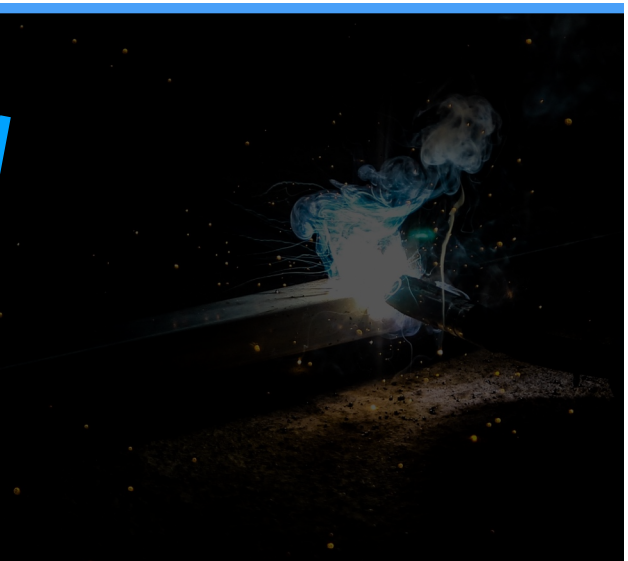
AI Landscape and Scientific Assessments



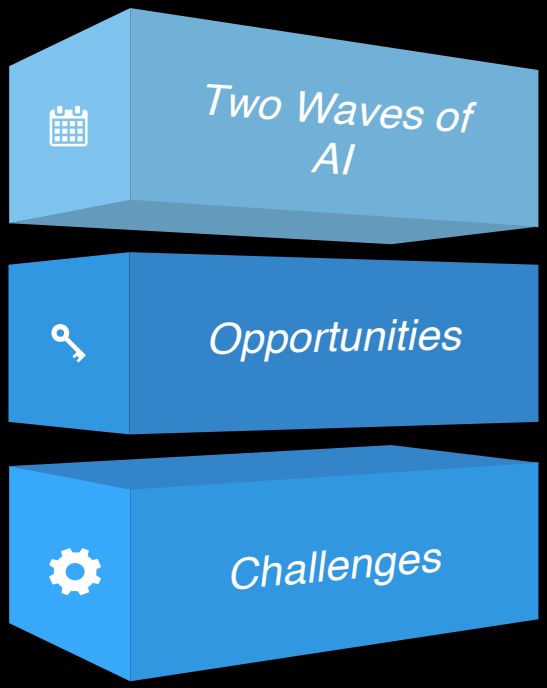
Heng Xu, PhD

Director, Center for AI, Cyber Governance & Privacy Management
Professor, Warrington College of Business, University of Florida

heng.xu@ufl.edu

robustanalytics.org/xu

linkedin.com/in/hengxu/



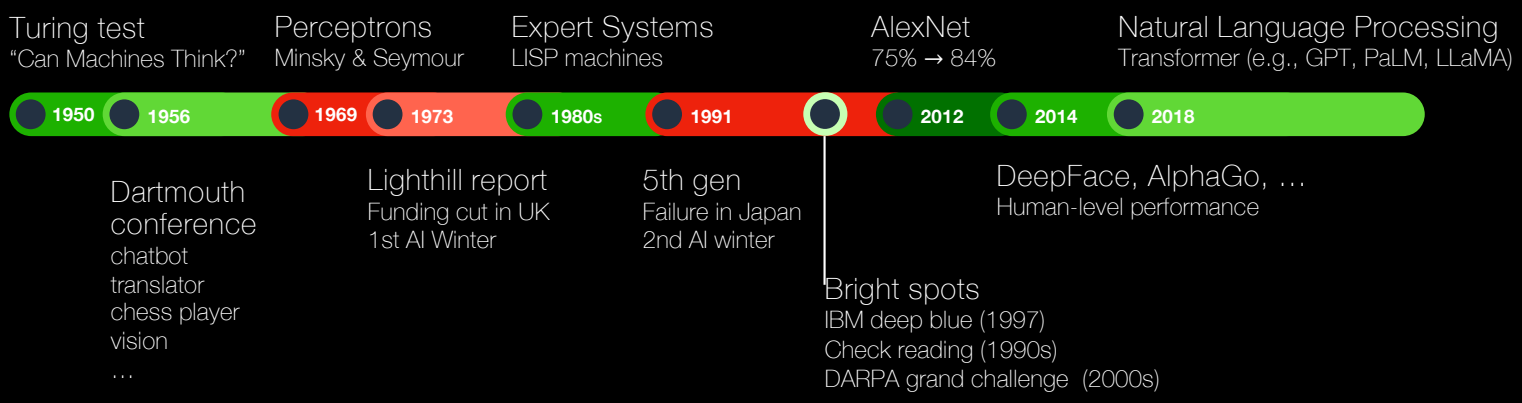
Symbolic AI → Machine Learning

Enabling Previously Impossible

Responsible AI for a Better Future

Robust Analytics Lab

What is AI?



{ Artificial Intelligence, AI, is the technology that enables machines to simulate human intelligence. }

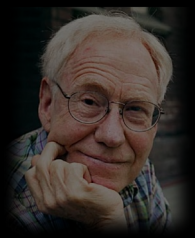
AI

1958

Herbert A. Simon:

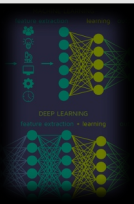
We can expect AI chess champion and theorem prover in 10 years because “A physical symbol system has the necessary and sufficient means for general intelligent action”

1965



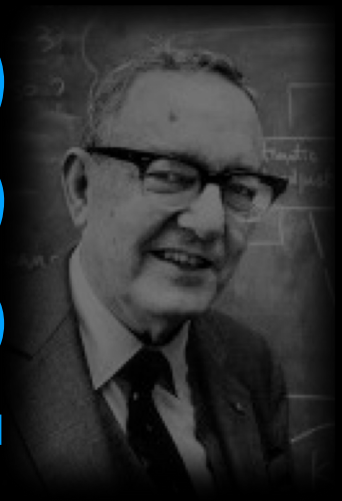
1988

2010



AI

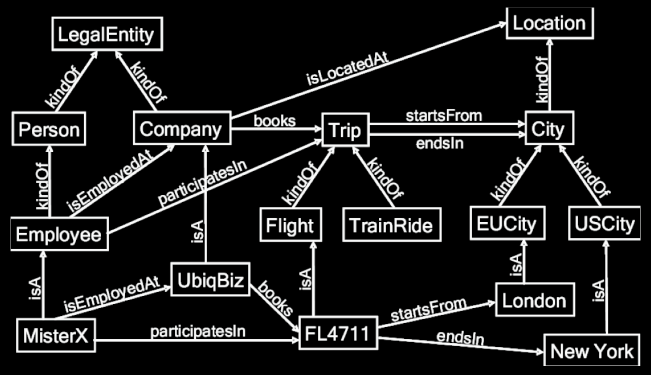
1958

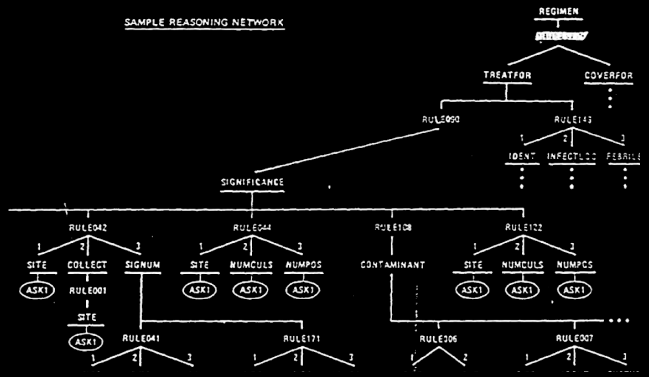


AI chess champion + theorem prover in 10 years thanks to ***Symbolic AI*** *where **RULES** are first-order drivers*

"A physical symbol system has the necessary and sufficient means for general intelligent action"

Symbolic AI was the dominant paradigm of AI research from the mid-1950s until the middle 1990s, which is based on high-level symbolic (human-readable) representations of problems, logic and search.





Newel, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. Communications of the ACM, 19(3), 113-126.

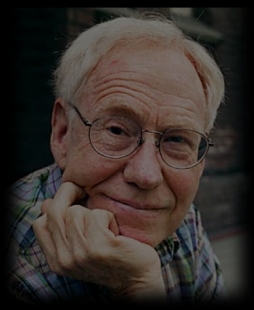
AI

1958

1965

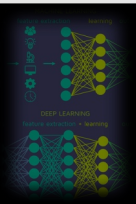
Hubert L. Dreyfus' critique of AI research:

"Our everyday coping cannot be understood in terms of rules or in terms of responses to fixed features of the context. Instead, it takes place on 'the background of ... primary familiarity, which itself is not conscious and intended but is rather present in [an] unprominent way'"



1988

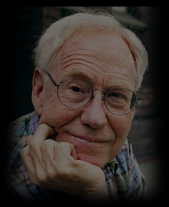
2010



AI

1958

1965

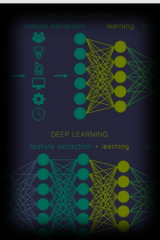


1988

Hans Moravec's paradox:

"it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility"

2010

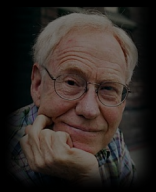


AI built via handcrafted rules couldn't learn a new rule on their own: it broke when it encountered a weird situation.

AI

1958

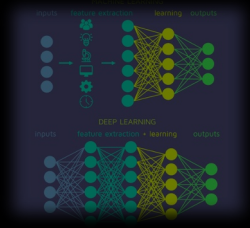
1965



1988

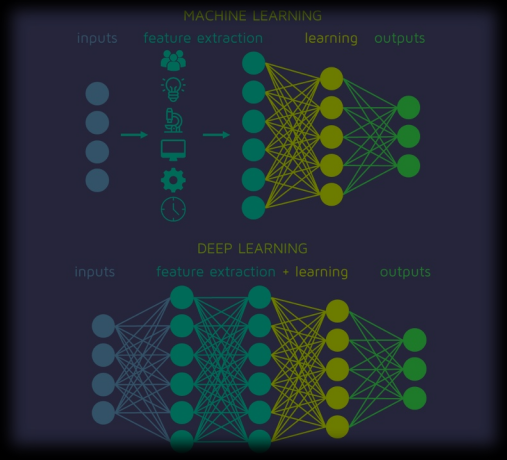
2010

Statistical machine learning
(especially deep learning) took off



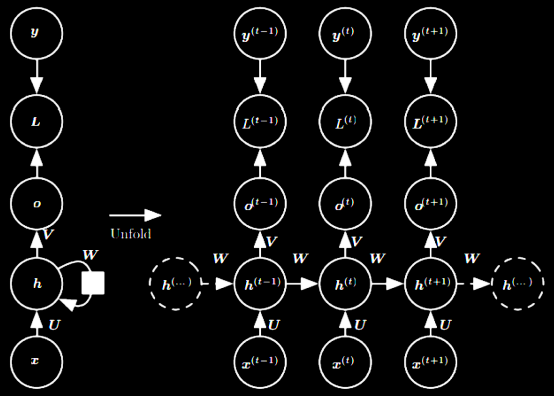
AI

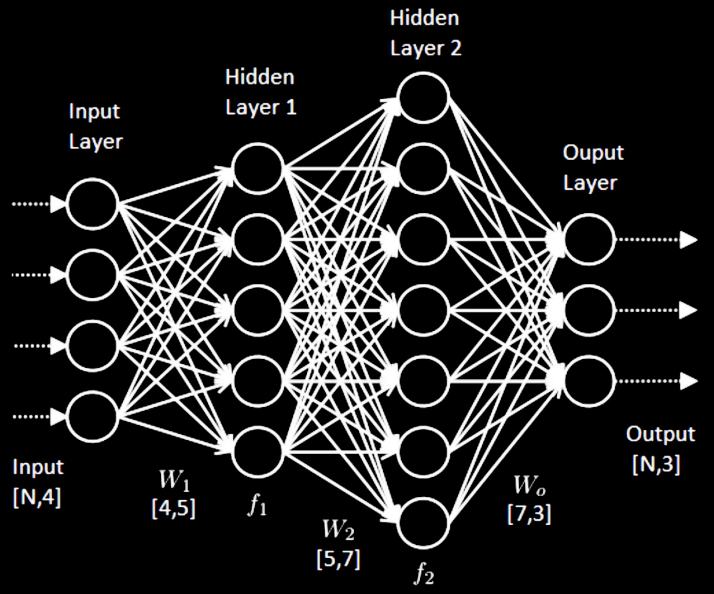
2010



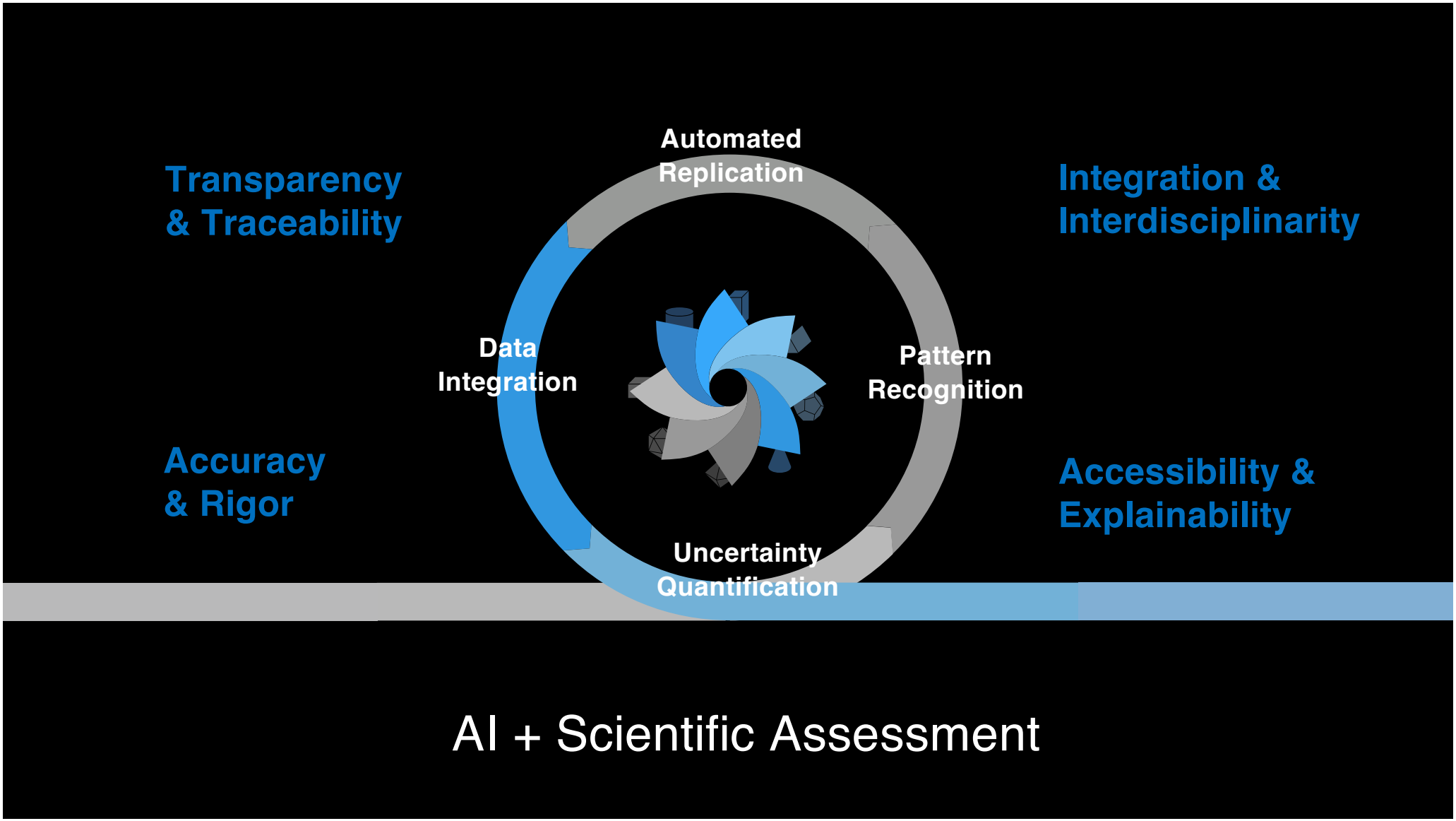
Statistical Machine Learning
where ***RULES*** become second-order approximations

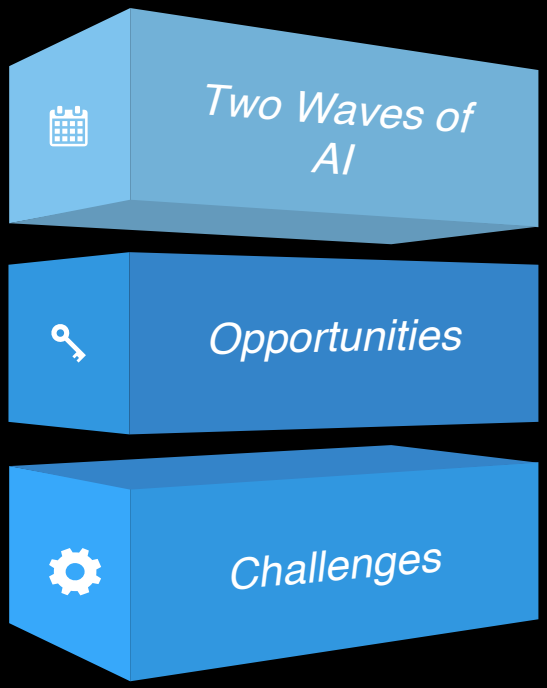
"*deep nets are known to be universal approximators, capable of representing arbitrarily complex functions given sufficient capacity*" (Arpit et al., 2017)





Arpit, D., Jastrzębski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M. S., ... & Lacoste-Julien, S. (2017). A closer look at memorization in deep networks. In International Conference on Machine Learning (pp. 233-242).


Transparency
& Traceability
Automated
Replication
Integration &
Interdisciplinarity
Data
Integration
Pattern
Recognition
Accuracy
& Rigor
Uncertainty
Quantification
Accessibility &
Explainability
AI + Scientific Assessment



Symbolic AI → Machine Learning

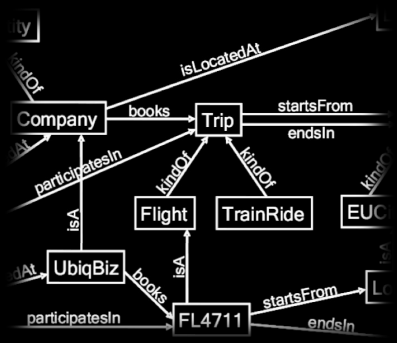
Enabling Previously Impossible

Responsible AI for a Better Future

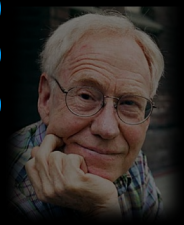
Robust Analytics Lab

AI

1958

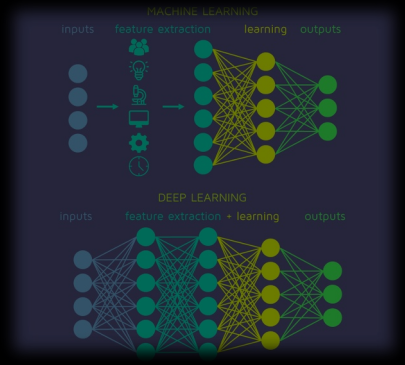


1965



1988

2010



Symbolic AI ⟶ Statistical Machine Learning

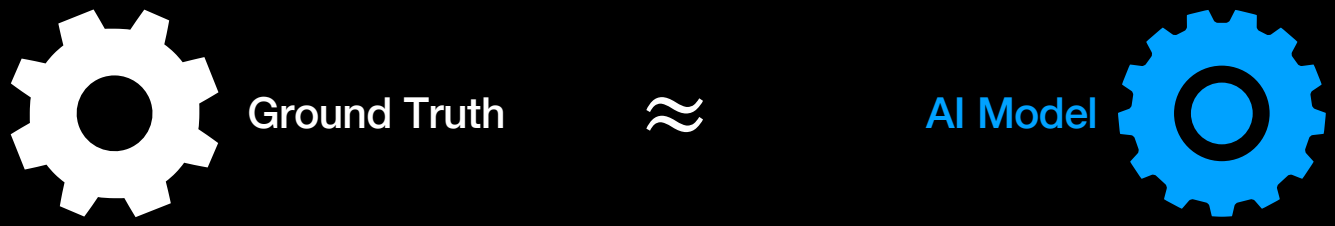
where rules are first-order drivers

where rules are second-order approximators

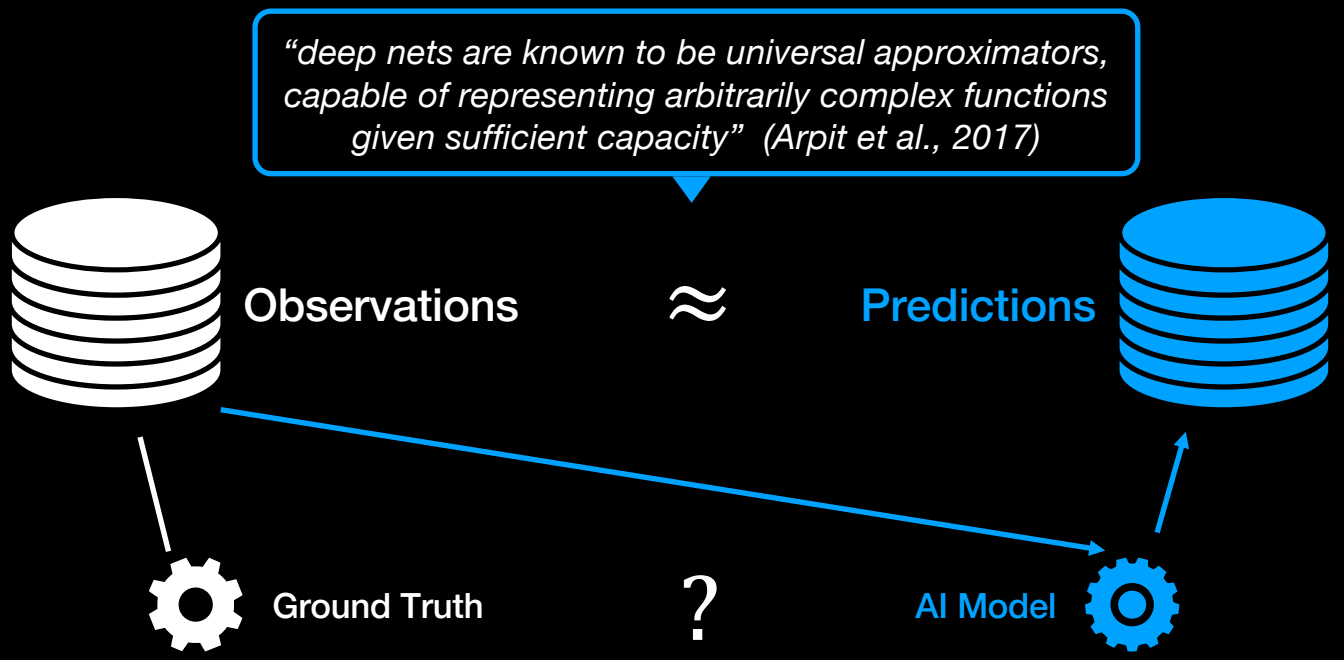
{ Rule-Based AI }

{ Black-Box AI }

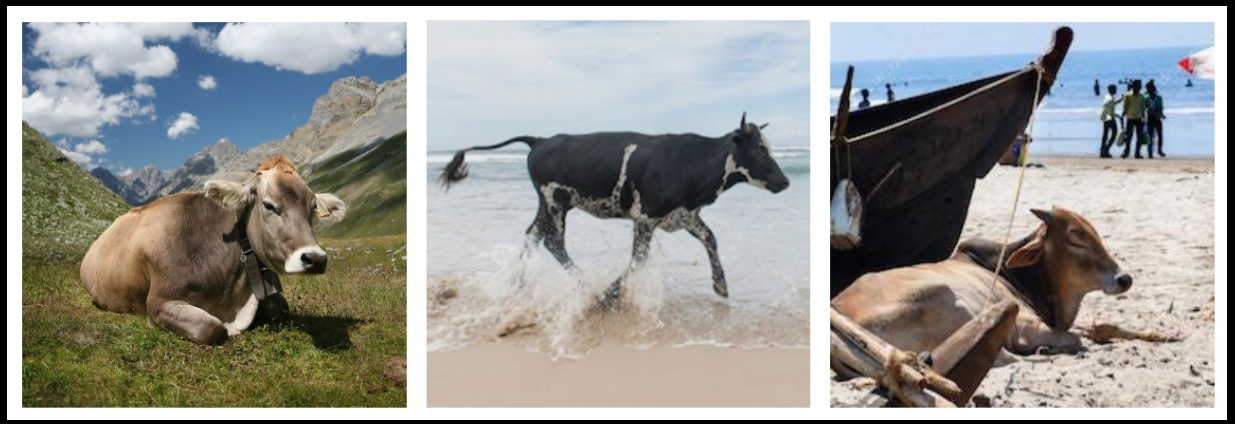
Back Then: The Objective of Symbolic AI



Now: The Objective of AI Today



Robust Analytics Lab



clarifai

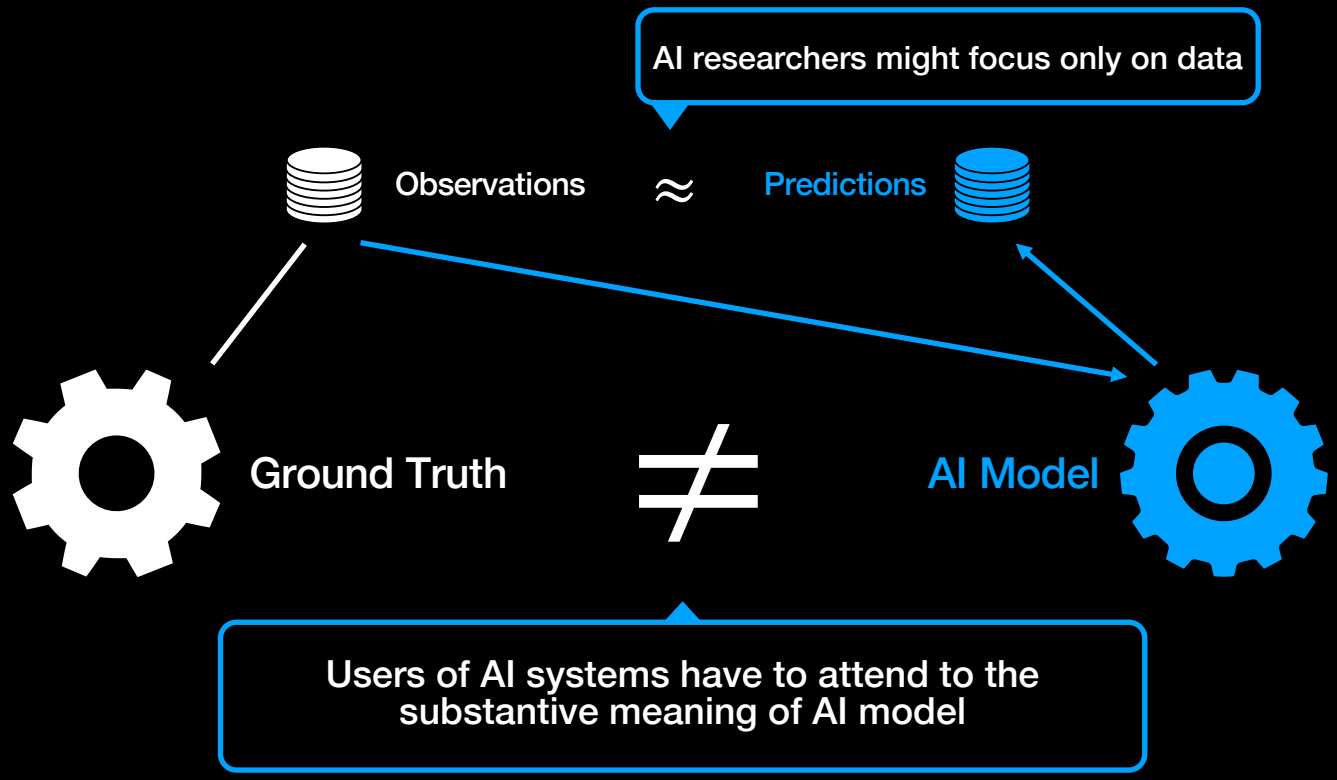
Cow: 0.99
Pasture: 0.99
Grass: 0.99
No Person: 0.98
Mammal: 0.98

No Person: 0.99
Water: 0.98
Beach: 0.97
Outdoors: 0.97
Seashore: 0.97

No Person: 0.97
Mammal: 0.96
Water: 0.94
Beach: 0.94
Two: 0.94

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in terra incognita. In Proceedings of the European conference on computer vision (ECCV) (pp. 456-473).

AI Today



Fundamental challenge for AI governance

Robust Analytics Lab

How is my data used by AI?

Privacy & Transparency

Why does Amazon's AI hiring tool show bias against women?

Fairness & Inclusiveness

How to ensure the outputs of AI models are trustworthy and useful?

Trust & Robustness

How do we state what we want from AI systems?

Specification Problem

Current Research Portfolio

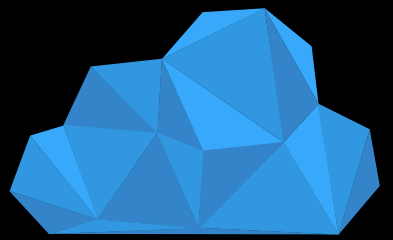
Robust Analytics Lab

Human-AI Teaming

"Specification Problem":

the problem of defining the behavior that we would want to see from AI systems

How do we state what we want?



Xu, H., and Zhang, N. (in press). Goal Orientation for Fair Machine Learning Algorithms, *Production and Operations Management*. https://journals.sagepub.com/doi/abs/10.1177/10591478241234998

ChatGPT

Draw roaring moon in space, surrounded by rocks.

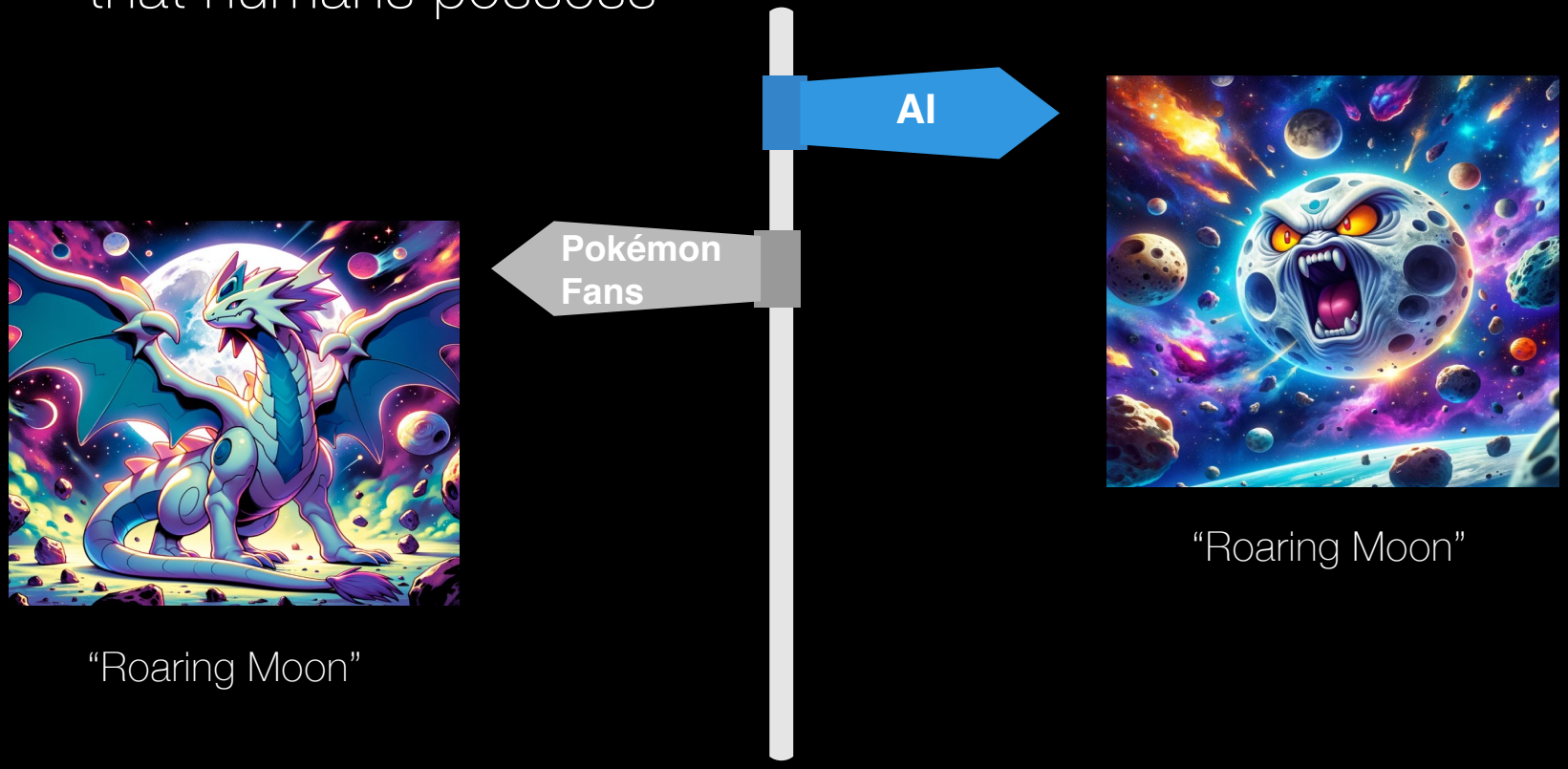
ChatGPT4 Output as of Jan 28, 2024

ChatGPT

This does not look like roaring moon in Pokémon. Roaring moon is a dragon in Pokémon. Can you make it more like a dragon?

ChatGPT4 Output as of Jan 28, 2024

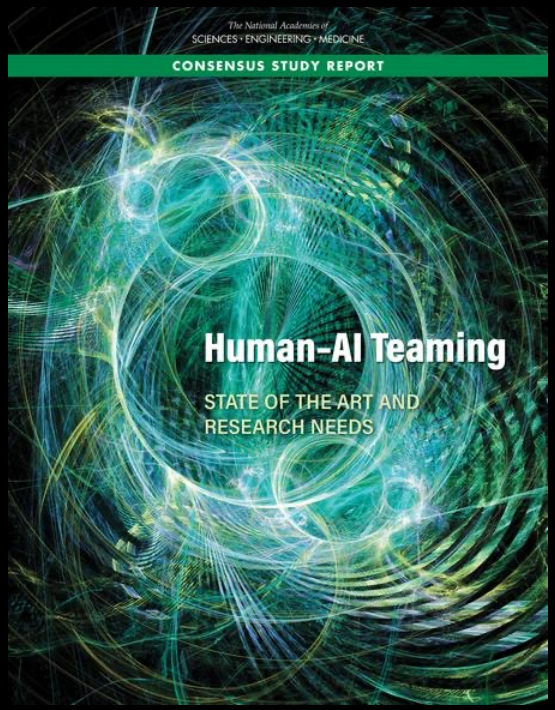
AI systems lack the common sense background that humans possess



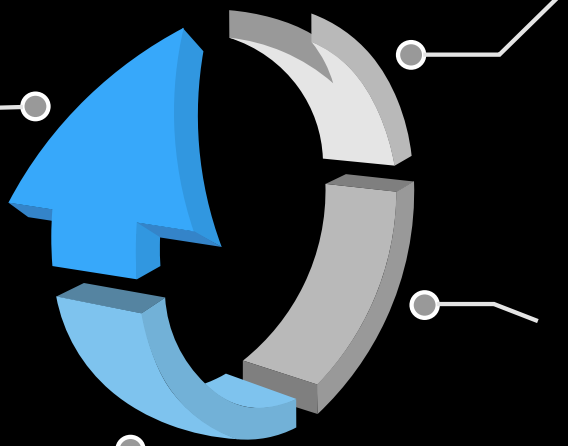
“Roaring Moon”

“Roaring Moon”

To effectively tackle the "specification problem," it is essential for domain experts to collaborate closely with AI system designers.



New Science of Human-AI Teaming (Task/Trust/Collaboration)



New Models and Metrics for Human-AI Performance Evaluation

Ensure the Safety and Security of AI Systems

Address the Ethical, Legal, and Societal Implications

thank you.

heng.xu@ufl.edu

robustanalytics.org

linkedin.com/in/hengxu/