International Council on Systems Engineering
A better world through a systems approach

Robust Testing and Simulation Frameworks for Artificial Intelligence Systems in Spacecraft Operations

Stephanie Anderson, *Colorado State University*

Dr. Steven Simske, *Colorado State University*

INCOSE International Symposium 2025 | Ottawa, Canada

Today's Agenda

- Background on Artificial Intelligence in Spacecraft
- Real World examples
- Systems Engineering Lifecycle
- Testing and Simulation
- Onboard Evaluation
- Long-Term Monitoring and Updates
- Future Work

Background on Artificial Intelligence in Spacecraft

Background on Artificial Intelligence (AI) in Spacecraft

- Motivation for AI in spaceflight
  - Deep space spaceflight operations involve extreme distances and time delays
  - Strong need for autonomous decision-making and adaptive system management
- AI vs. auto pilot
  - Both can reduce workload on the pilot
  - Auto pilot is incapable of making decisions or learning
- AI is becoming more prominent and desirable
- Key areas where AI is applied in spacecraft today
  - Autonomous navigation
    - Help spacecraft autonomously navigate unknown terrain (Moon, Mars, etc.)
  - Fault detection and diagnostics
    - Assist in predicting failures before they happen and suggest corrective actions when they do happen
  - Resource management
    - Assist in optimizing the use of power, fuel, and computing resources
  - Planning and scheduling
    - Automatically schedule tasks based on resource availability and mission priorities

https://theconversation.com/five-ways-artificial-intelligence-can-help-space-exploration-153664

Background on Artificial Intelligence (AI) in Spacecraft (cont.)

- Different types of AI
  - Expert systems
    - Rule-based systems used for problem-solving (fault protection, etc.)
  - Machine learning
    - Systems that can adapt based on new data, useful in dynamic/changing environments
  - Reinforcement learning
    - Allowing AI to learn from "trial and error"
  - Computer vision
    - Tasks such as terrain analysis and object recognition (landings, dockings, etc.)
- Challenges
  - Reliability
    - AI must be highly reliable and dependable in high-risk conditions
  - Verification and validation
    - Can be difficult to test AI for every possible situation astronauts may encounter, especially in deep space where humans have little to no experience
  - Ethical limitations and authority
    - Ensuring AI decisions do not override human authority inappropriately
    - Human override should always be possible and assumed to be the "final say"

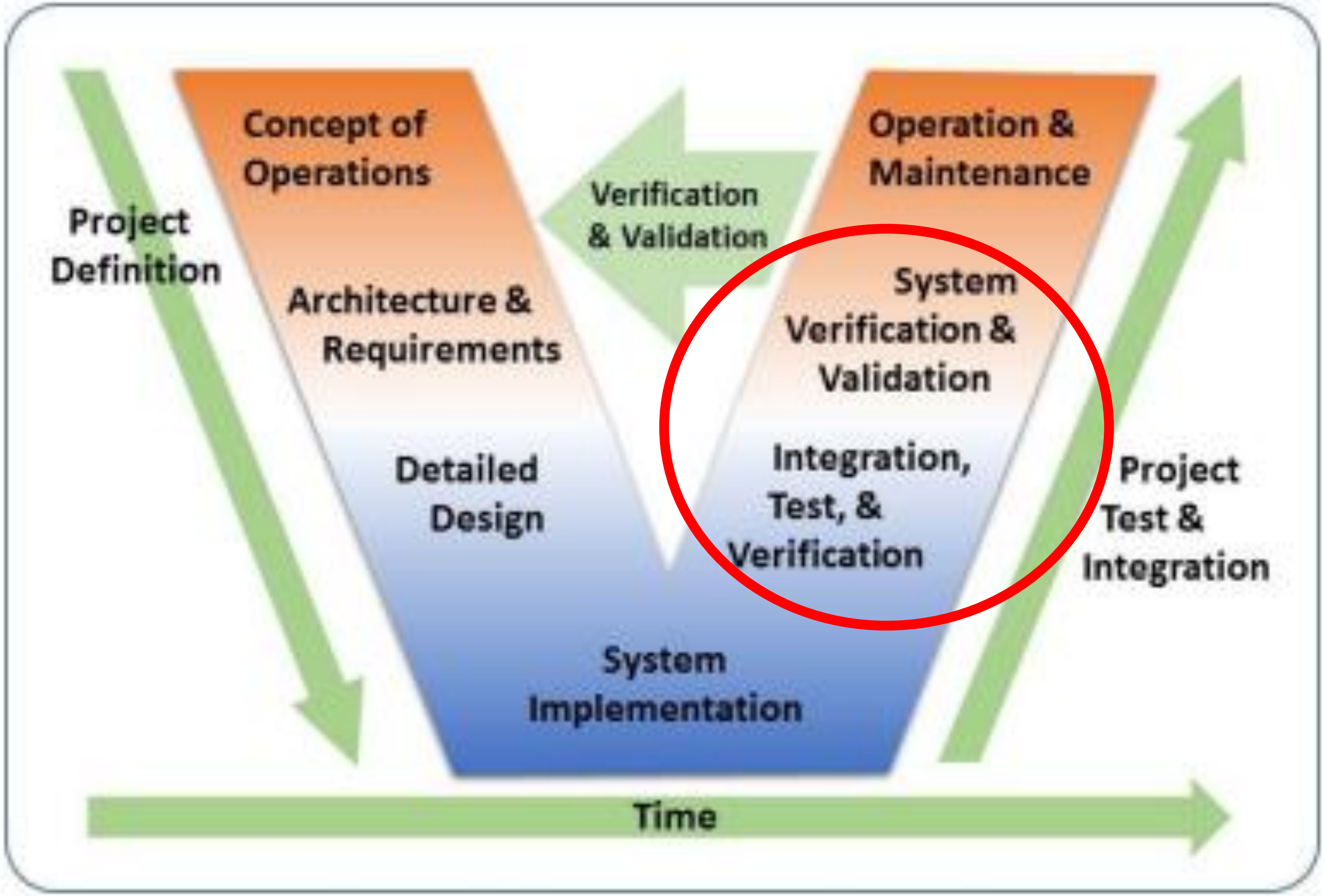
Real World Examples

Timeline of the Evolution of AI in Spacecraft

Decade	Key Development	Example
1970s	Basic fault protection	Voyager
1980s	Expert systems research	Deep Space Network
1990s	Full onboard AI	Deep Space One (Remote Agent)
2000s	AI for planning and scheduling	Mars Rovers (MAPGEN)
2010s	Machine learning and autonomy	Curiosity Rover, Orbital Debris AI (Agatha)
2020s	Deep learning, cognitive AI, swarm AI	Perseverance Rover, ASTROBEE, Lunar Anomaly Detection

Systems Engineering Lifecycle

Systems Engineering Lifecycle

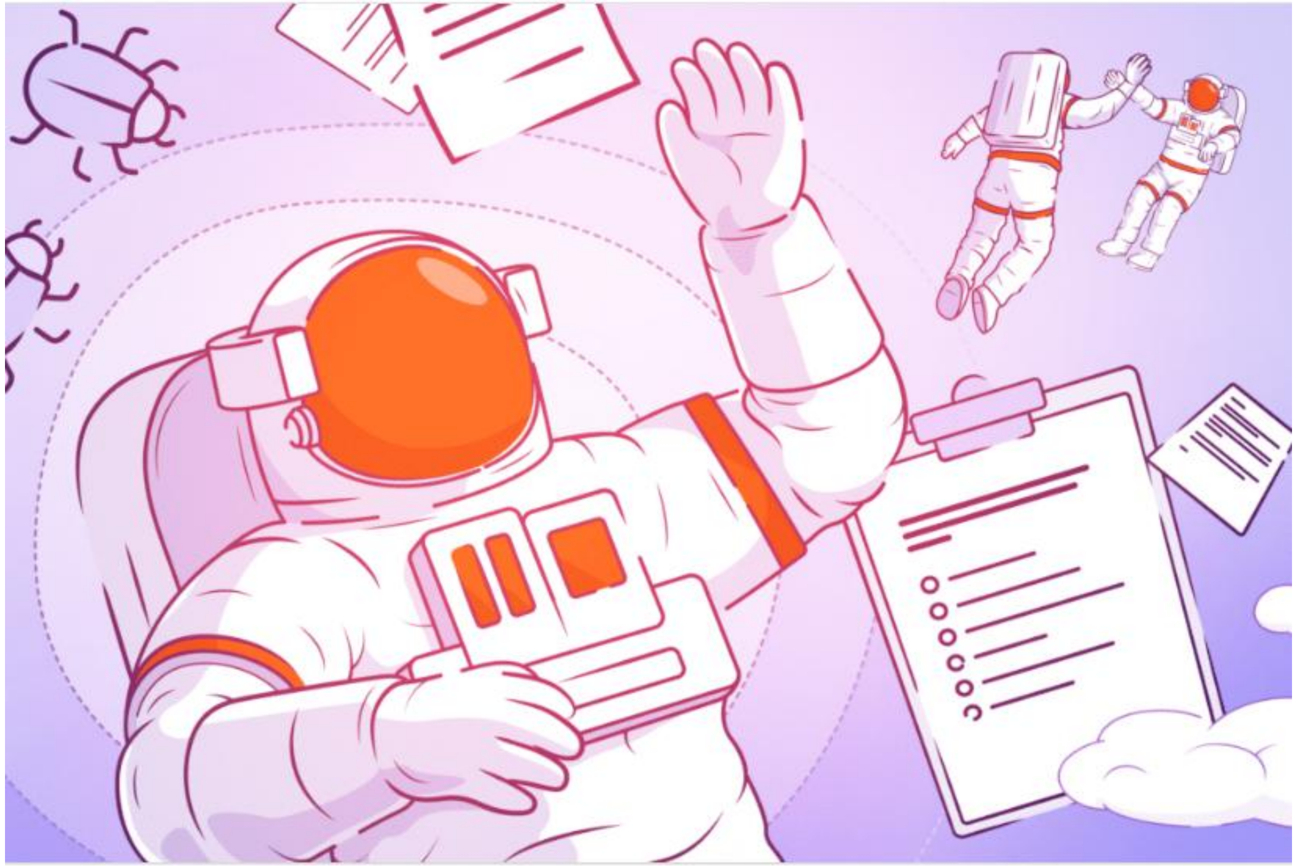


https://www.incose.org/docs/default-source/los-angeles/libraryresources/se-jr-handdbook-2015.pdf?sfvrsn=1d3c80c6_2

Testing and Simulation

Ground Based Testing

- Hardware-in-the-Loop (HIL)
  - Test AI with actual spacecraft hardware to assess integrated interactions
    - How hardware reacts to AI commands
    - Determine potential timing limitations
- Human-in-the-Loop
  - Astronauts interact with AI in simulators to evaluate usability, communication clarity, and trust
    - Used to determine where and how AI should be used
    - Fine tune manual human override



https://deviniti.com/blog/application-lifecycle-management/software-testing-challenges/

Ground Based Testing (cont.)

- Simulations
  - Types of simulations
    - High-fidelity mission simulations
      - Reproduce full mission profiles including launch, orbital insertion, orbital maneuvers, free-flight, docking, landing, etc.
      - Within a realistic environment and environmental conditions, such as zero gravity, radiation, micrometeoroids, and erratic system behaviors
      - Used to “stress test” the AI decision-making under both expected and edge-case conditions
    - Failure mode simulations
      - Evaluate how AI specifically handles anomalies or emergency cases such as system malfunctions, communication blackouts, sensor degradation, power shortages, software glitches, etc.
      - Train AI to recognize the signs of these failures, isolate, and respond to the failures autonomously or in collaboration with human interaction
    - Human-AI interaction simulations
      - Focus on testing crew member interactions with AI during spacecraft operations
      - Real-time problem solving alongside crew members, voice and/or text commands, and behavioral understanding
      - Ensures that AI responses are clear, timely, and built on trust with the human users

Ground Based Testing (cont.)

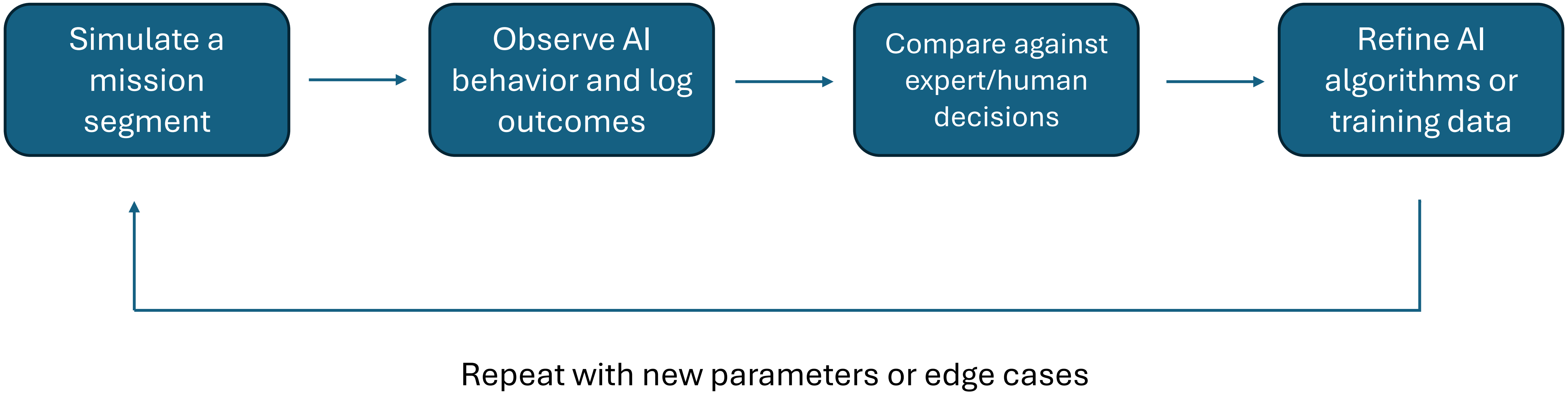
- Simulations (cont.)
  - Simulation environments
    - Digital twin
      - Virtual replicas of spacecraft systems with real-time data syncing – accurate representation is key
      - Allows for AI to interact with a realistic representation of the actual spacecraft hardware and software, and simulate responses to changes and commands
    - Augmented/Virtual reality
      - Enables immersive training environments for astronauts to practice scenarios with AI
      - Enhances the intuitive understanding of special relationships and tool use, which can be especially useful for EVA preparations/training
    - Monte Carlo
      - Used to run the AI through thousands/millions of randomized variables (sensor noise, environmental changes, combinations of system failures, etc.)
      - Statistically validate reliability and robustness under uncertainty

Ground Based Testing (cont.)

- Simulations (cont.)
  - Evaluation metrics
    - Accuracy
      - Does the AI make correct decisions based on simulated input?
    - Response time
      - How fast does it react, especially in emergencies?
    - Stability
      - Can it operate continuously over long periods of time (long duration missions) without drifting or degrading?
    - Adaptability
      - Can the AI handle new or unexpected data without failing or overfitting?
    - Crew feedback
      - Is the AI behaviour helpful, trustworthy, and not overwhelming?

https://www.americaspace.com/2013/09/20/lockheeds-gps-iii-prototype-successfully-connects-with-raytheon-ground-control-system-in-milestone-test/

Iterative Testing Cycle



Onboard Evaluation

Onboard Evaluation

- Incremental deployment and pilot studies
  - First deploy in a passive mode where AI runs in parallel with human or automated systems, but without any control authority
    - Compare outcomes to human decisions
    - Can be first deployed on an uncrewed cargo vehicle, with the human interaction coming from the ground
      - Validate performance without risk to crew
  - Begin with low-risk tasks such as environmental monitoring or scheduling
  - Gather feedback from astronauts to refine interaction models
  - These tasks/pilot studies can be conducted on the International Space Station
  - Once astronauts have built trust in the AI system, AI can be allowed to take part in decision-making
- Implement adaptive learning boundaries
  - If the AI adapts over time, control how much it can change and ensure updates are logged and can be reversed
    - Trust is earned and can be lost
- Evaluation by astronauts will be paramount

Long-Term Monitoring and Updates

Long-Term Monitoring and Updates

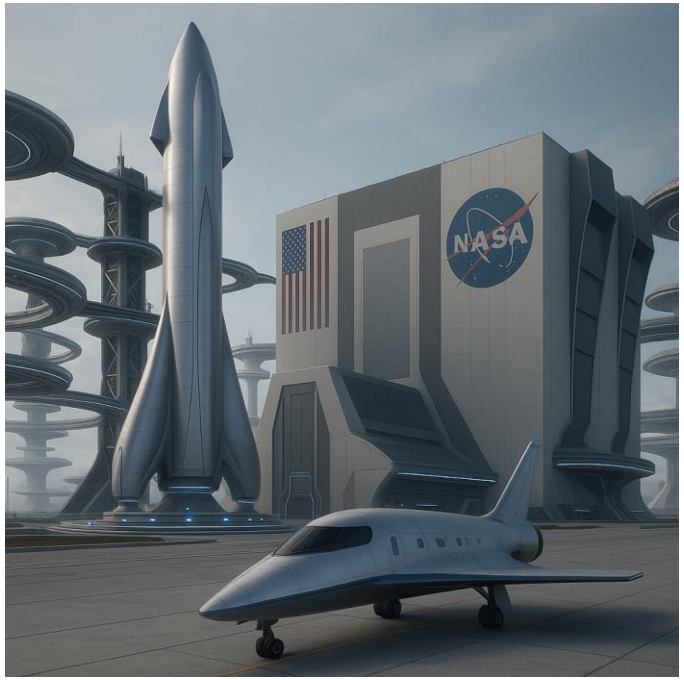
- Systems should always be monitored after deployment to assess any failures in the system or possible areas for improvement
  - Ground teams on Earth should continuously monitor AI performance via data downlink
  - If ground teams discover a previously untested combination of parameters, AI should be put through additional simulations on the ground
- Update protocols
  - Updates to the system should be thoroughly tested (through the iterative testing cycle) and pass any levied verification requirements
  - Updates should be logged in a secure tracking system
  - Previous versions of the AI should be kept in archive in case an update needs to be reversed

https://img.freepik.com/premium-vector/system-monitoring-performance-icon_996135-40157.jpg

Future Work

Future/Forward Work

- Artificial Intelligence in human-rated spacecraft is in its infancy
- Some thoughts on using AI to help review results of flight software automation testing (proprietary)
- As the lunar program and Moon2Mars evolves, so will the use of AI in spacecraft
- Dissertation will focus on studying astronaut trust in AI, and how they want AI to be implemented in the spacecraft they fly for long duration missions



(OpenAI, 2025)