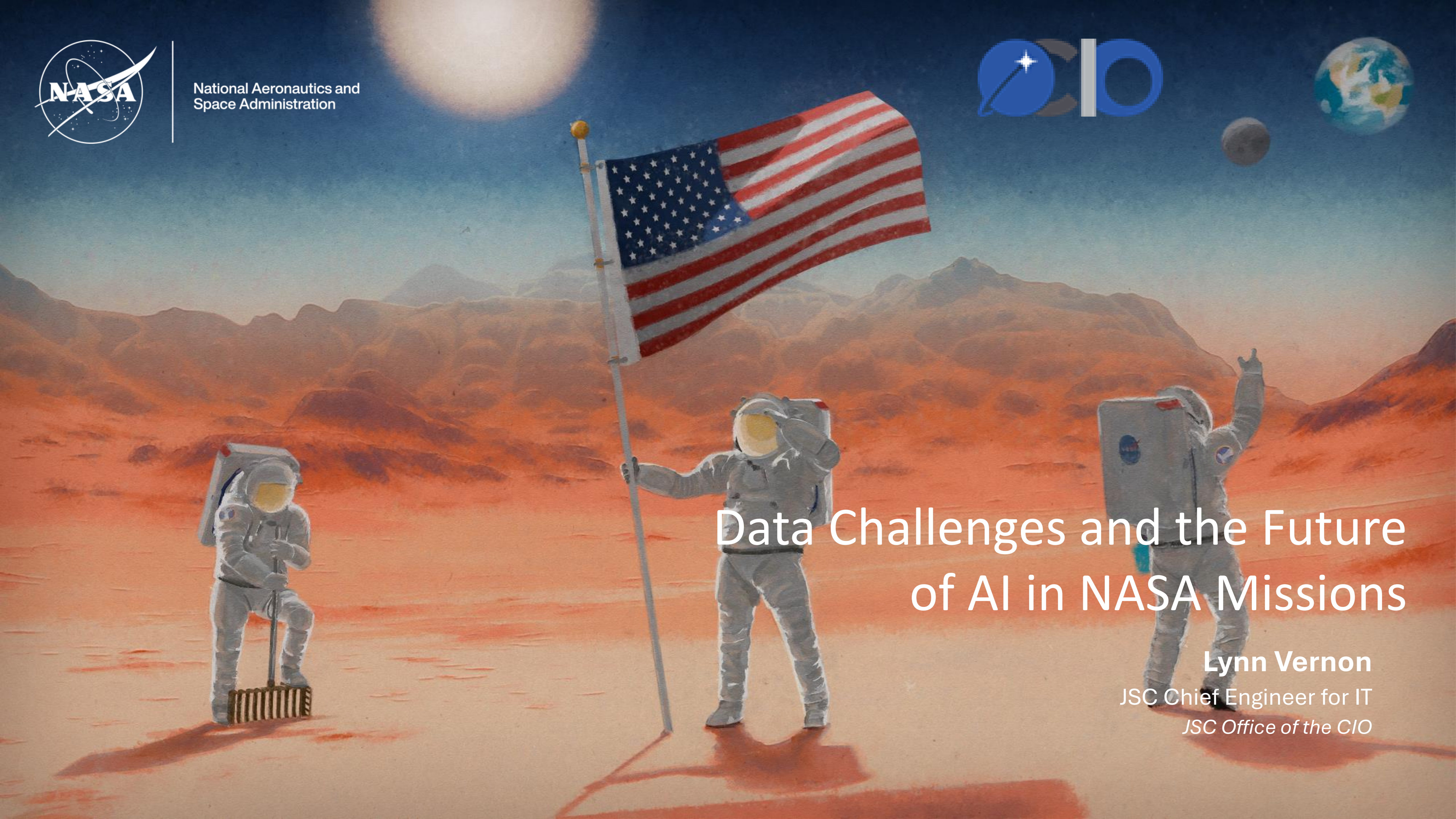


National Aeronautics and Space Administration

Data Challenges and the Future of AI in NASA Missions

Lynn Vernon
JSC Chief Engineer for IT
JSC Office of the CIO

NASA Vision

- “To discover and expand knowledge for the benefit of humanity”

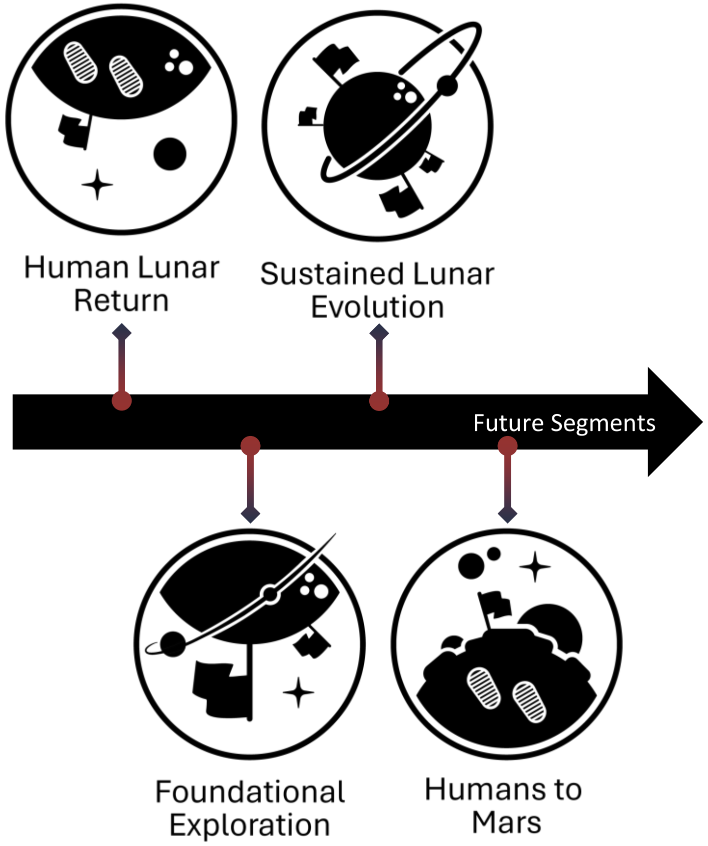
NASA Space Act

- The NASA Space Act, originally enacted in 1958, provides the framework for NASA's operations and partnerships.
- Regarding data, the Act emphasizes the importance of sharing and disseminating information. Here are some key points:
  - Public Availability: NASA is required to make its scientific and technical information widely available to the public. This includes data from its missions and research.
  - Partnerships: The Act allows NASA to enter into agreements with various entities, including private companies and international organizations, to share data and collaborate on projects.
  - Geospatial Data: Recent agreements, like the one with Esri, enhance the use of NASA's geospatial data for research and exploration.
- These provisions ensure that NASA's data contributes to scientific advancement and is accessible for various applications.

We Are Going to the Moon To Prepare for Mars

- Artemis, named after the **twin sister of Apollo** who is also the Goddess of the Moon and the hunt, encompasses all of our efforts to return humans to the Moon – which will prepare us and propel us on to Mars.
- Through the Artemis program, we will see the first woman and the next man walk on the surface of the Moon.

M2M Architecture Components

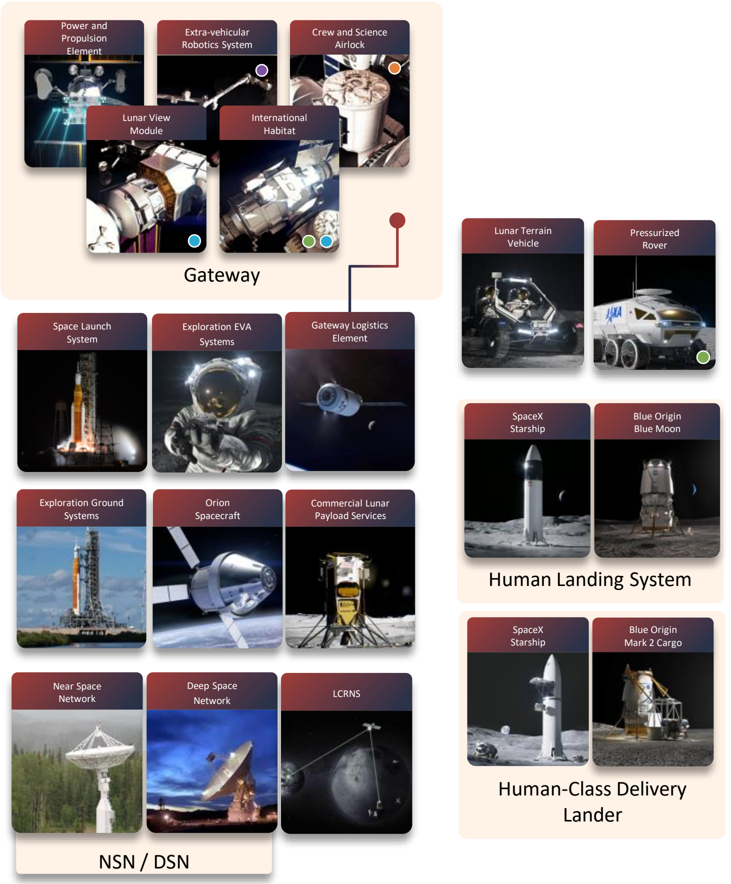


Segments

A portion of the architecture that integrates sub-architectures and progressively increases in complexity and objective satisfaction.

Sub-Architectures

A group of tightly coupled elements, functions, and capabilities that work together to accomplish one or more objectives.



Elements

A notional exploration system that enables a set of functions.

International Partnerships

- Canadian Space Agency
- European Space Agency
- United Arab Emirates
- Japan Aerospace Exploration Agency

Data Systems and Management

Sub-Architecture Definition

Data Systems
and Management

<u>Definition</u>: ***The group of capabilities that works together to <u>transfer, distribute, receive, validate, secure, decode, format, compile, and/or process data and commands</u> for use throughout the architecture. This includes future capabilities such as internet of things (IoT), cloud computing, servers, etc.***

The Data Systems and Management sub-architecture includes capabilities that work together to move, manage, and secure/protect data within acceptable latency constraints for use throughout the architecture. This sub-architecture is tightly coupled with the CPNT and human system sub-architectures to ensure data is shared and made useful across the architecture. Future capabilities may include data fusion, internet of things (IoT), cloud computing, and servers. The implementation of this sub-architecture spans from Earth systems, space, and planetary surfaces. Not all capabilities are expected to reside in-situ; each domain will include a mixture of assets.

Data systems and management play a pivotal role in modern information-driven landscapes. This area encompasses the intricate framework of tools, processes, representations, and technologies designed to capture, store, process, and retrieve data efficiently and securely. From small-scale payloads to large, complex mission sequences, effective data management across the Moon to Mars Architecture ensures that valuable insights can be derived from raw data, driving informed decision making and providing broad access as allowed.

These systems can consist of databases, their management systems, data warehouses, and data lakes that collectively organize and maintain data integrity. With the advent of big data, cloud computing, and advanced analytics, modern data systems not only handle structured information, but also embrace unstructured and semi-structured data formats. A robust data management strategy considers data availability, quality, interoperability, security, privacy, compliance, and access to ensure that we can harness the full potential of the expansive amount of lunar data, fostering innovation across the architecture.

Architecture Definition Document (ADD)
Revision A — Section 1.3.2.1

What does our Data look like?

Shift of How we do things!

- Historically focused on system centric
  - System controls
  - Data loaded and access through system

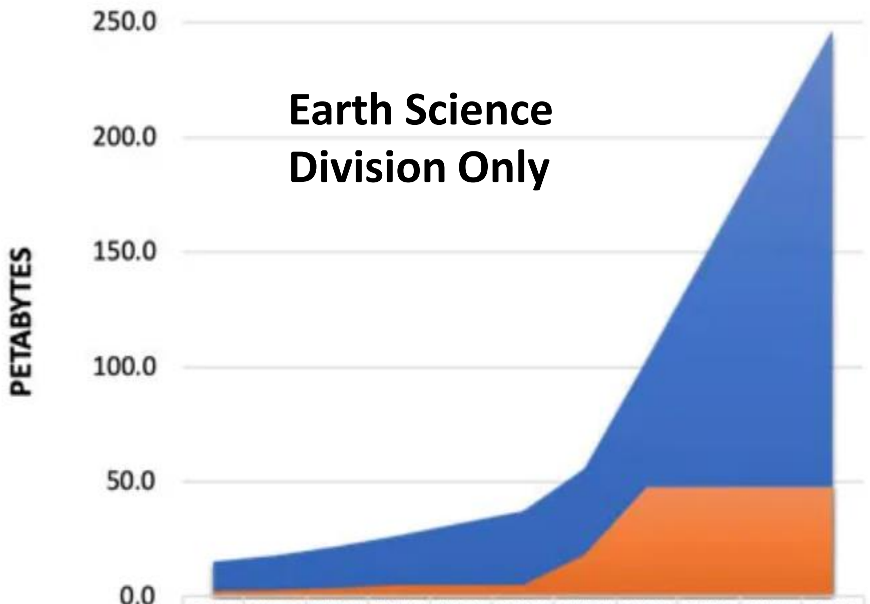
- Today, need to be data focused
  - What data do we have?
  - Where is the data?
  - Who can access the data?

WE ARE AWASH IN DATA

- Within 5 years, all four science divisions are cumulatively projected to generate over 100 PB of data per year and continue to grow rapidly as additional missions are launched and new models are run.
- Velocity & veracity are challenges
- If you feel there is a problem now, it will not get easier
- More data does not equal more insight

Data Growth:

2015 – 2025



	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Archive Size (PB)	15.0	17.7	21.6	26.8	32.0	37.2	55.6	103.4	151.1	198.9	246.6
Annual Growth (PB)	2.6	2.8	3.9	5.2	5.2	5.2	18.4	47.7	47.7	47.7	47.7

This chart shows the projected growth of the Earth Science Division database from 2015 until 2025. The orange area shows the order of magnitude increase. (Courtesy NASA's Earth Observing System Data and Information System).

https://www.marketplace.org/shows/marketplace-tech/nasas-data-is-headed-for-the-cloud/

1 Petabyte =
1,000 terabytes
a mile long beach—
100' wide , 1' deep

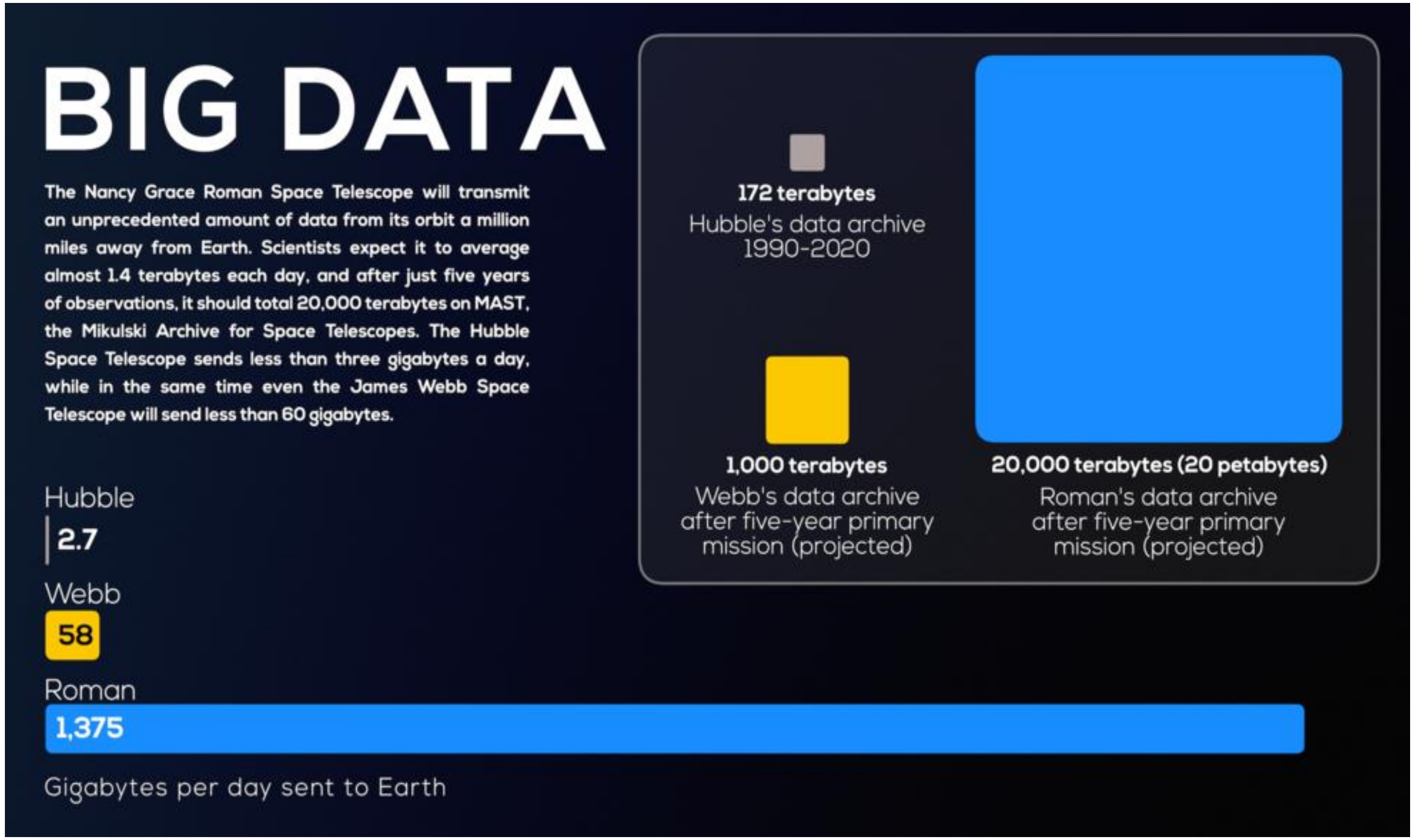
1 Exabyte =
1,000 petabytes
the same beach—
from Maine to North Carolina

1 Zetabyte =
1,000 exabytes
the same beach—
along the entire US coast

1 Yottabyte =
1,000 zetabytes
enough info to bury the entire
US under 296 feet of sand

History and Future of Spacecraft Data

- Apollo: 51.2 kbps (with 'unified' return link)
- Shuttle S: 192 kbps
- Shuttle K: up to 50 Mbps based on configuration
- Station K: Rated at 600 Mbps but using 300 Mbps today
- Station S: 23.6 Mbps

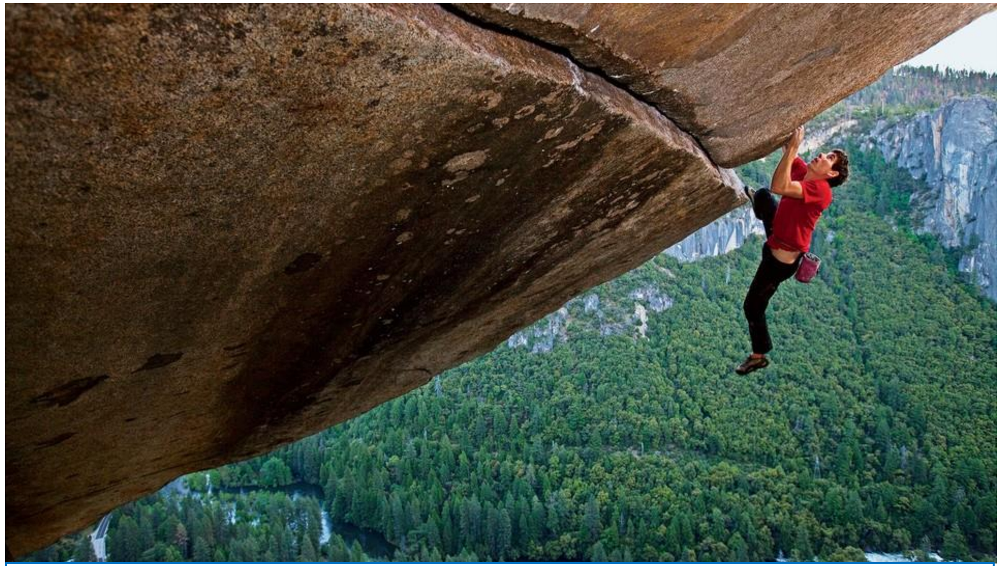


TRENDING DATA CHALLENGES → ORGANIZATIONAL RISK

At this year's Gartner Data & Analytics Summit, ATLAN surveyed over 600 data leaders. Their biggest priority, across industries and companies, is Data Governance.

With the explosive growth of AI, CIOs and CDOs are constantly being asked about their AI roadmap. The main hurdle isn't the AI models. It's the lack of AI-ready data, or data that is trustworthy, secure, and enriched with business context.

According to studies from IBM and AWS, the key obstacles hindering generative AI implementations are data privacy, data lineage, and data quality — all key elements of data governance.



The lack of Data Management and Governance could be viewed as free climbing with no harness or safety ropes

Data Management process discipline is a critical path dependency for responsible and transparent AI

TOP 10 BARRIERS TO IMPLEMENTING DATA GOVERNANCE

TOP FIVE (5)

- Lack of commonly understood approach across organization
- Lack of key roles
- Lack of skills / experience to drive initiative
- Inadequate funding (tools)
- Resistance to governance adoption from areas (bottom-up)

ROADMAP FOR CREATING VALUE FROM DATA

Data Disorder
- Same type of data means different things in different systems
- Example: Artemis is the same as Artemis 1 & Artemis Themis

Master / Ref. Data
- Publish and subscribe to master /reference data
- Ex: Single view of Program's across all information systems

BI/ML/Data Analytics
- Analyzing / Clustering data
- Looking for trends and correlations – machine learning
- Prescriptive, Predictive Analytics

PROCESSES: Data Discovery | Profiling | Data Governance | Data Products

Data Disorder → Defined Data → Master / Ref. Data → Documented & Governed Data → Smart Data BI/Analytics → AI Ready Data

TOOLS: AI Classification & Matching | Data Catalog & Data Stewardship Tools → | BI Integration | AI / Analytics

Insight

Defined Data
- Data is discovered, profiled and classified according to type.
- Ex: Determine the sources, types, and properties of Program's records and locations

Documented/Governed Data
- Data is continually onboarded with artifacts (Lineage, Quality, etc.)
- Bring metadata together with ontologies for drilling and hierarchies
- Ex: Drill Into Program's

AI Ready Data
- Using the cleansed & enhanced data to enable productivity
- Wiki Type Sharing Of Self-Provisioned Environments
- Atomic Data Analytics

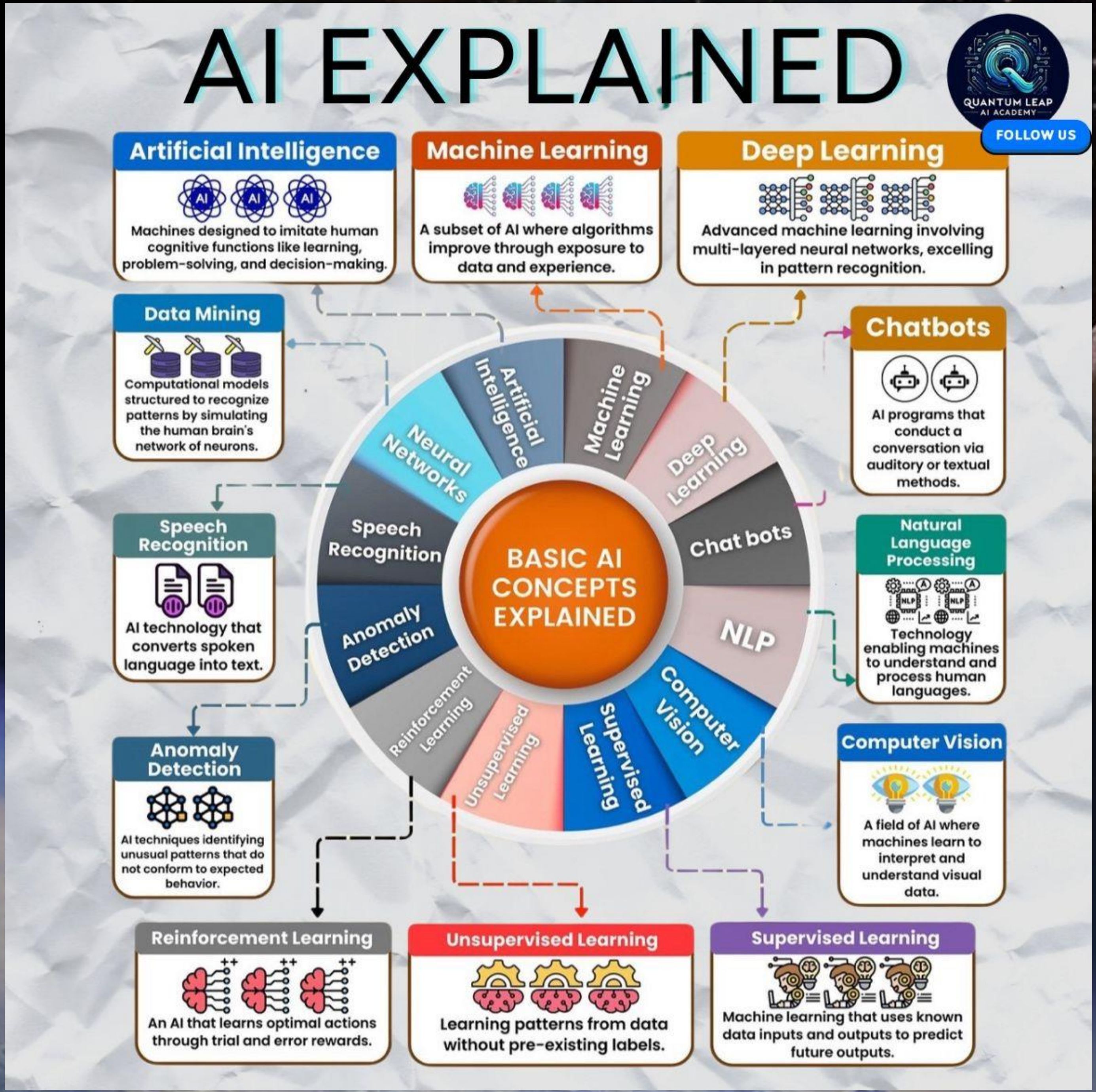
AI Considerations

The AI Platform Shift

- As adoption continues and generative AI applications become more powerful, the question is no longer whether organizations will embrace generative AI—but how they can use it most effectively
- Moore's Law is a longstanding observation in the field of computer science, named after Gordon Moore, cofounder of Intel.
  - Under Moore’s Law, computing power roughly doubled every two years with the number of transistors in an integrated circuit.
  - Under new scaling laws, AI capabilities can double every six months.
  - As organizations use AI to tackle complex challenges, the impact of this accelerated pace will be felt everywhere, and it’s this rate of change that makes the AI platform shift uniquely transformative

Tech paradigm shifts of the last 50 years

Client server → Internet → Mobile and Cloud → AI


AI EXPLAINED
QUANTUM LEAP AI ACADEMY
FOLLOW US
Artificial Intelligence
Machines designed to imitate human cognitive functions like learning, problem-solving, and decision-making.
Machine Learning
A subset of AI where algorithms improve through exposure to data and experience.
Deep Learning
Advanced machine learning involving multi-layered neural networks, excelling in pattern recognition.
Data Mining
Computational models structured to recognize patterns by simulating the human brain's network of neurons.
Chatbots
AI programs that conduct a conversation via auditory or textual methods.
Speech Recognition
AI technology that converts spoken language into text.
Natural Language Processing
Technology enabling machines to understand and process human languages.
Anomaly Detection
AI techniques identifying unusual patterns that do not conform to expected behavior.
Computer Vision
A field of AI where machines learn to interpret and understand visual data.
BASIC AI CONCEPTS EXPLAINED
Artificial Intelligence
Machine Learning
Deep Learning
Chat bots
NLP
Computer Vision
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Anomaly Detection
Speech Recognition
Neural Networks
Reinforcement Learning
An AI that learns optimal actions through trial and error rewards.
Unsupervised Learning
Learning patterns from data without pre-existing labels.
Supervised Learning
Machine learning that uses known data inputs and outputs to predict future outputs.

AI: “GARBAGE IN → GARBAGE OUT”

- Three common data quality issues: inaccuracy, inconsistency, incompleteness (and others)
- Humans make up for data issues in explicit and implicit ways
  - Need for Human Oversight
    - ✓ Good example of where bias can be an issue: Resume reviews. If you ask/answer a question with specific wording, an AI tool will prefer your phrasing. The way you would ask/answer might not be the same as how someone else would.
    - ✓ Consider intent: Is the point of your activity to create a record of something happening, or is it for you to think through the exercise? (Performance review example). Think: If you have to ask a machine to write a report - is that report worth writing?
    - ✓ "What we use these tools for needs to be considers why we are doing the activity" e.g. performance reviews vs a performance report

AI: "GARBAGE IN → GARBAGE OUT" (cont.)

- AI doesn't (yet) have those human workarounds
  - ✓ How many R's are in Strawberry?
- High Quality AI-Ready Data Is The Foundation For AI & Machine Learning
  - ✓ Clean, complete, authoritative all support FAIR data (findable, accessible, interoperable, reusable)
  - ✓ AI-ready data accelerates time to insight

Human Oversight of AI

- ***Scientific best practices*** to verify and validate
- ***CIO Guidance*** to verify and validate
- What have YOU seen?
  - Six-fingered astronauts
  - Mis-captioned photos
  - Incorrect math
  - Fabricated references

Example Frameworks for AI Data Readiness

Are we evolving to AI Engineering?

ACCURACY

Predictions or classifications closely match real world outcomes.

ROBUSTNESS

AI performs well across a range of inputs;

Clean, well-prepared data contributes to a robust model.

FAIRNESS

Biases are understood and mitigated to manage fairness issues across a range of demographics.

EFFICIENCY

Models are optimized to provide accurate, robust, fair results without excess compute.

[Clean data is the foundation of machine learning | TechTarget]

*As a community, we may adapt this or other AI Data Readiness Frameworks

Data Expectations of Artificial Intelligence

Artificial Intelligence

The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. *(Oxford Dictionary)*

Data

Visual – imagery/video
Speech – audio
Decision – telemetry, commands

Processing

Perception – ocular/optics inputs with pattern recognition
Recognition – acoustic or word matching
Decision analysis – data handling & pattern recognition

AI In-Flight

Determine how to ensure crew safety, vehicle integrity, & mission success while realizing benefits of *learning* AI technology to accomplish that mission more safely, effectively, and efficiently.

AI Use Cases

1) AI Model for Code Review
2) Space Medicine Playbook (RAG)
3) Space Medicine Chatbot
4) Health Surveillance
5) Code modernization/generation
6) Predictive maintenance
7) Grant Analysis/Grading
8) Scientific Data Analysis
9) XR Environment – Geology Computer Vision
10) EVA Tool Project – Identify Necessary Tools for an EVA
11) Cognitive Estate Determination – Computer Vision Project
12) MCC Ops console that identifies out-of-family / anomalous events
13) Reduces the loss of SME knowledge because of workforce attrition
14) Supports the Vehicle Development Process
15) Performance and Awards assessment
16) Large data analytics

Human Health Examples

What will AI allow HRP to do?

01

Leverage AI and Data Science to drive research to **help mitigate human space flight risk, moving from LEO to Long Duration Exploration Planetary Missions**:

- Adverse Health Outcomes
- Long Term Health Outcomes
- Adverse Mission Outcomes
- Performance and Behavioral Health Decrements
- Performance Decrement and Crew Illness due to Inadequate Food and Nutrition
- Ineffective or Toxic Medications

Capabilities

02

- **Interpret large data sets (**Omics, Pharmacology, etc.) from multiple sources and using Small Language Modeling
- **Decision Intelligence**: Predictive analytics
- **Immersive technologies**: Extended reality
- **Critical condition su**rveillance
- **Algorithm medicine:** Evidence-based practices
- **Predictive models:** Monitor anticipated medical change from the crew

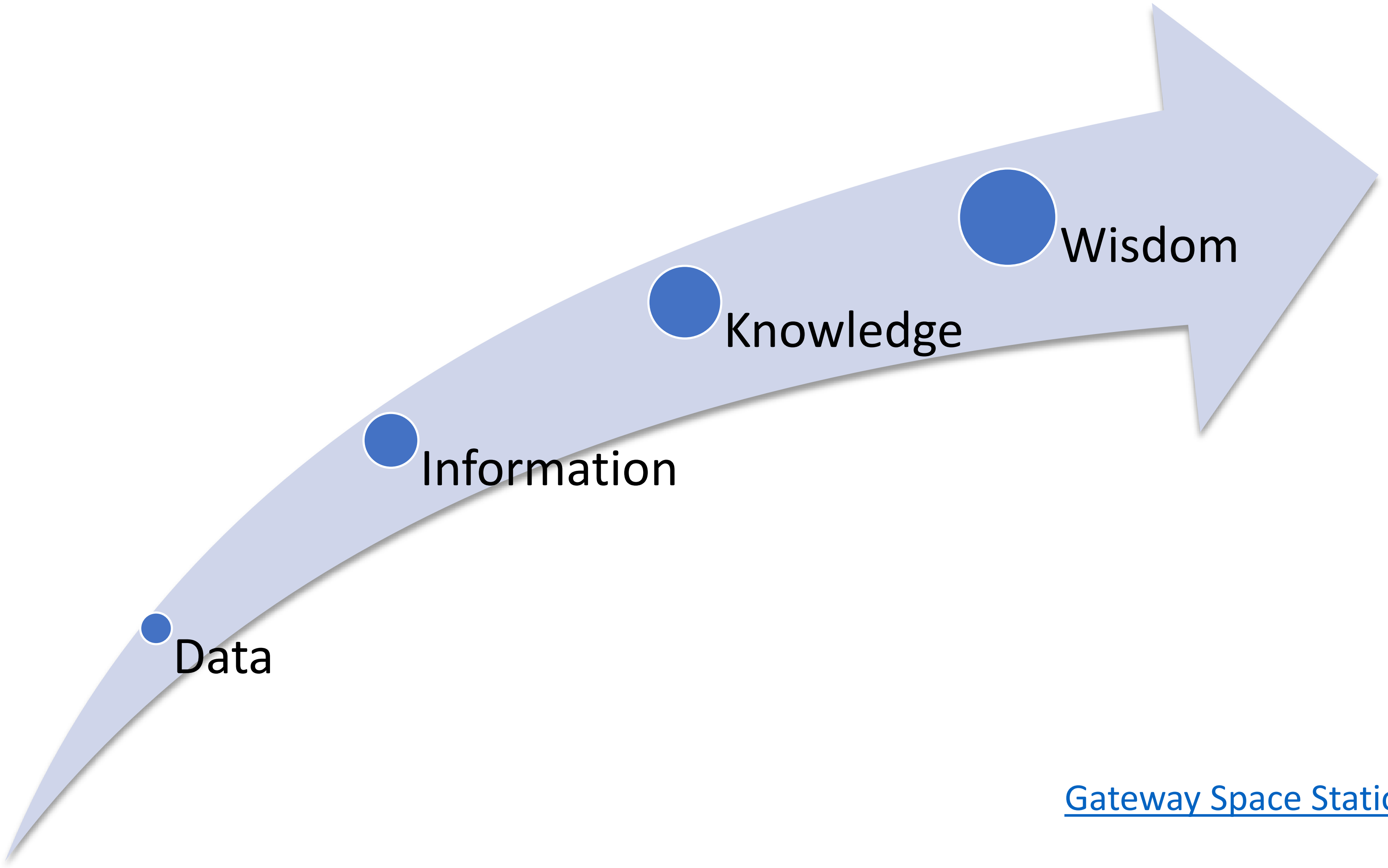
Create a highly capable team

03

Delivers disruptive, game-changing AI/ML and **data science solutions**

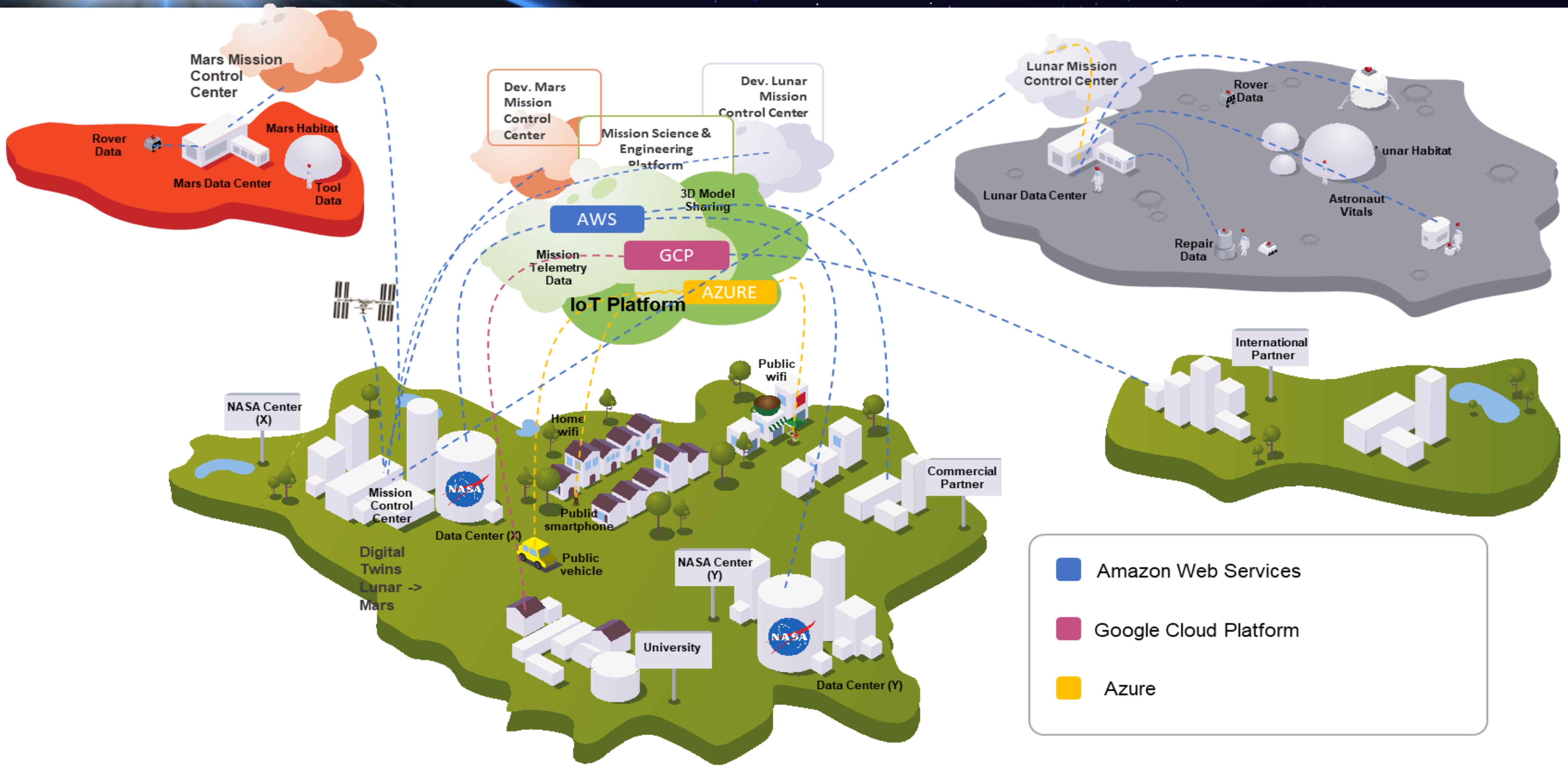
- A combination of Agency, HHPD-HRP and Vendors AI skillsets
- Anticipate service needs
- Leverage applied research and clinical data to answer the complex questions for the betterment of humanity

Thoughts?



[Gateway Space Station in 3D - NASA]

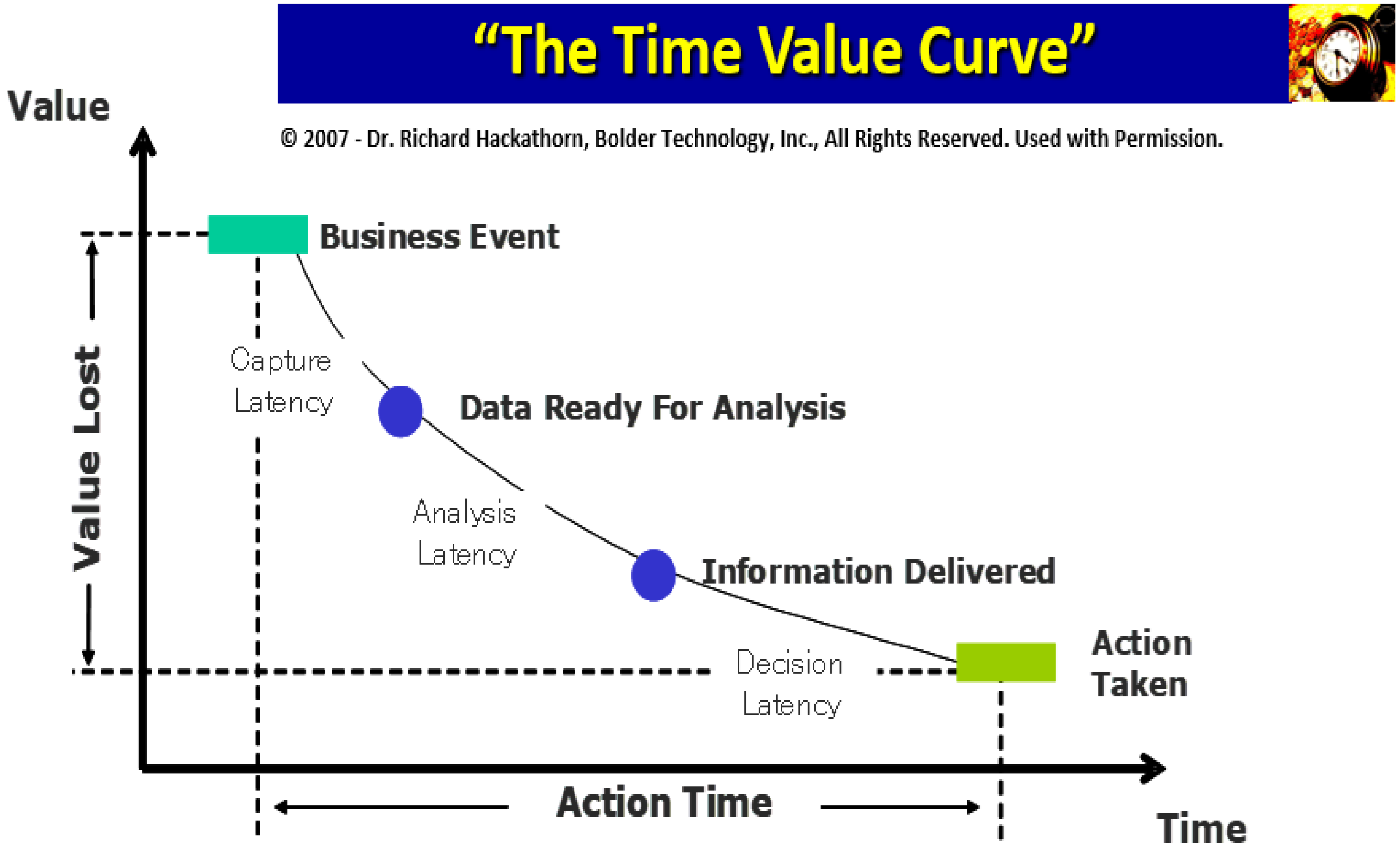
A View of the Future



DATA TIME VALUE IMPLICATIONS

Not Just Data Growth,

Need for Speed!



Data at the speed of machine learning

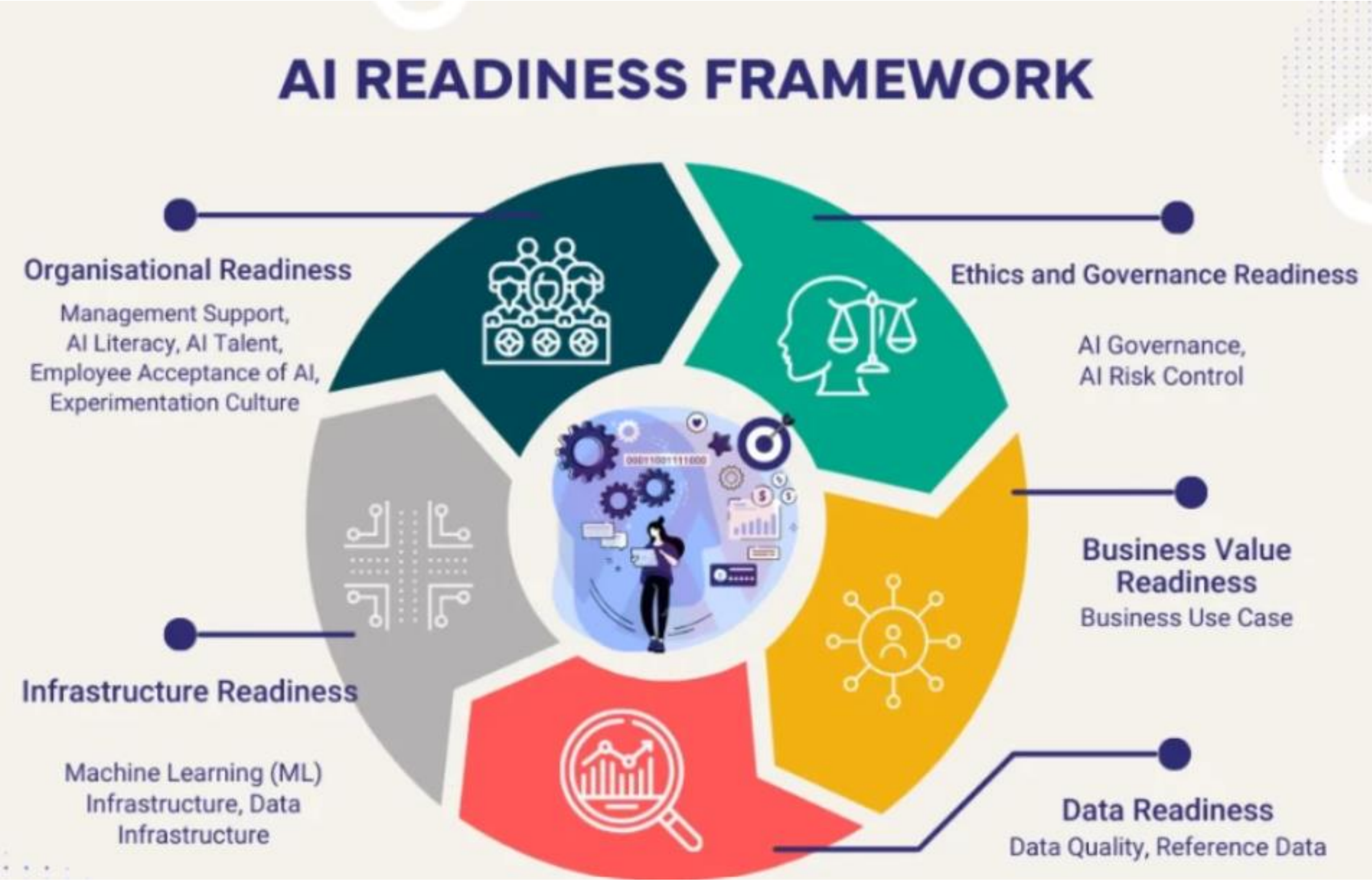
AI Ready Framework (Research in Progress)

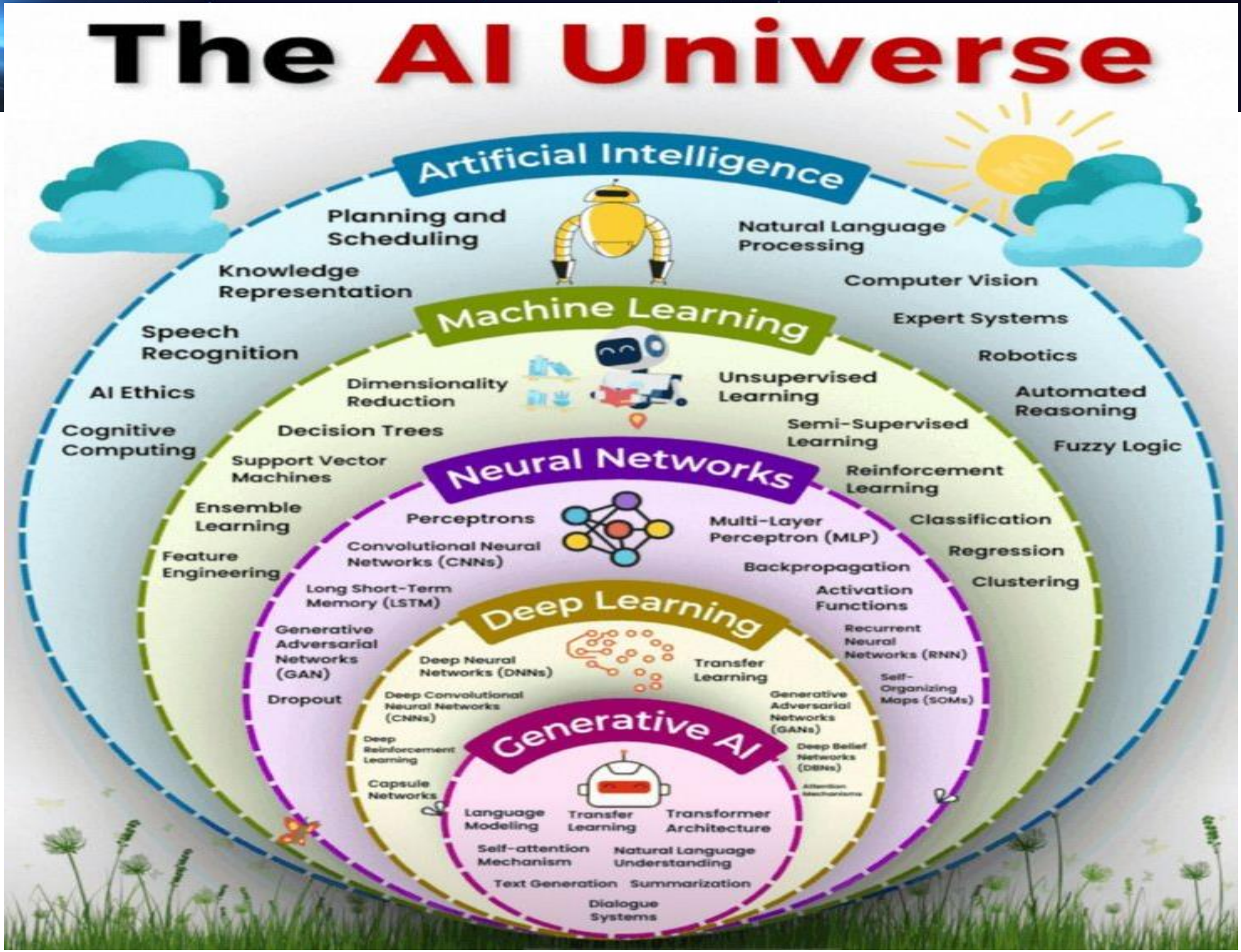
"The measure of intelligence is the ability to change"

- Albert Einstein

"It's likely that machines will be smarter than us before the end of the century—not just at chess or trivia questions but at just about everything, from mathematics and engineering to science and medicine."

- Gary Marcus




The AI Universe
Artificial Intelligence
Planning and Scheduling
Knowledge Representation
Speech Recognition
AI Ethics
Cognitive Computing
Natural Language Processing
Computer Vision
Expert Systems
Robotics
Automated Reasoning
Fuzzy Logic
Machine Learning
Dimensionality Reduction
Decision Trees
Support Vector Machines
Ensemble Learning
Feature Engineering
Unsupervised Learning
Semi-Supervised Learning
Reinforcement Learning
Classification
Regression
Clustering
Neural Networks
Perceptrons
Convolutional Neural Networks (CNNs)
Long Short-Term Memory (LSTM)
Generative Adversarial Networks (GAN)
Dropout
Multi-Layer Perceptron (MLP)
Backpropagation
Activation Functions
Recurrent Neural Networks (RNN)
Self-Organizing Maps (SOMs)
Deep Learning
Deep Neural Networks (DNNs)
Deep Convolutional Neural Networks (CNNs)
Deep Reinforcement Learning
Capsule Networks
Transfer Learning
Generative Adversarial Networks (GANs)
Deep Belief Networks (DBNs)
Generative AI
Language Modeling
Transfer Learning
Transformer Architecture
Self-attention Mechanism
Natural Language Understanding
Text Generation
Summarization
Dialogue Systems