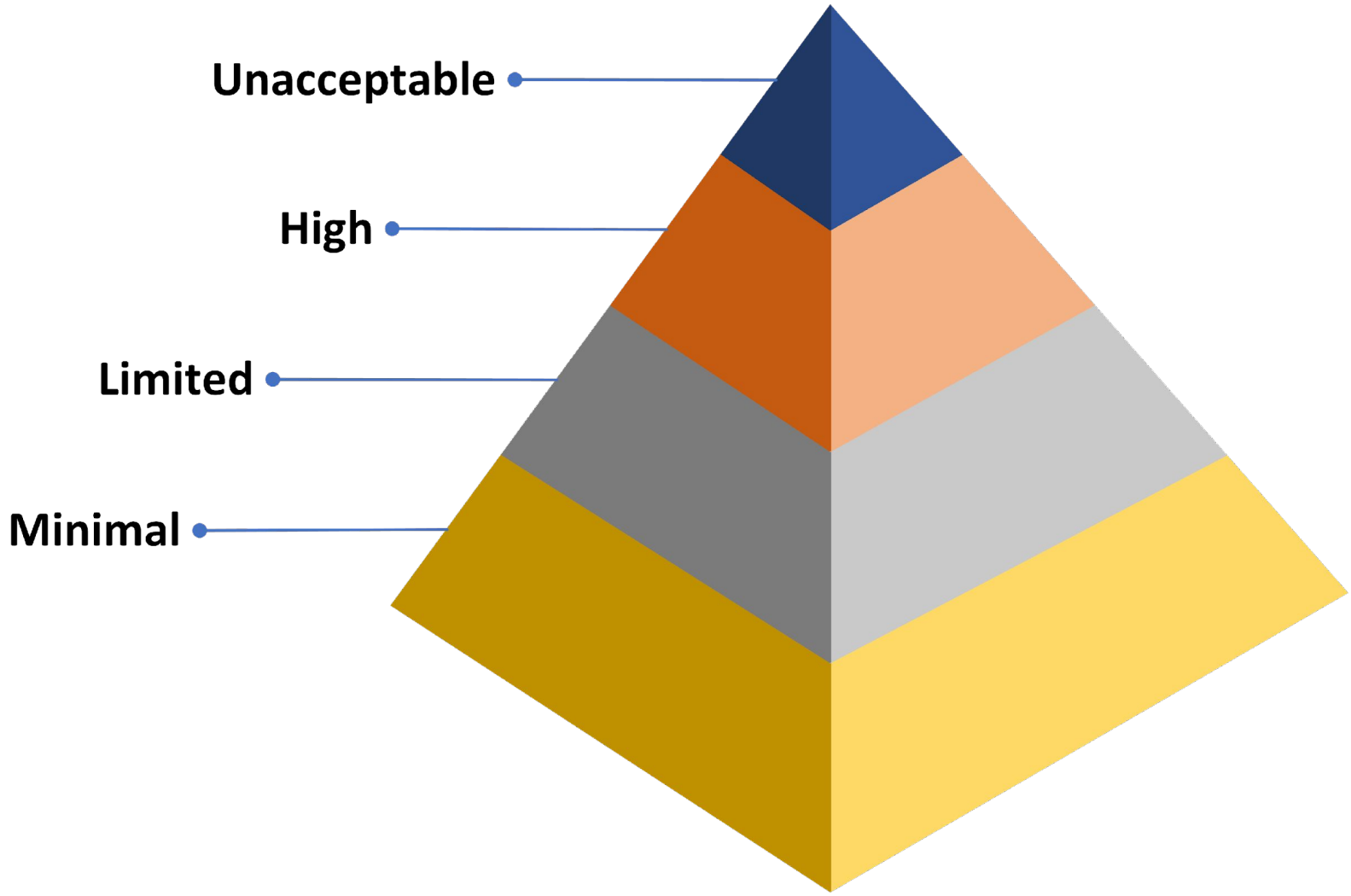
All the rest of the rules will kick in after two years, except for the classification of AI systems that have to undergo third-party conformity assessment under other EU rules as high-risk, which was delayed by one additional year.

[Edited by Alice Taylor]

EU Regulatory framework proposal on artificial intelligence

https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

A Risk-Based Approach



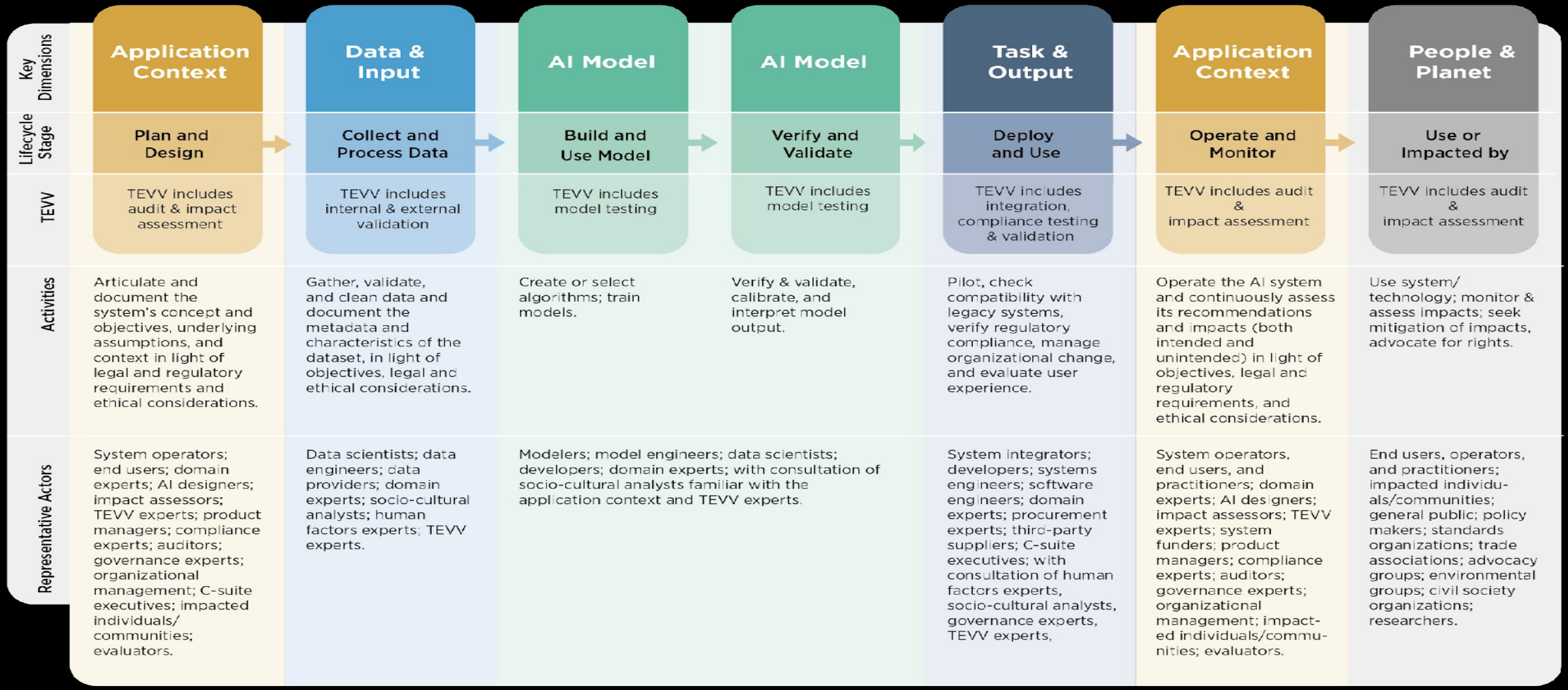
Documentation based on Risk Level: The EU Act

High Risk *Compliance Obligations*	• Conformity assessments (against the EU Act) • Fundamental rights impact assessments
Limited Risk *Transparency Obligations*	• Informing users they are interacting with an AI system • Marketing synthetic audio, video, text, and images content as AI-generated or manipulated
Minimal or No Risk *Free Use*	• Voluntary codes of conduct are encouraged

Artificial Intelligence Risk Management Framework (AI RMF 1.0)

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

January 2023
US Department of Commerce



AI actors across AI lifecycle stages. Note that AI actors in the AI Model dimension are separated as a best practice--those building and using the models are separated from those verifying and validating the models.

Terms Employed in Documents that fall within the concept of Paradata: NIST RMF

Govern	• Applicable laws and regulations, AI policies, risk management policies, impact assessments, data governance practices, training programs.
MAP	• Information disclosure plan; risk profile (tolerance); AI system profile, knowledge limits, output; datasheets for datasets; TEVV metrics
MEASURE	• Stakeholder engagement plans; algorithmic methodology; usability testing, system test results to measure limitations and errors
MANAGE	• Privacy impact assessments; incidence response plans; model cards and fact sheets; procedures for retiring the system; third-party contracts/terms of service

Canada

Bill C-27, with the text of the Government's **proposed amendments** to the *Artificial Intelligence and Data Act* (AIDA).

artificial intelligence system means a technological system that, using a model, makes inferences in order to generate output, including predictions, recommendations or decisions.

machine learning model means a digital representation of patterns identified in data through the automated processing of the data using an algorithm designed to enable the recognition or replication of those patterns.

High Impact AI Systems

- Employment-related decisions
- Provision of services
- Biometric information processing
- Content moderation & prioritization on communications platforms
- Healthcare and emergency services
- Court or administrative body decision-making
- Law enforcement

Artificial Intelligence and Data Act (AIDA) - Canada (6 Mandated Risk Reduction Tasks)

Anonymization of data and management of anonymized data
Assessment of high-impact status
Assessment of risks and mitigation for possible harms or biased outputs
Ongoing monitoring and risk mitigation measures
Publish plain-language description of systems and risk reduction measures taken
Notification to Minister of Innovation, Science, and Industry in case of demonstrated or likely harms emerging from system use.

Document: Overview Model Artificial Intelligence Governance Framework-Singapore. Second Edition

- The Model Framework is based on two principles:
  - Organizations using AI in decision making should ensure that the decision-making process is explainable, transparent, and fair.
  - AI solutions should be human-centric (protect the interests of human beings and amplify their capabilities).

MODEL AI Governance Framework -- Key Areas

- Internal governance structures and measures
- Determining the level of human involvement in AI-augmented decision-making
- Operations management
- Stakeholder interaction and communication

Operations Management, which includes data reparation, algorithms, and model, is the most salient section regarding requirements that imply a need for paradata.

Terms Employed in Document that fall within the concept of Paradata

Explicit

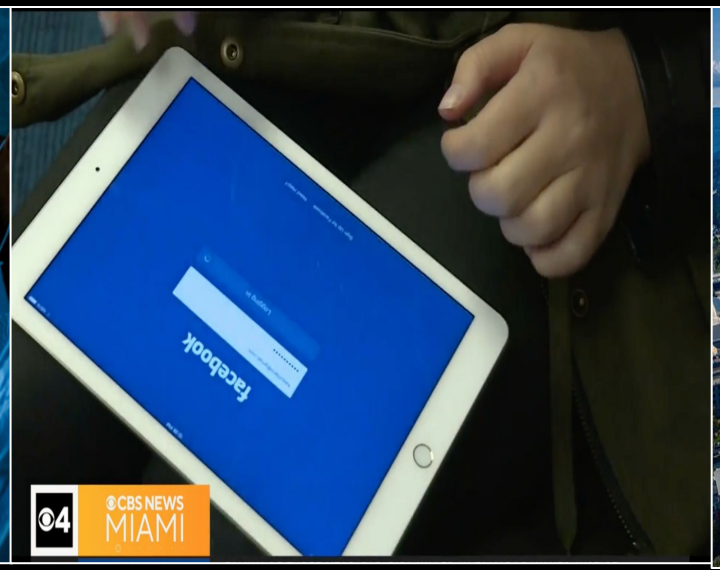
- Audit trail
- Data provenance record
- Dataset
- Decisions documented
- Internal Policy
- Repeatability assessments
- Technical specifications

Implicative

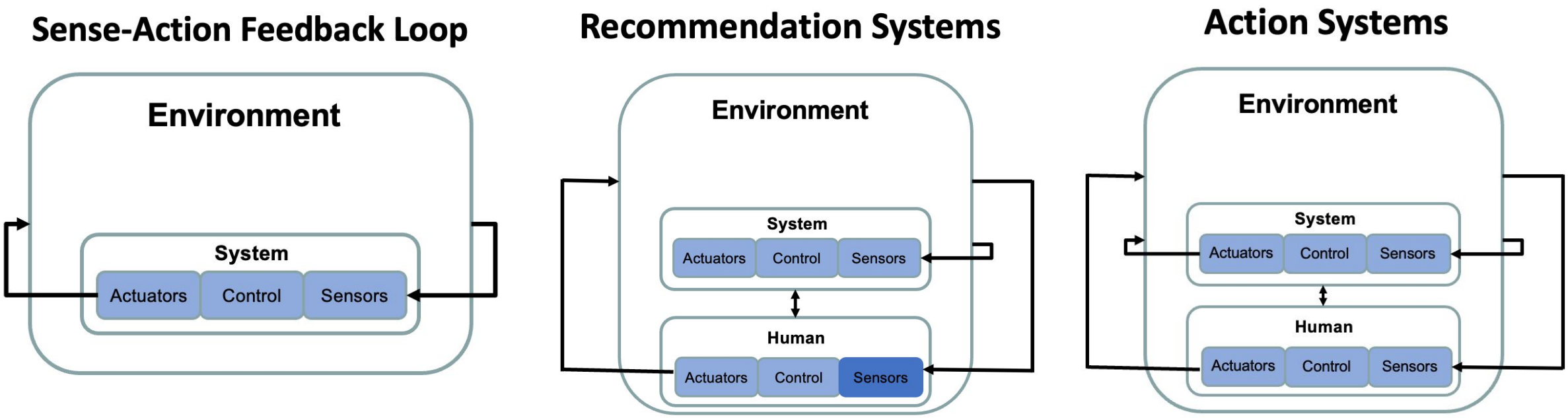
- Accounting for changes
- Black box recorder
- Counterfactual fairness testing
- Traceable

Algorithm audits: Carried out at request of a regulator or Technology Provider on behalf of its customer who must respond to such a request. The goal: To discover the actual operations of algorithms comprised in models

Preserving paradata for accountability of semi-autonomous AI agents in dynamic environments: An archival perspective



Excerpt from paper by Scott Cameron and Babak Hamidzadeh - Pre-print is available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4681230



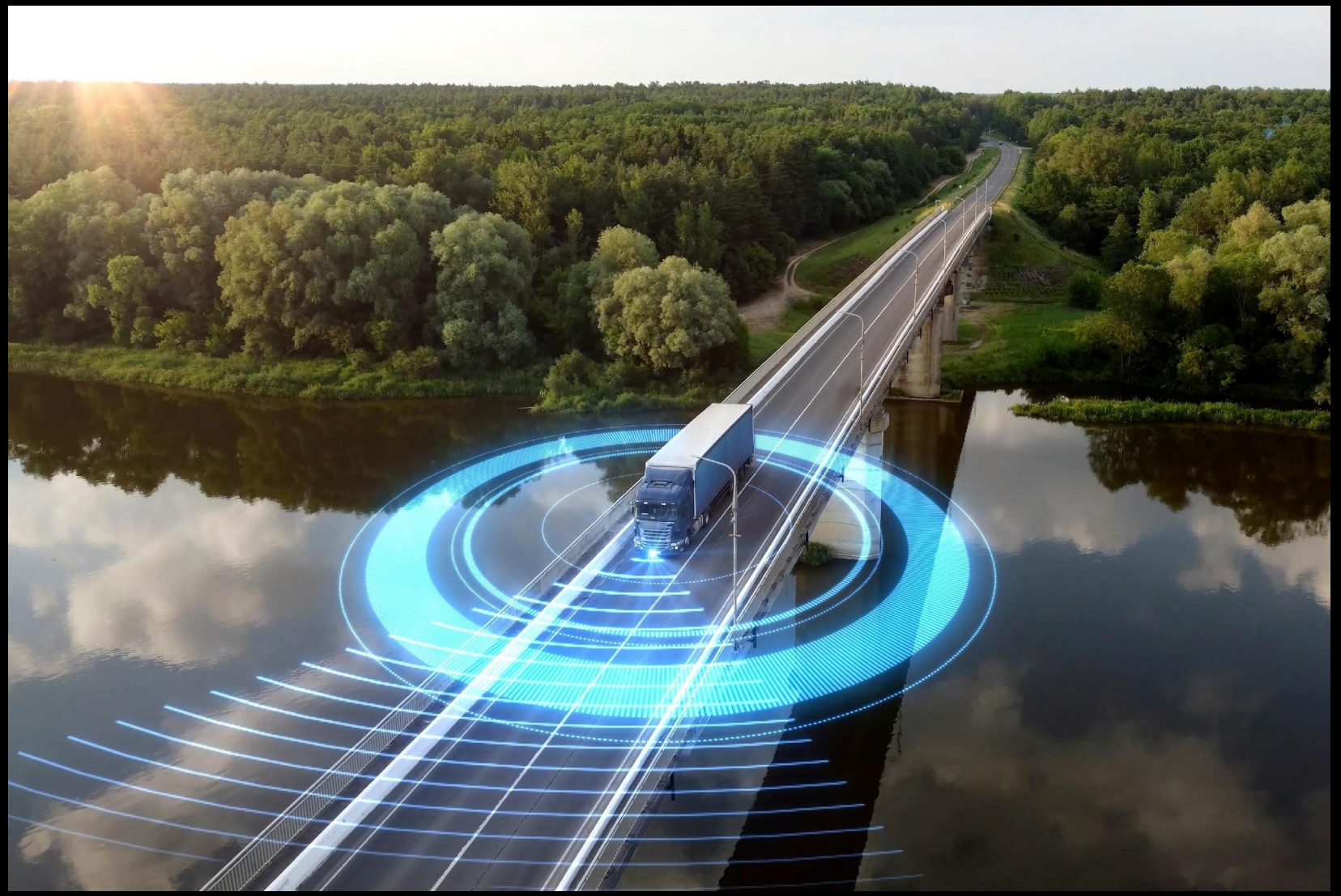
Continuous cycle:

- Sensors (HW or SW) measure the real world,
- Measurements are fed into & inform the control (AI) processes
  - Control processes determine responses to real-world stimuli
- Actuators (HW or SW) execute or effect the responses to real-world stimuli
- Consequence of system actions are measured by the system's sensors
- and the cycle continues …

General Paradata

What is documented	Example documents/records
System by itself (independently of its specific uses and operation)	Preservation of the system itself and its versions; HW/SW architecture and design diagrams; Code, model, algorithms, logic and executables; Maintenance and upgrade documentation; Training data, test data and results, validation data and results; Means for running the system
Governance and compliance information	Organizational records documenting self-auditing processes, acceptance tests, change control; impact assessments, employee training

Autonomous semi-truck with a trailer, controlled by artificial intelligence



Operational Paradata: Autonomous Vehicles

Sensor Input (real world inputs): Log of sensor data (speedometer, GPS data, thermometers, steering mechanisms, etc.); camera footage used for computer vision systems.	**Actuators Mechanisms to put control decisions into effect):** Log of human control actions; log of automated system's control actions as implemented; log of messages communicated from system to human controller and external parties.
Controller (control directions system is using): Log of control directions; relevant settings of control systems; paradata from intermediary processes leading up to a decision; post-facto AI explanations of these processes; log of warning notifications and control handover notifications.	**Effects (real word changes into effect by system):** Log of sensor data; camera footage.

Surgeons Perform Heart Surgery Using Augmented Reality Technology

Mixed Reality Surgical Assistance

Sensor Input (real world inputs): Real-time camera footage.	**Actuators Mechanisms to put control decisions into effect):** Surgeon retains full control and assesses the real-word plus the AR's interpretation; additional camera footage , audio, and written records may record their actions.
Controller (control directions system is using): Screen capture of the real-time interpretations provided by computer vision system; paradata from intermediary processes leading up to a decision; post-facto AI explanations of these processes.	**Effects (real word changes into effect by system):** Consequences of the surgeon's actions visible to surgeon and recorded by camera. Additional follow-up assessments with patient may be documented as well.

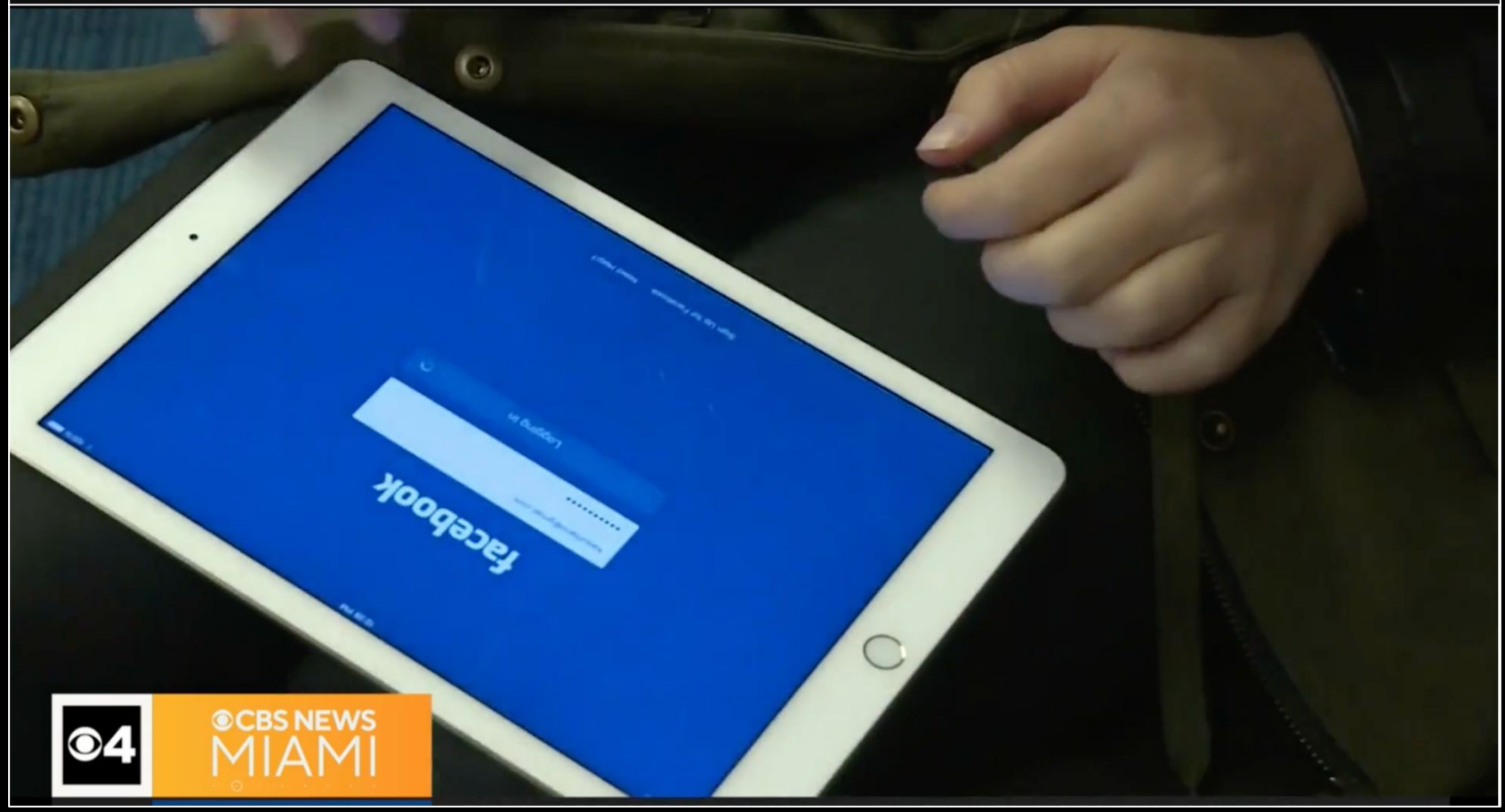
Social Media Content Targeting

MONEYWATCH ›

Meta deliberately targeted young users, ensnaring them with addictive tech, states claim

MONEY WATCH

By **Khristopher J. Brooks**
Updated on: November 27, 2023 / 3:07 PM EST / MoneyWatch



Social Media Content Targeting Systems

Sensor Input (real world inputs): Log of advertisements show to user and user's interaction with system during each session (ex. mouse use and scroll behavior).	**Actuators Mechanisms to put control decisions into effect):** Log of system's decisions of advertising and content to display.
Controller (control directions system is using): Log of selections of ads to display; paradata from intermediary processes leading to decision; post-facto AI explanations of processes; records of progression of a user's algorithmic profile upon which decisions are made; records of the advertising available at a given time.	**Effects (real word changes into effect by system):** System may influence user's behavior while using platform. It may only measure changes through data captured by sensors; screen capture records users' experiences.

Chattanooga Digital Twin Project

https://www.nrel.gov/news/program/2023/digital-twin-project-green-lights-traffic-congestion-improvements.html

Digital Twin (YVR airport facility management)

Sensor Input (real world inputs): Real-time camera footage, sensor data (pedestrian/vehicle traffic counters, weather data, maintenance and infrastructure data).	**Actuators Mechanisms to put control decisions into effect):** Orders and notifications issued to human controllers; traffic control directions issued.
Controller (control directions system is using): Log of decisions produced by the system; paradata from intermediary processes leading up to a decision; post-facto AI explanations of these processes.	**Effects (real word changes into effect by system):** Changes in traffic patterns (pedestrians, ground vehicles, planes), security or maintenance actions. Recorded through sensor inputs and by camera.

Conclusions

- Decisions made and actions taken by AI-enabled systems must be documented.
- Some of the documentation will be automatic as part of the AI system; some will be human-created prior to or after the creation and implementation of the AI system.
- Paradata is recommended to document the AI process and promote transparency and accountability.
- The archival perspective supports the capture and preservation of paradata—and is necessary to ensure the AI process is captured in a way that preserves the characteristics of authoritative records: authenticity, reliability integrity, and usability.
- Guidance in the form of laws, regulations, and frameworks must be monitored and applied to the AI process.

Conclusions

- A risk-based approach is recommended for AI governance; high-impact, high-risk AI implementations are the most challenging.
- Dynamic systems that rely on real-time data to make decisions and perform actions may involve both the AI agent and a human agent; the capture of paradata in these environments is complex and requires attention.
- Standards are emerging.
- A full-fledged information and AI governance structure is necessary.
- Risk-value-cost trade-offs must be considered.
- Every organization should begin to discuss the topic of AI Governance.

Thank you!

Dr. Patricia C. Franks
CA, CRM, IGP, CIGO, FAI
Professor Emerita
San Jose State University
Patricia.franks@sjsu.edu

Resources

- Bunn, J. (2020), "Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI)", *Records Management Journal*, Vol. 30 No. 2, pp. 143-153. https://doi-org.libaccess.sjlibrary.org/10.1108/RMJ-08-2019-0038
- **AI Lifecycle.** Source: *AI Guide for Government: A Living and Evolving Guide to the Application of Artificial Intelligence for the U.S. Federal Government*, GSA, Centers of Excellence. https://coe.gsa.gov/coe/ai-guide-for-government
- Lise Jaillant, Katherine Aske, Annalina Caputo. (2020). AEOLIAN: Artificial Intelligence for Cultural Institution, The National Archives (UK) Case Study, accessed February 9, 2024, https://www.aeolian-network.net/wp-content/uploads/2021/11/AEOLIAN-Case-Study-1-The-National-Archives-UK.pdf
- Burke, Frank G. 1997. *Research and the Manuscript Tradition.* Lanham, Md.: Scarecrow Press.
- CLIR (n.d.) The Archival paradigm: The Genesis and Rationales of Archival Principles and Practices, accessed Feb. 9, 2024, https://www.clir.org/pubs/reports/pub89/archival/#:~:text=The%20integrity%20of%20the%20evidential,and%20tracking%20all%20preservation%20activities
- Ibid. (Duchein 1983)
- Ibid. (provenance)
- Frank Upward and Sue McKemmish. (1994). Somewhere Beyond Custody, accessed February 9, 2024, https://publications.archivists.org.au/index.php/asa/article/download/8403/8397
- EU AI Act, https://artificialintelligenceact.eu/the-act/
- NIST. (2023, January). *Artificial Intelligence Risk Management Framework (AI RMF 1.0, https://www.nist.gov/itl/ai-risk-management-framework*
- Artificial Intelligence and Data Act (AIDA). (2022_. https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act
- Infocomm Media Development Authority and Personal Data Protection Commission. (2020). Model Artificial Intelligence Governance Framework Second Edition, https://iapp.org/media/pdf/resource_center/pdpc_model_framework_ai_governance_second_edition.pdf
- Scott Cameron and Babak Hamidzadeh, (2024). Preserving paradata for accountability of semi-autonomous AI agents in dynamic environments: An archival perspective. [Preprint] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4681230