Presenters: Ali Raz, Ramakrishnan Raman, Barclay Brown

Artificial Intelligence – Systems Engineering Primer

www.incose.org/IW2023

AI-SE Primer Team (Active Contributors)

Ali K. Raz
George Mason University

Barclay Brown
Raytheon Technologies

Tom McDermott
Stevens Institute of Technology

Ramakrishnan Raman
Honeywell.

Phill Barry
L3 Harris

Rael Kopace
Robotic Research

Kevin Robinson
Shoal Group

Kerry Lunney
Thales

Alice Squires

Steve Doskey
MITRE

AI-SE Primer Outline

- **Introduction & Scope**
  - Value between AI and Systems Engineering
  - Scope of AI Primer
  - Audience

- **Foundations of Artificial Intelligence**
  - What is AI?
  - Rule Based Systems
  - Machine Learning
  - Typical Application Areas of AI

- **Systems Engineering for Artificial Intelligence Systems**
  - The SE Process
  - Challenges facing Engineering of AI Systems
  - New Failure Modes
  - Unpredictability & Fragility of Behavior
  - Robust design & trustworthy

- **Artificial Intelligence for Systems Engineering**
  - AI for SE Processes
  - Concept Stage
  - Development Stage
  - Verification & Validation
  - Production
  - Technical Management Processes

- **Key Concepts/ Pain Points when applying AI**

- **Myths & Misunderstandings**

- **Further Reading**

*This is also our session outline for today!*

Introduction & Scope

- This Primer sets forth to demystify AI and provide several recommendations and considerations to adopt, adapt and refine the incorporation of AI into the world of systems
  - It introduces AI, covering the different types of AI, as each has evolved over time
  - It outlines potential application areas of AI, followed by the twofold interactions and influence of AI for Systems Engineering (AI4SE) and Systems Engineering for AI Systems (SE4AI)

- The guidance and descriptions provided are through the lens of a systems engineer, but the messaging is still very relevant to any engineering discipline working with AI

- From this Primer, systems engineers will also be able to identify areas in their AI knowledge that may require further research, education, training, coaching and/or mentoring as needed

Foundations – AI Categories

- Artificial Intelligence (AI) is perhaps best defined as the ability of a system to exhibit behavior, which if exhibited by a human being, would be considered *intelligent*.
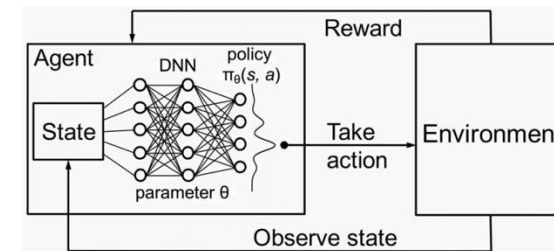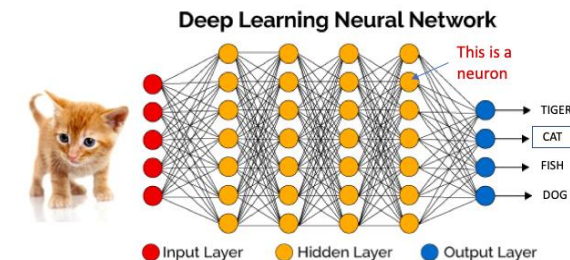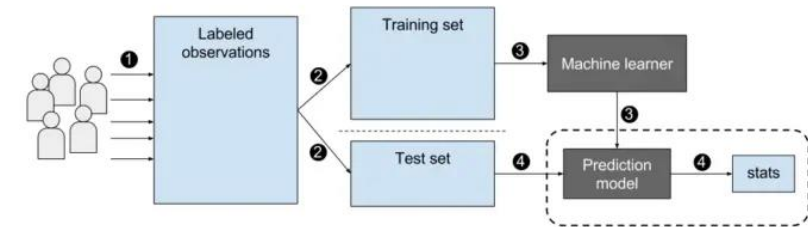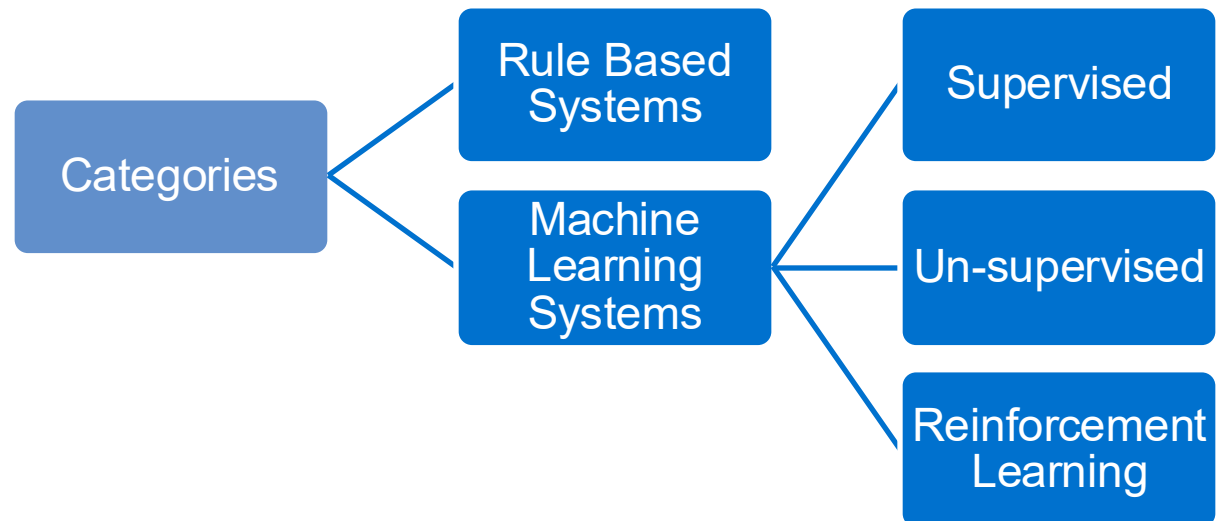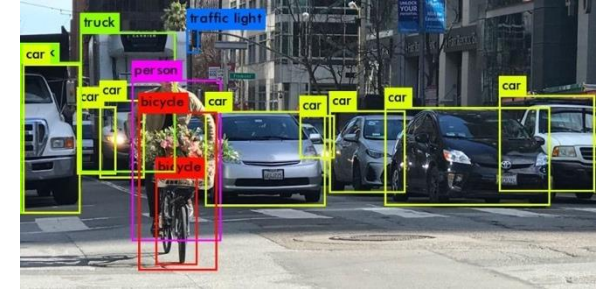
Foundations: Application Areas

Pattern Matching

- Classification of objects in images or video streams.
- E.g. recognizing vehicles, pedestrians, and other obstacles in images from car-mounted cameras,

Regression & Prediction

- Predictive modelling of dependent variable based on the relationships identified between independent variables.
- E.g. .Learn from sensor data such as shaft vibration and predict future failures based on the sensor data characteristics
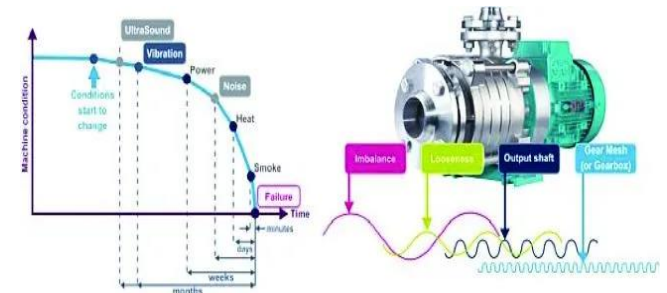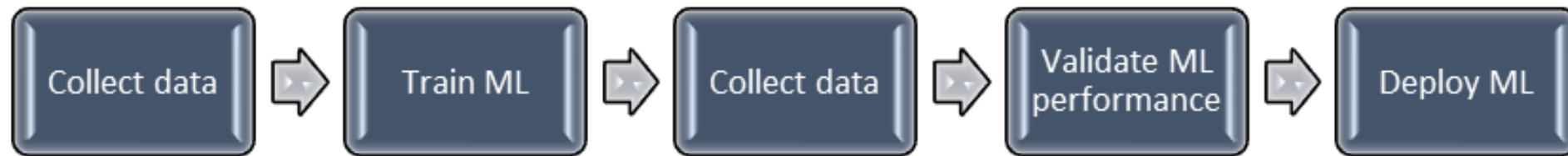

Natural Language Processing

- Written text form or speech sound patterns can be analyzed using ML
- E.g., the meaning of a consumer's inquiry when calling an automated customer service, or a request for a style of music from a smart speaker

src: https://languagegeekblog.wordpress.com/

Systems Engineering for AI Systems

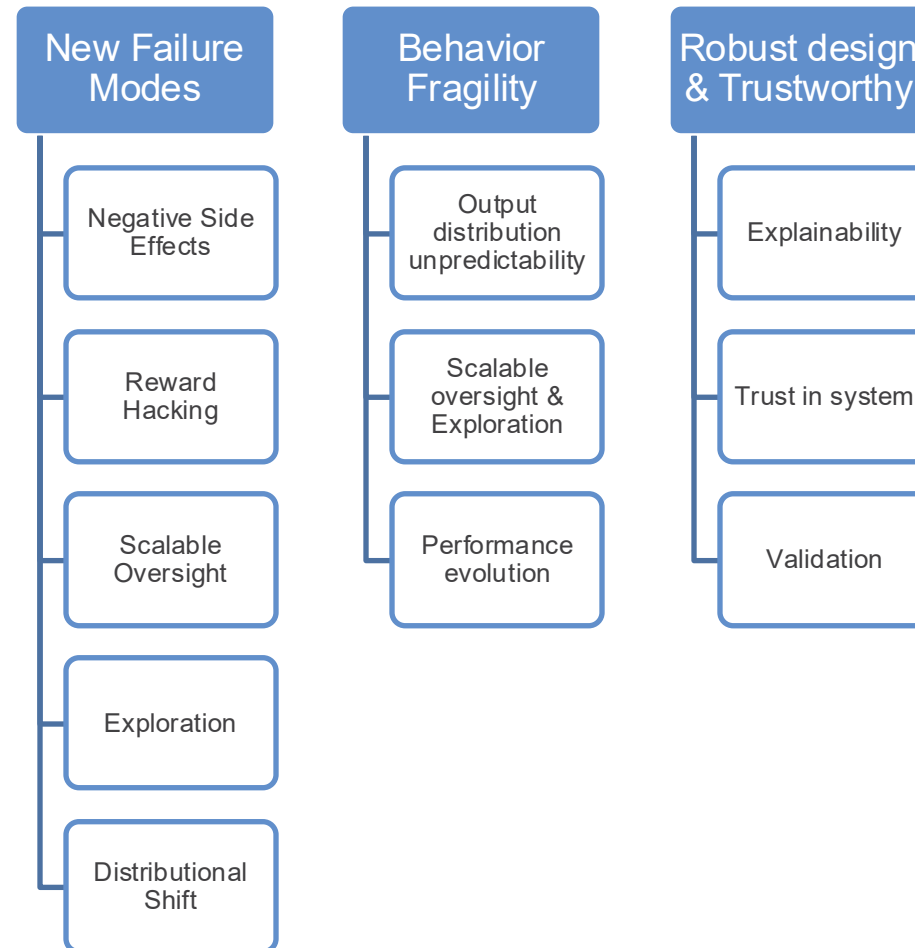| Collect data | → | Train ML | → | Collect data | → | Validate ML performance | → | Deploy ML |

- What is the intelligence that is required for the system? What are the desired behaviors? What are the desired performance levels? What are the acceptable tolerances for errors/ false-positives/ false negatives?

- What is the source data on which the intelligence can be built? What are the data types? What are the restrictions on usage of the data? What should be the training, validation & test data sets?
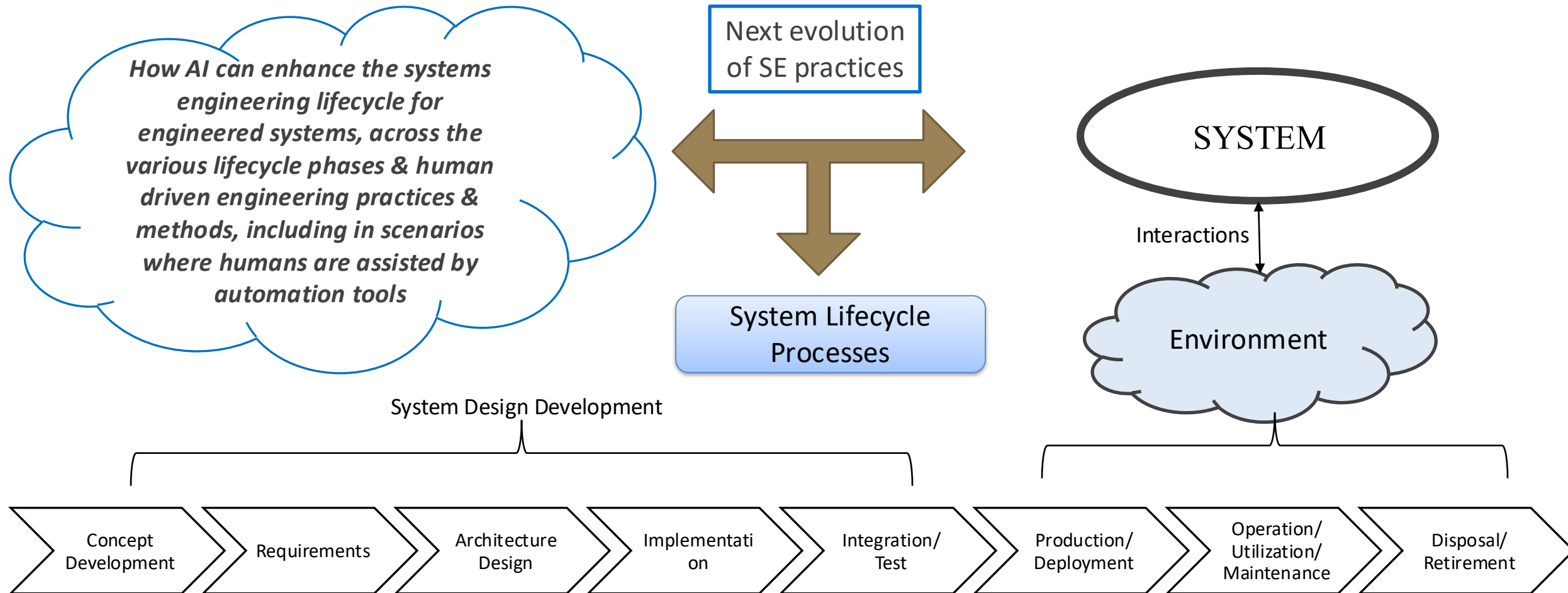
New Failure Modes

1. The new failure modes not previously experienced in the engineering of systems
2. The unpredictability and fragility of behavior due to the inconsistencies in the input data
3. The inability to explicitly validate system performance due the level of trust in the systems and its robustness to future environments

New Failure Modes
- Negative Side Effects
- Reward Hacking
- Scalable Oversight
- Exploration
- Distributional Shift

Behavior Fragility
- Output distribution unpredictability
- Scalable oversight & Exploration
- Performance evolution

Robust design & Trustworthy
- Explainability
- Trust in system
- Validation

AI for Systems Engineering

How AI can enhance the systems engineering lifecycle for engineered systems, across the various lifecycle phases & human driven engineering practices & methods, including in scenarios where humans are assisted by automation tools

Next evolution of SE practices

System Lifecycle Processes

SYSTEM

Interactions

Environment

System Design Development

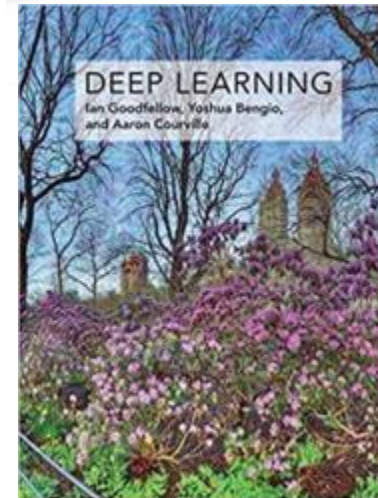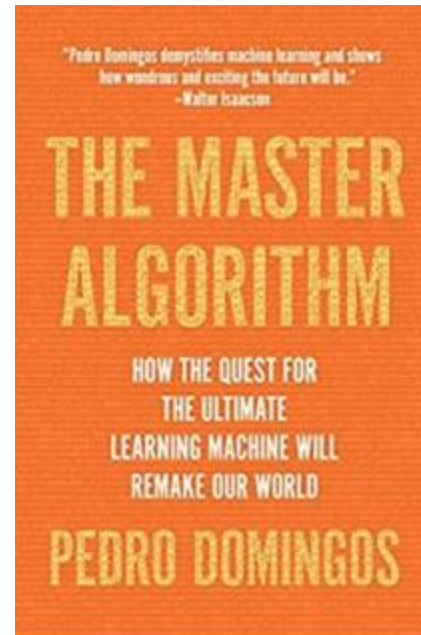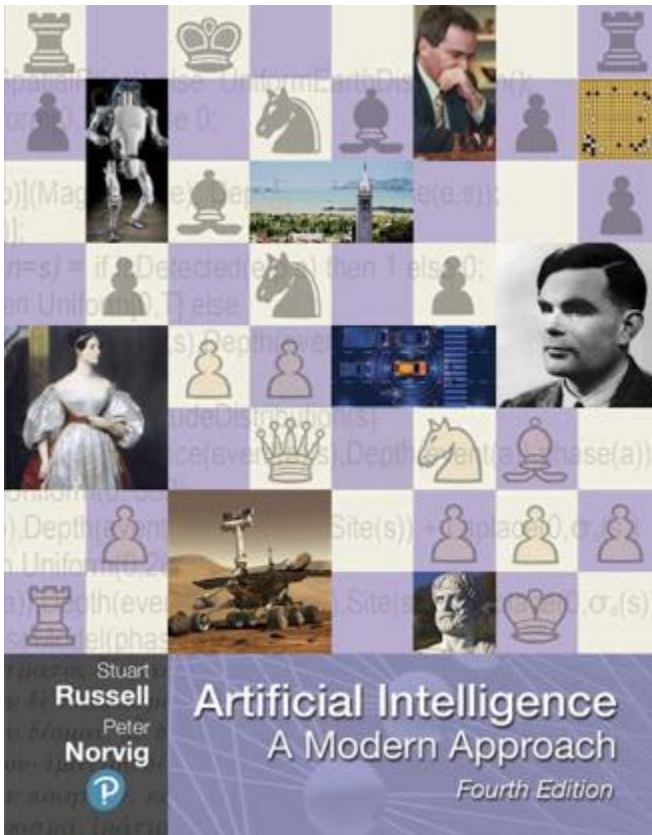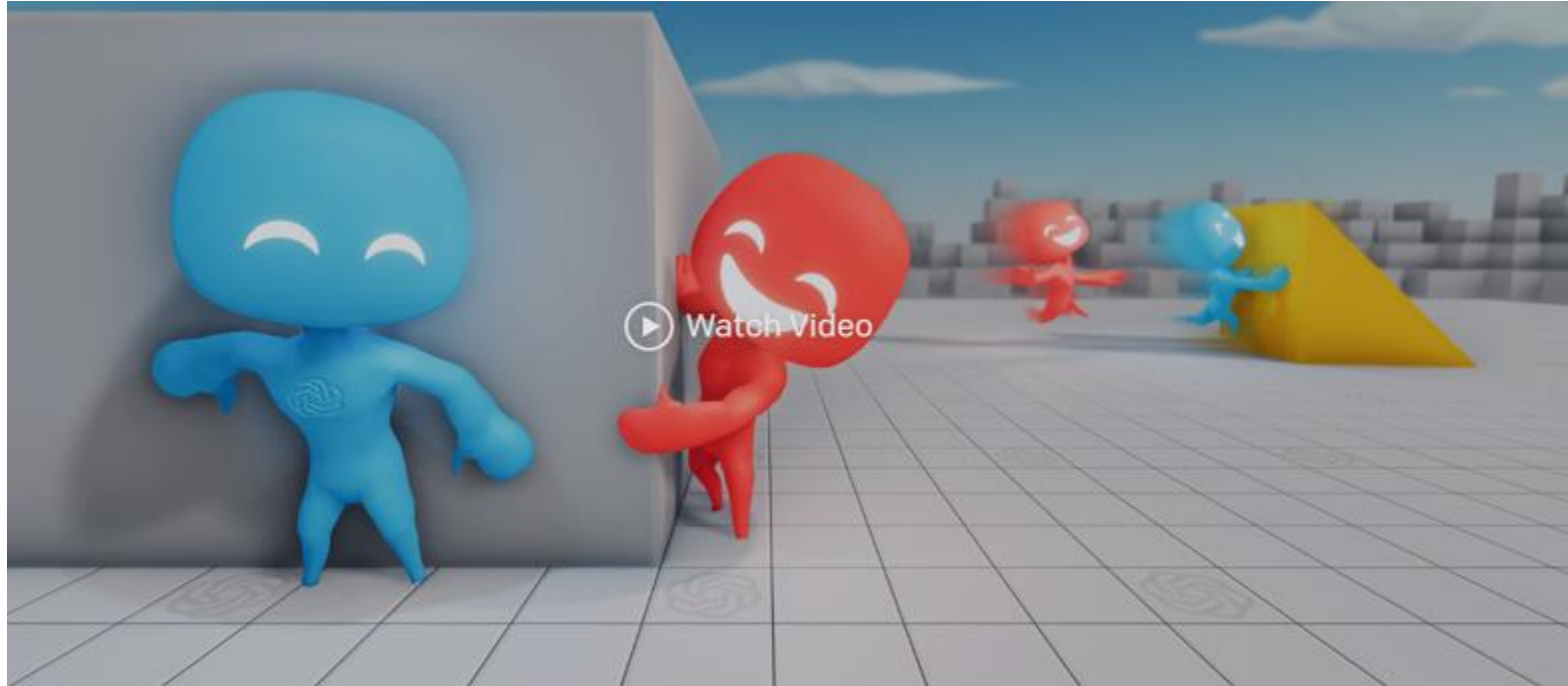| Concept Development | Requirements | Architecture Design | Implementation | Integration/ Test | Production/ Deployment | Operation/ Utilization/ Maintenance | Disposal/ Retirement |

Myths & Misunderstanding

- **AI Super Intelligence**
  - "A superintelligence is a **hypothetical** agent that possesses intelligence far surpassing that of the brightest and most gifted human minds." (Wikipedia)

- **Not all AI is driven by Deep Neural Networks:**
  - Statical regression and Rule-based expert systems are also a form of AI

- **Deep Neural Network Myths**
  - ❑ They work just like the brain does
  - ❑ Busted: we don't KNOW how the brain produces decisions, opinions, emotions, thoughts

  - ❑ As they grow in size and speed they get smarter
  - ❑ Busted: often better-designed smaller networks outperform larger ones. Size alone doesn't matter.

  - ❑ Solving a hard or complex problem shows that an AI is nearing human intelligence
  - ❑ Busted: Deep Blue mastered chess but can do NOTHING else remotely resembling human behavior or thought

Technical and non-Technical Readings

Some fun with AI – Play Hide and Seek

https://openai.com/blog/emergent-tool-use/