



# Paradata: *Documenting the AI Process for Transparency and Accountability*

Patricia C Franks, PhD, CA,CRM, IGP, CIGO  
San Jose State University

InterPARES 2023 Summer School

July 7, 2023





*The future is here—  
it's just not evenly  
distributed yet*

*~William Gibson*

## **Topics:**

**1. Basics**

**2. Applications & Risks**

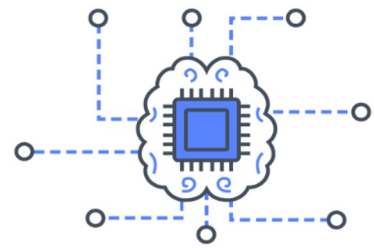
**4. Paradata**



The background of the image is a complex, glowing network of blue and white lines and dots, resembling a neural network or a data mesh. The lines are thin and connect various points, creating a dense, interconnected pattern. The dots are small and vary in brightness, with some appearing as bright white points and others as dimmer blue or grey points. The overall effect is a sense of depth and complexity, with the network appearing to recede into the distance towards a bright, hazy light source at the top center. The text "AI Basics" is centered in the middle of the image, rendered in a clean, white, sans-serif font. The text is slightly larger than the surrounding network elements, making it the focal point of the composition. The lighting is soft and diffused, with a subtle gradient from the bright center to the darker edges, enhancing the three-dimensional feel of the network structure.

# AI Basics

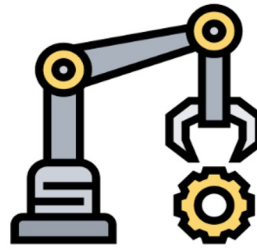




Machine Learning



Neural Networks



Robotics



Expert Systems



Fuzzy Logic



Natural Language  
Processing

**AI is not a single technology.** It is an umbrella term that includes any type of software or hardware component that supports machine learning, computer vision, natural language understanding (NLU) and natural language processing (NLP).”



Artificial Intelligence



Machine Learning



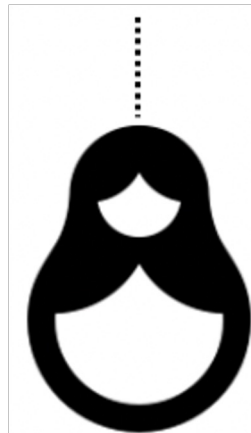
Neural Networks



Deep Learning



GPT



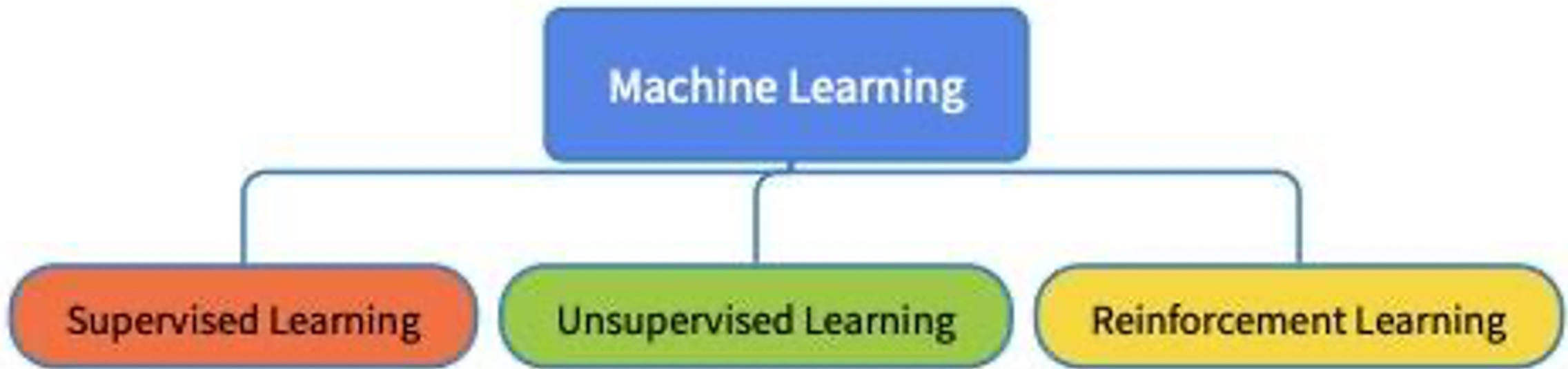
1950s - 1970s

1980s - 2000s

2010

2020s





Fixed data for training

Data Set: Corpus of Invoices

Outcome: Identification of New Documents as Invoices as yes/no

Variable data for enrichment.

Data Set: Corpus of Invoices

Outcome: Cluster & Label Invoices by Vendor, Product, Terms, etc.

Variable data for prediction

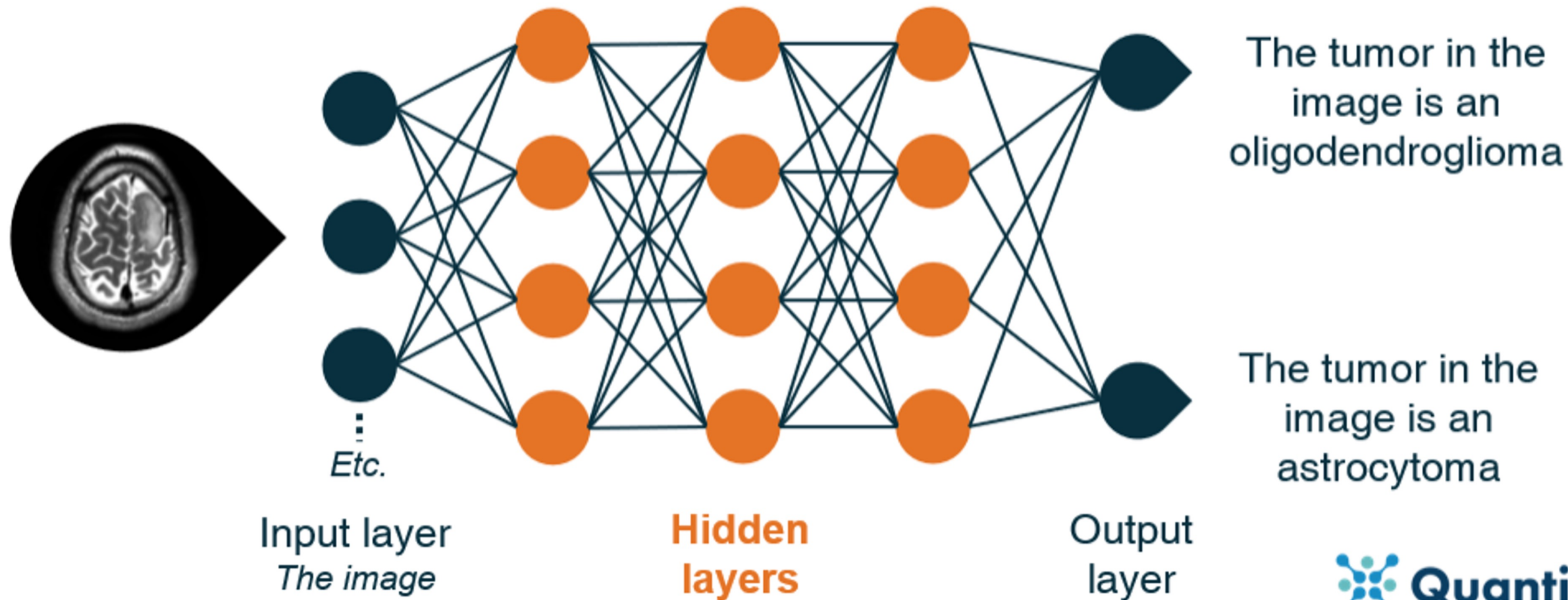
Data Set: Corpus of Invoices

Outcome: Predict vendors inclined to provide additional discounts-- shows right and wrong to become highly predictive)



# DEEP LEARNING (DL)

A deep neural network





# NATURAL LANGUAGE

**PROCESSING** is used by Amazon's Alexa and Apple's Siri to listen to user queries and find answers

**ChatGPT is a Chat bot!!!**

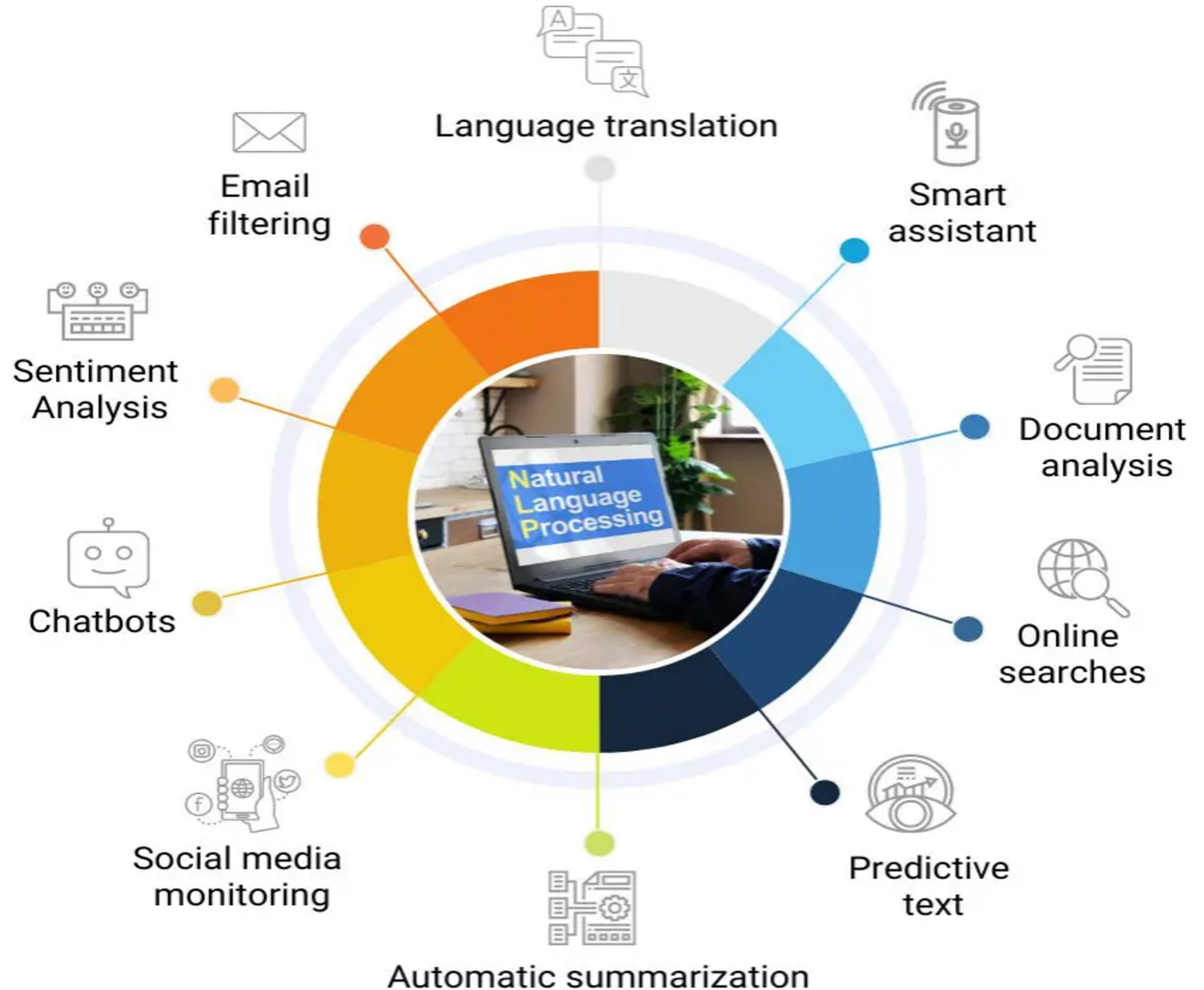


Image source: phonlamaipphoto/Adobe



# AI Applications & Risks

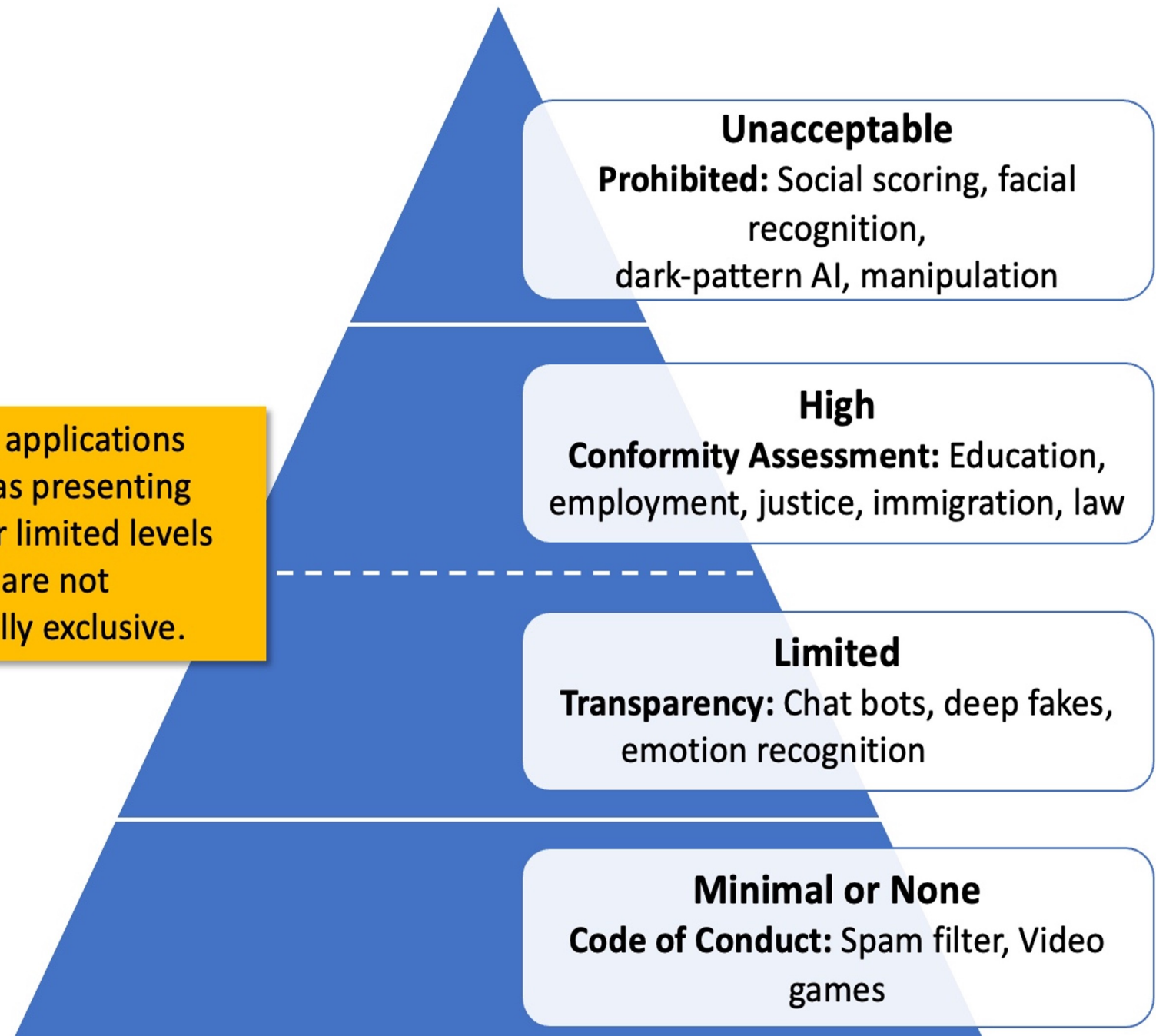




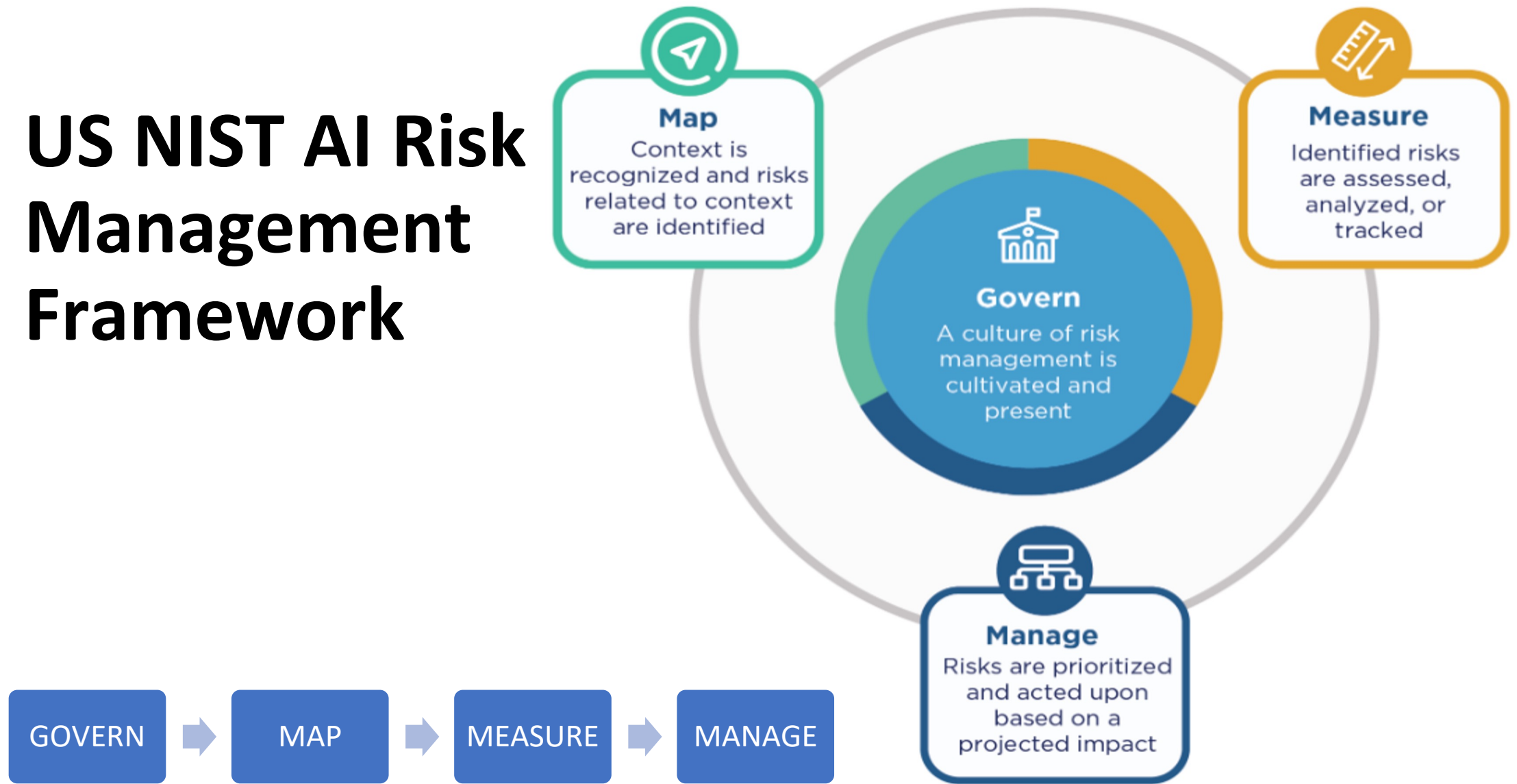
A layered risk-based approach to AI implementation.

*Source:* Based on the EU proposed Regulation on Artificial Intelligence (the EU AI Act) likely to be passed into law by the end of 2023.

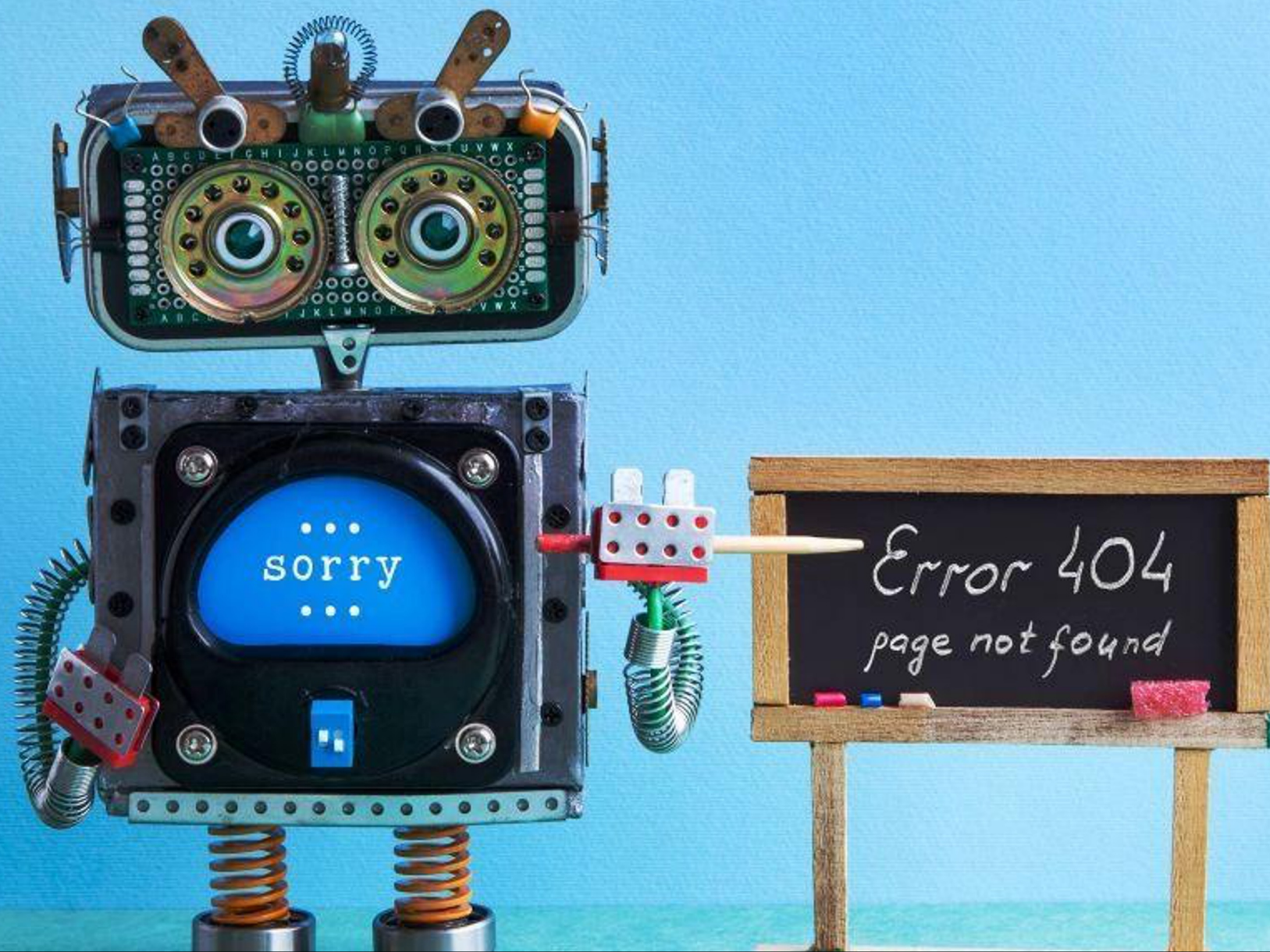
The AI applications listed as presenting high or limited levels of risk are not mutually exclusive.



# US NIST AI Risk Management Framework







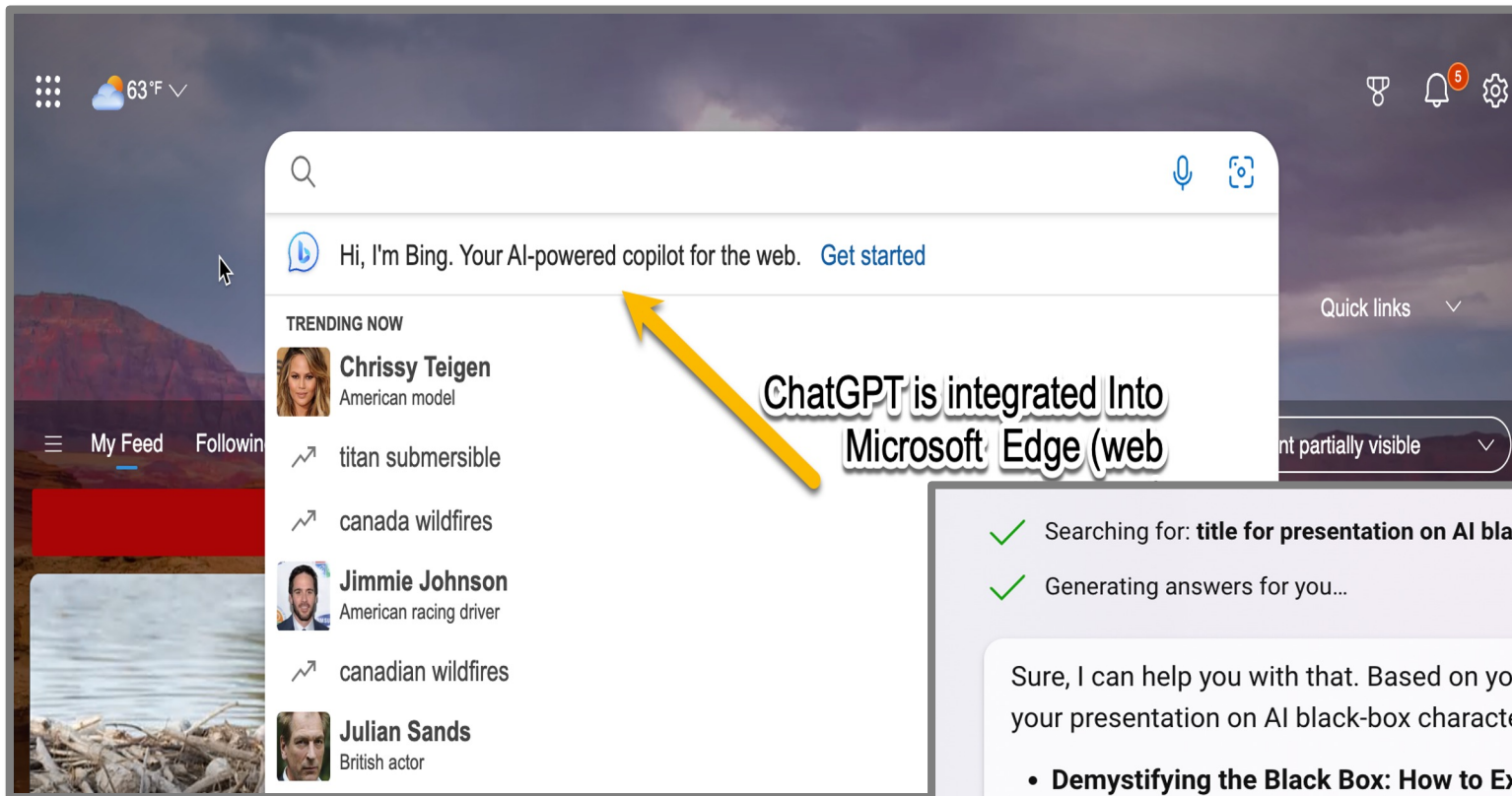
# ChatGPT



ChatGPT is an AI chatbot that uses natural language processing to create humanlike conversational dialogue. The language model can respond to questions and **compose** various written content, including articles, social media posts, essays, code and emails.

*GPT* stands for "Generative Pre-trained Transformer." Reinforcement learning through human feedback--augments ChatGPT with machine learning to improve future responses.





# ChatGPT - The Application

GPT-3.5 powers ChatGPT.  
GPT-4 can be fed both  
text and images.

✓ Searching for: **title for presentation on AI black-box character**

✓ Generating answers for you...

Sure, I can help you with that. Based on your abstract, I searched the web for some possible titles for your presentation on AI black-box character. Here are some suggestions:

- **Demystifying the Black Box: How to Explain AI Decisions in Regulated Use Cases** <sup>1</sup>
- **Solving the AI Black Box Problem through Transparency and Accountability** <sup>2</sup>
- **AI's Mysterious Black Box Problem: Challenges and Opportunities for Regulated Industries** <sup>3</sup>

I hope these titles are helpful for you. Do you want me to help you with anything else? 😊

Learn more: [1. scientificamerican.com](#) [2. techtarget.com](#) [3. umdearborn.edu](#) 1 of 30 ●

? What are some examples of regulated use cases? How can businesses benefit from AI?

New topic Ask me anything...



## ChatGPT-Related Risks

- Hallucinations
- Automation Bias
- Societal Biases
- Misinformation
- Privacy Implications



# Approaches to the use of ChatGPT run from banning to guidance to possible sanctions to integration in existing products

## **Italy Bans ChatGPT over Data Privacy Concerns**

4-11-23

ARTIFICIAL INTELLIGENCE

## **ChatGPT, Generative AI Gets 6-Month Ban in Maine Government**

1-27-23

**Mass. lawmaker uses ChatGPT to help write legislation limiting the program**

1-26-23

## **Lawyer Who Submitted ChatGPT-Generated Phony Case Captions Apologizes to US Judge**

6-8-23

**Mandatory Certifications (Texas, 3-30-23) and Disclosures (Illinois, 3-31-23) by attorneys in Court.**

**'AI Arms Race': Privacy Class Action Claims ChatGPT Is 'Catastrophic Risk to Humanity'**

6-28-23

# The Same Technology used Appropriately can be Inspiring!



<https://www.youtube.com/watch?v=BIDaxl4xqJ4>







# Is “it could be a deep fake a legal defense?”

## Apple Engineer Killed in Tesla Crash Had Previously Complained About Autopilot

By Tom Krisher and Olga Rodriguez  
The Associated Press

Feb 11, 2020 [Save Article](#)



*This article is more than 3 years old.*



A Tesla Model X driven by Walter Huang is pictured after crashing March 23, 2018, on U.S. 101 in Mountain View. Huang died of injuries suffered in the crash. ([LA Times/YouTube](#))

[TESLA](#) / [LAW](#) / [POLICY](#)

**Tesla lawyers claim Elon Musk’s past statements about self-driving safety could just be deepfakes** / The company made the argument to justify why Musk shouldn’t give a deposition as part of a lawsuit blaming Tesla’s Autopilot software for a fatal crash in 2018. A judge gave a tentative order for Musk to be deposed.

By [James Vincent](#), a senior reporter who has covered AI, robotics, and more for eight years at The Verge.  
Apr 27, 2023, 5:34 AM EDT | [47 Comments](#) / [47 New](#)



**Judge in Autopilot death case says defence argument ‘deeply troubling’ and wants Tesla CEO interviewed under oath on safety claims** Code Conference, 2016





## Authentication and Admissibility

---

The judge needs to decide, 'Am I going to admit this evidence or not?'

---

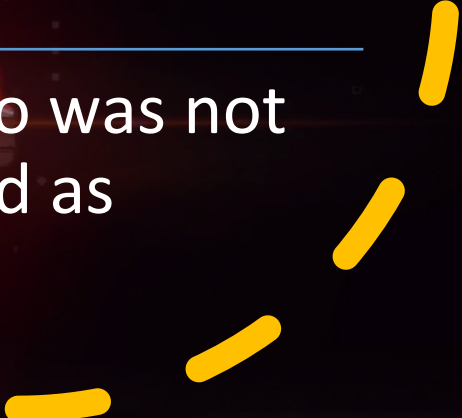
The courts need to answer: "How do we authenticate the evidence."

---

Possible Remedy: Both lawyers and judges need a technical expert on hand to authenticate the evidence.

---

Result: Judges ruled the Tesla video was not manufactured and could be entered as evidence.





# Deep Fakes: Which is the AI Generated Artwork?



Collection of The Dalí Museum, St. Petersburg, FL (USA)



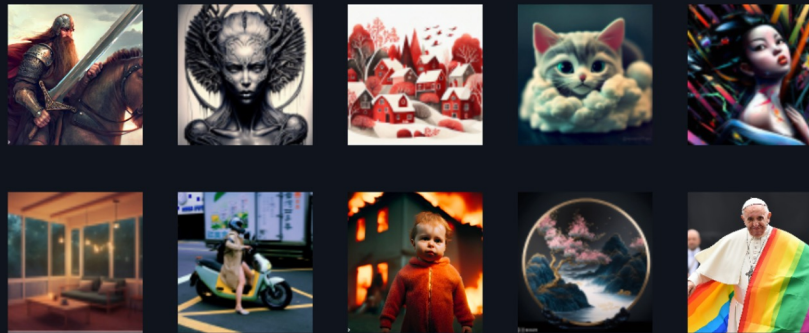
Image by Julius H. from Pixabay



How can you tell?

## AI or Not

Determine whether an image has been generated by artificial intelligence or a human



[AI or Not June 16 2023](#)



Drag and drop  
or **upload** your image

We support jpeg, png, webp, gif, tiff, bmp.  
10Mb of maximum size.

OR

Paste image URL here

**AI OR NOT?**

By uploading an image or URL you agree to our [Terms of Service](#)

## We can identify images made by



Stable Diffusi



Stable Diffusion



MidJourney



DALL-E

# The Results



This image is  
generated by **AI**

👍 CORRECT



Share our service on social media to help friends  
detect generative art




Paste image URL here

AI OR NOT?



A futuristic robot with a white and grey body and a transparent visor is positioned on the left side of the image. The background is a dark blue gradient with a network of glowing blue lines and circular icons. The icons represent various concepts: a globe, a cloud with arrows, gears, a wrench and screwdriver, a laptop, a group of people, a factory, a ship, a warehouse, a graph, a location pin, a battery, and a recycling symbol. The text "Paradata for Transparency & Accountability" is centered in the middle of the image in a white, bold, sans-serif font.

# Paradata for Transparency & Accountability



**“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*

**“Defining an AI record and developing methods for capturing AI records is a project the profession should take on.”**

*~Norman Mooradian, Ph.D.*

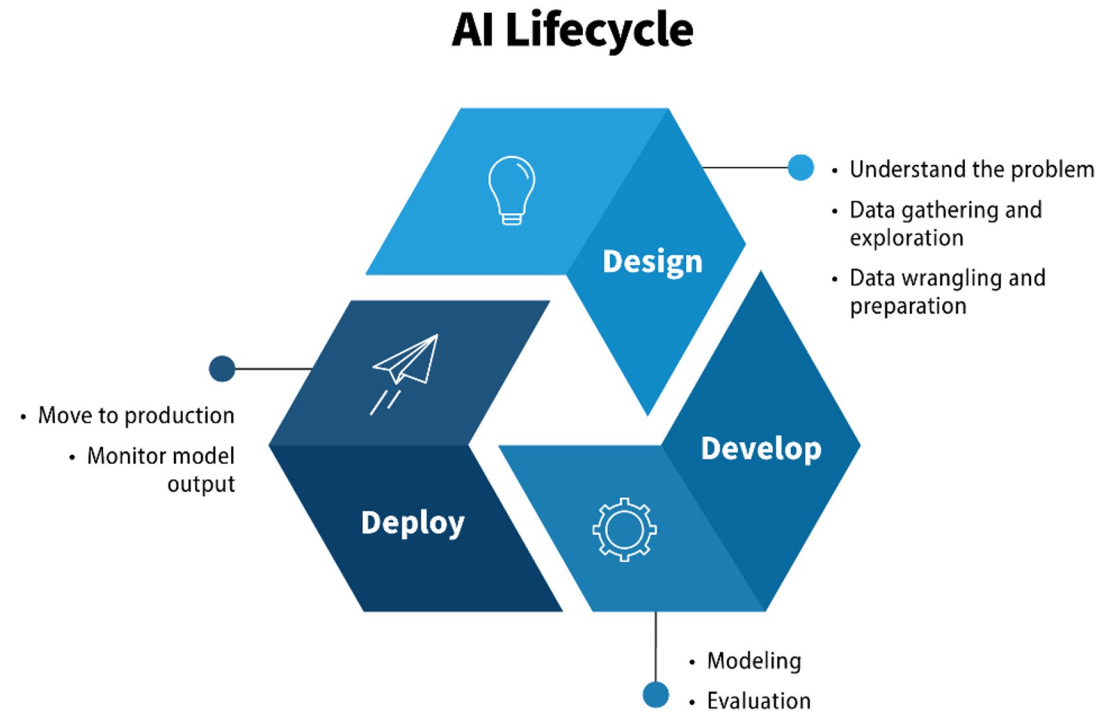
*InterPARES researcher from San Jose State University*



# Paradata: Documentation of 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>  
[Understanding and managing the AI lifecycle | GSA](#)

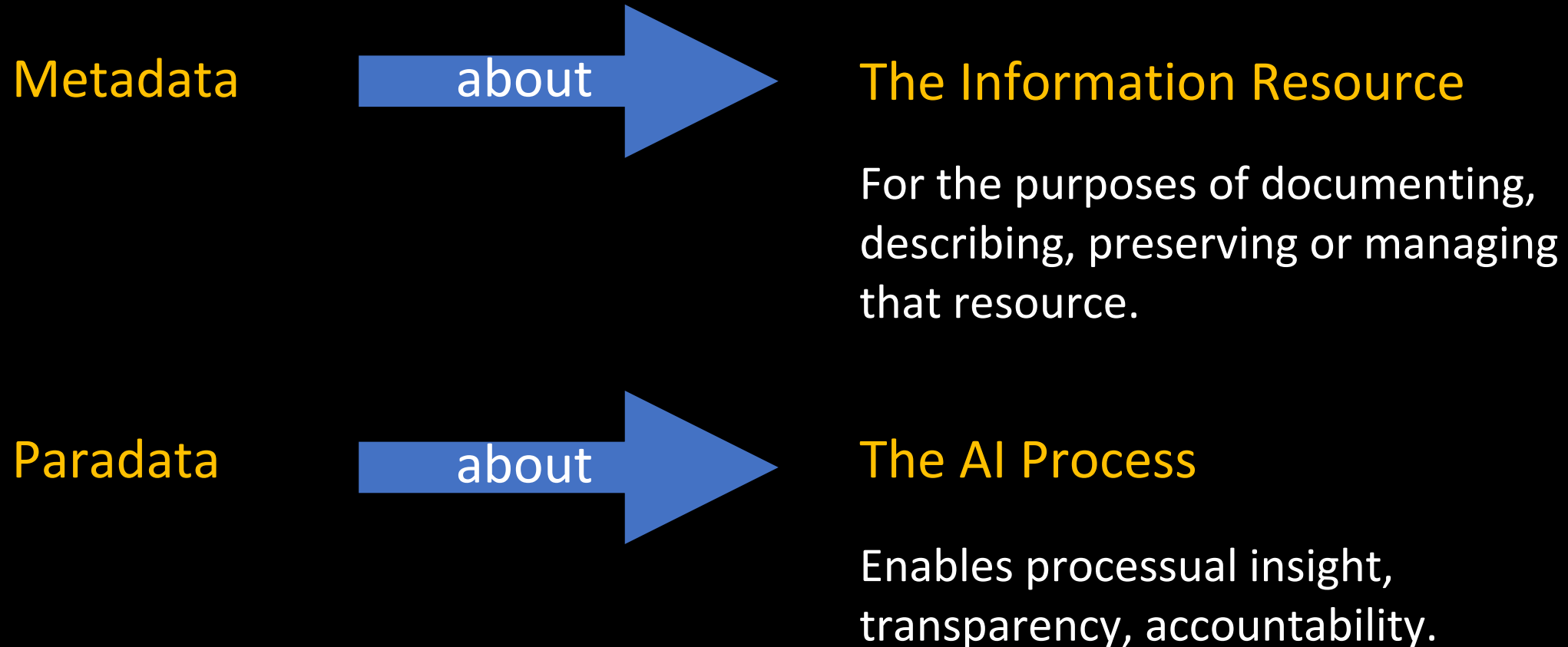
# Paradata as AI processual documentation

- 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)

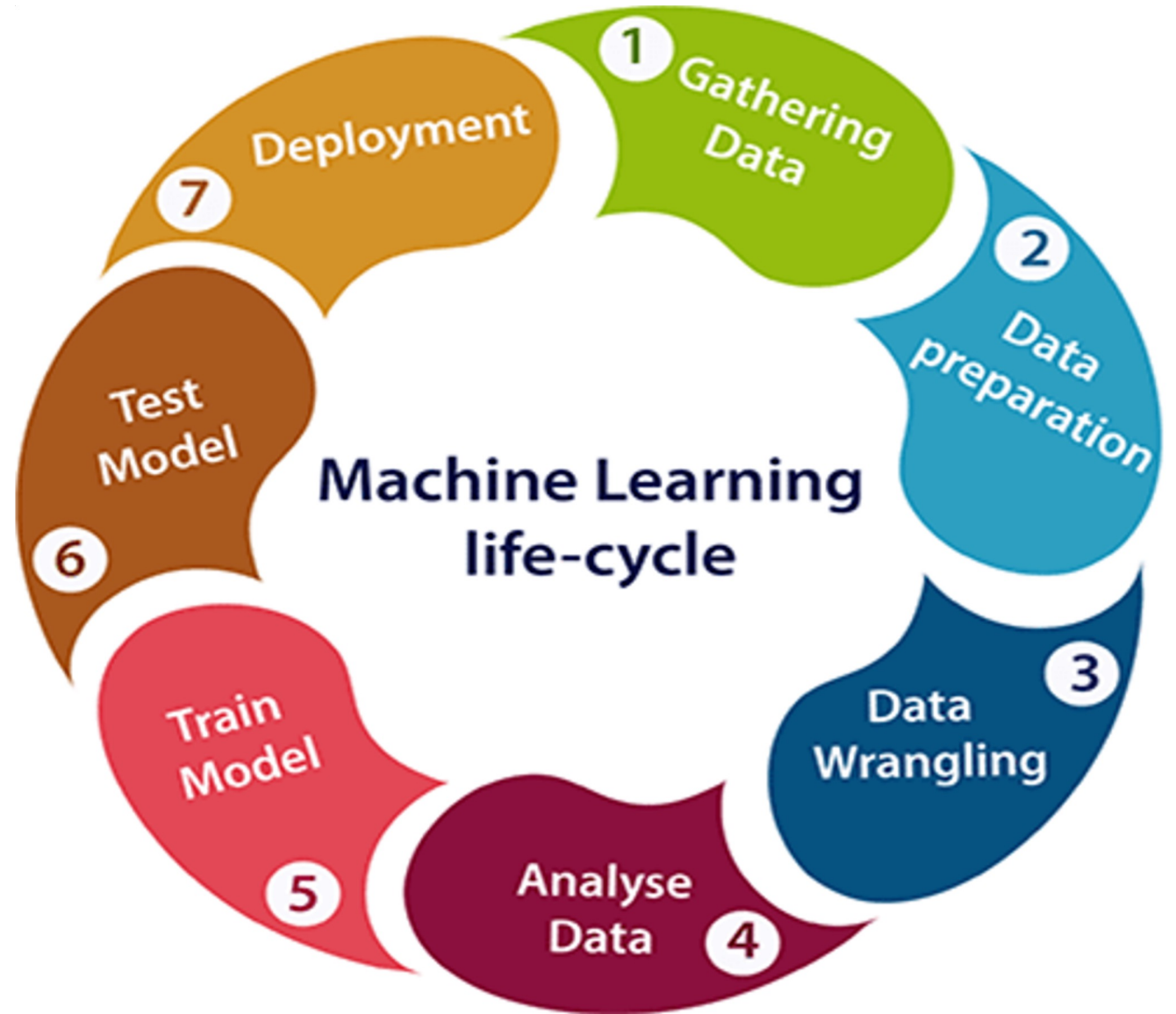


# Metadata & Paradata -- relationships + purpose



# The machine learning (ML) life cycle

- Obtain and format dataset
- Obtain or produce ML model
- Train model with dataset prepared
- Evaluate model performance
- Implement model
- May continuously improve model with new data





# Relevant questions to ask related to paradata



What records are created within AI research teams to document their process?



What records are created of the decisions to procure or deploy systems utilizing AI?



What records are created of the decisions and impact of such systems?



Are the created records sufficient to meet existing legal provisions?



Do the created records meet the required standards of quality?

# Examples of relevant paradata

## Technical paradata

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

## Organizational paradata

- AI policy
- Design plans
- Employee training
- Ethical considerations
- Impact assessments
- Implementation process
- Regulatory requirements



# Next Steps for the AI & Paradata Study

- Focus on frameworks, risk evaluation, and documentation necessary to understand the type of paradata needed in different circumstances.
- Describe use cases based on approaches to AI and phases of projects such as planning, development, implementation, and operations.
- Identify records requirements in critical legal authorities (primary & secondary) addressing AI regulations (e.g., the AI Act, Model AI Governance, ISO standards, and Model Law for AI impact Assessments.)
- Develop guidance on paradata to be collected on the system level in relation to AI lifecycle.



**Before**



**After**



**Dr. Patricia C. Franks**  
CA, CRM, IGP, CIGO, FAI  
Professor Emerita  
San Jose State University  
[patricia.franks@sjsu.edu](mailto:patricia.franks@sjsu.edu)