Digital Systems Safety Research

A11D.SDS.6

Complex Digital Systems

Srini Mandalapu
Manny Rios
Data Science and Cyber Security Section (ANG-E271)
WJH FAA Technical Center, Atlantic City

SAS Meeting
August 2-3, 2022

Complex Digital Systems, A11D.SDS.6

Research Project Description

- Using the Overarching Properties in Novel Examples
  - Expanding the OP concepts
- Requirements Modeling
  - Identify candidate methods for formalizing requirements and design a proof-of-concept study to assess candidate MBSE methods for requirement formalization
- Estimation of Worst Case Execution Time using AI/ML
- AI/ML Verification Framework
  - Formal and continuous verification of the design

Sponsor Anticipated Outcome

- Expected outcome is the implementation of a less prescriptive, risk-based guidance for assurance approaches, methodologies, and techniques and propose assurance criteria to assure complex digital systems.
- Training material to train certification engineers on the new assurance approaches and safety impact of new technologies

Critical Milestones

- Overarching Properties in Novel Examples, phase 1 report along with training material, 9/2022.
- Requirements Modeling Phase I Final report , draft released 7/14/2022.

Research Accomplishments in FY22

- Published “An Introduction to Constructing and Assessing Overarching Properties Related Arguments (OPRAs)” (NASA DocID: 20210025425)
- Published “Safety Verification of Autonomous Systems: A Multi-Fidelity Reinforcement Learning Approach”, 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems

Complex Digital Systems, A11D.SDS.6

Using the Overarching Properties in Novel Examples

Prime Performer: C. Michael Holloway, Langley Research Center, Hampton VA
Status: On-going, Phase I ends in Sep. 2022
Phase II starts in Sep 2022 (one year)
Major Accomplishments:

Defined the Overarching Properties
- Published "Understanding the Overarching Properties (NASA/TM–2019–220292)

Developed Concepts on the Construction and Assessment OP Related Arguments (OPRAs)
- Published "An Introduction to Constructing and Assessing Overarching Properties Related Arguments (OPRAs)" (NASA DocID: 20210025425)

The Overarching Properties:

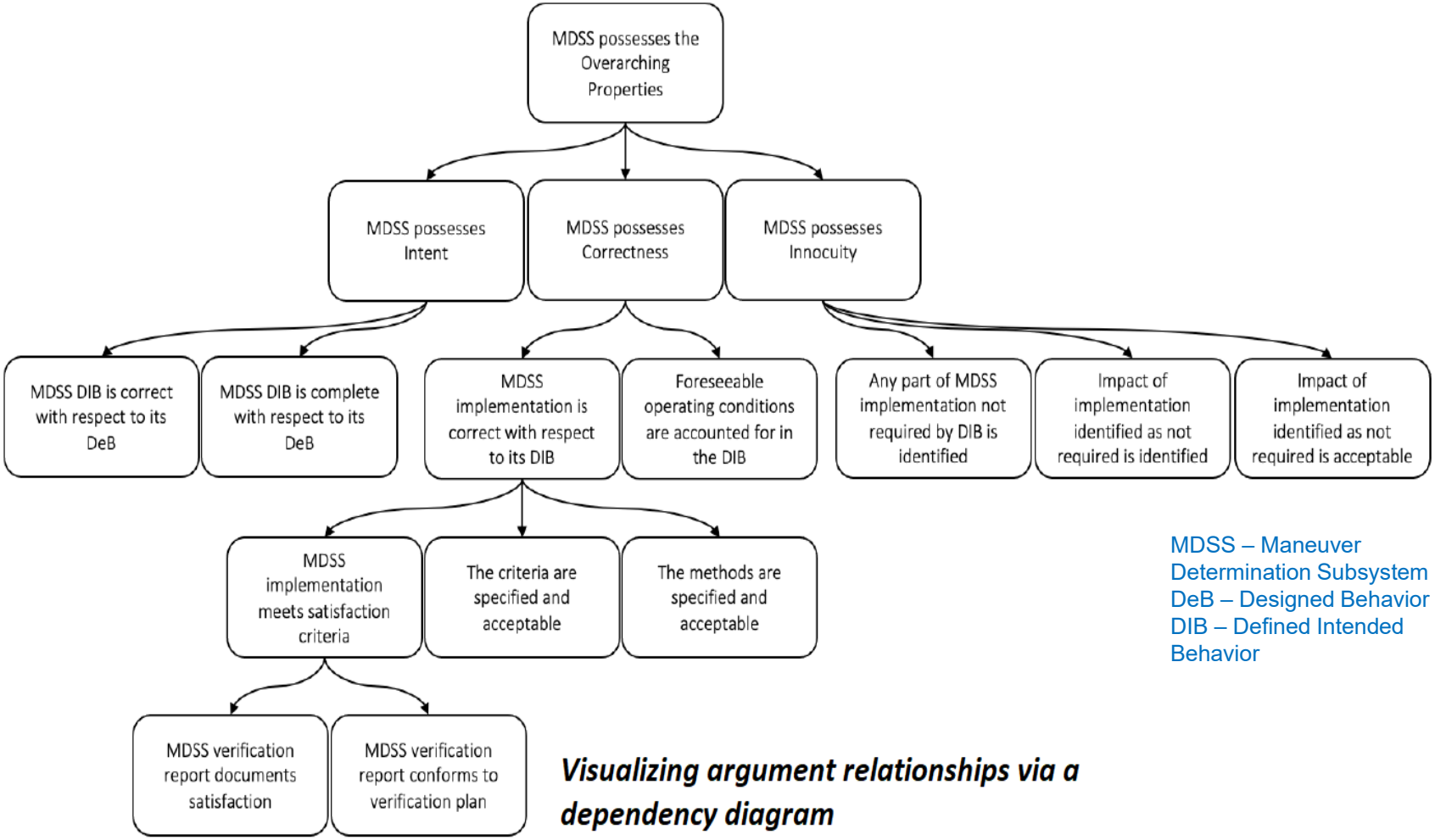
Intent: The *defined intended behavior* is correct and complete with respect to the *desired behavior*.

Correctness: The *implementation* is correct with respect to its *defined intended behavior*, under *foreseeable operating conditions*.

Innocuity: Any part of the *implementation* that is not required by the *defined intended behavior* has no *unacceptable impact*.

Complex Digital Systems, A11D.SDS.6

Using the Overarching Properties in Novel Examples



Visualizing argument relationships via a dependency diagram

MDSS – Maneuver Determination Subsystem
DeB – Designed Behavior
DIB – Defined Intended Behavior

Complex Digital Systems, A11D.SDS.6

Using the Overarching Properties in Novel Examples

ARGUMENT
your attempt to convince others to BELIEVE a CONCLUSION through REASONING from one or more PREMISES

A BINDING associates a term used in an ARGUMENT and the real-world information to which that term refers.

CONCLUSION
the statement you want your audience to BELIEVE

PREMISE
a statement you think your audience BELIEVES

REASONING
why you think the PREMISES should cause your audience to BELIEVE your CONCLUSION

BELIEVE
accept as true

An ATOMIC ARGUMENT consists of a single CONCLUSION together with its immediate REASONING, PREMISES, and BINDINGS (if any).

A COGENT ARGUMENT rationally justifies BELIEVING its CONCLUSION.

A COMPOUND ARGUMENT is an

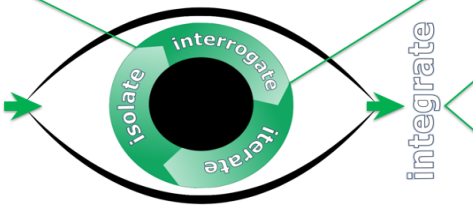
iTest for assessing the COGENCY of a COMPOUND ARGUMENT

Syntax proper?
- a. propositional CONCLUSION?
- b. a statement of REASONING?
- c. at least one PREMISE?
- d/e. PREMISES (DEFEATERS) propositional?
- f. *OP compliant* BINDINGS
- g. Proper *OP compliant* BINDINGS for artifacts?

Premises acceptable?
- a. each PREMISE BELIEVABLE?
- b. each PREMISE relevant?

Reasoning acceptable?
- a. required level of confidence known?
- b. *OP CONSTRAINT COMPLIANT* ARGUMENT engenders required level?

Yes justified?
- a. REASONING relevant?
- b. REASONING consistent with current knowledge *and OP constraints?*



ATOMIC ARGUMENTS COGENT?
Terminal PREMISES BELIEVED?
OP constraints satisfied?

C. Michael Holloway 2022

Complex Digital Systems, A11D.SDS.6

Using the Overarching Properties in Novel Examples+-

SAM and IAM Doing no Harm

The final example arose from trying to create a simple, nearly realistic illustration of what a partial argument related to the Overarching Properties [5] might look like. Its sole purpose here is to illustrate the use of FAN as it might occur during development. Note the current incompleteness of the arguments. Much remains to be done, such as expanding the *BINDING BLOCKs*, determining some of the reasoning, figuring out some necessary premises, and providing additional arguments. As with the previous example, this one is presented without additional commentary.

Example 46. SAM and IAM are harmless

```
Believing
   Subsystems SAM and IM both possess /Innocuity/ {1}

is justified by applying
   the principle of conjunction {2}

to
   SAM possesses /Innocuity/ {3}
   IAM possesses /Innocuity/ {4}
   SAM and IAM are /independent/ {5}

with
   Innocuity: definition in the OP description \
              <https://hdl.handle.net/2060/20190029284> {6}
   independent: to be defined {7}


Believing
   SAM possesses /Innocuity/ {3}

is justified by applying
   the meaning of /Innocuity/ {8}
```

*FAN – Friendly Argument Notation

Complex Digital Systems, A11D.SDS.6

Requirements Modelling

Cooperative Research: AVSI (FAA, NASA, DoD, Boeing, GE, Honeywell, Airbus)

Status: Phase I ended in March. 2022

Phase II planning to start in Sep 2022

Major Accomplishments:

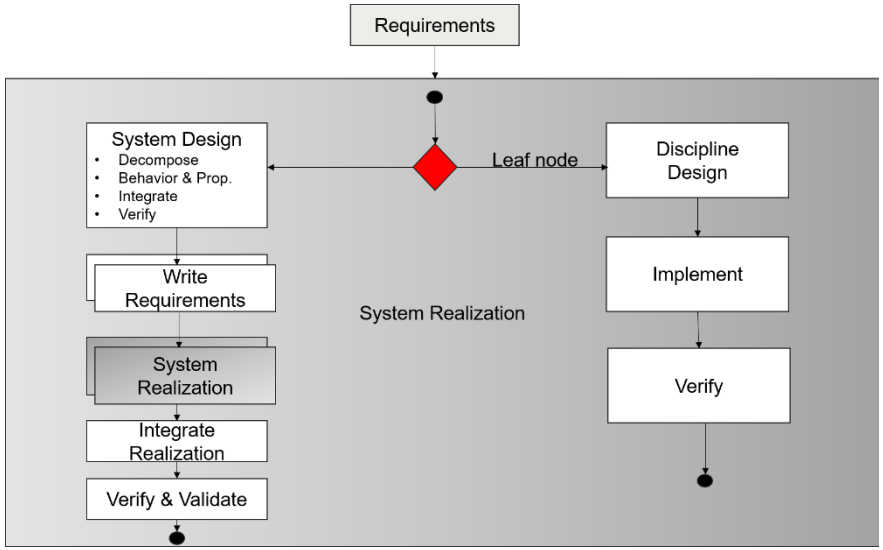
- Established a baseline success criteria for the specific MBSE activity of formalizing requirements
- Identified candidate methods for formalizing requirements
- Designed proof-of-concept study to assess candidate MBSE methods for requirement formalization

Phase II Tasks Identified:

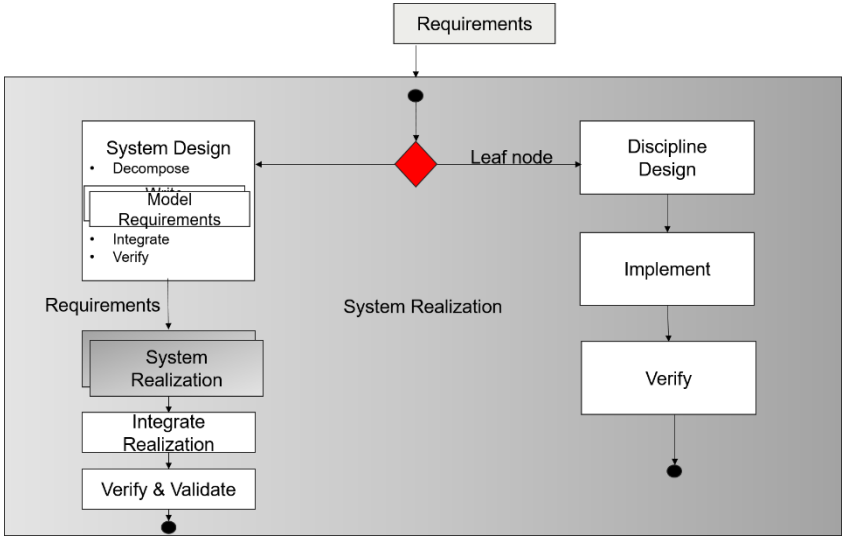
- Execute/document the proof-of-concept study as per selected MBSE requirement formalization methods
- Develop Recommendations

Complex Digital Systems, A11D.SDS.6

Requirements Modelling



Traditional System Realization Process



System Realization Process Using Logical Modelling Method (LMM)

Complex Digital Systems, A11D.SDS.6

Estimation of Worst Case Execution Time (WCET) using AI/ML

Performer: Dr. Bjorn Andersson, Software Engineering Institute
Status: Planned end date December, 2022

Major Accomplishments:

Report on the State-of-the-art in Machine Learning and its use in WCET Analysis

Report on the Software Architectures for WCET Analysis based on AI/ML

Generated Synthetic Data for training and testing the models

Selected a model for final training and testing.

Complex Digital Systems, A11D.SDS.6

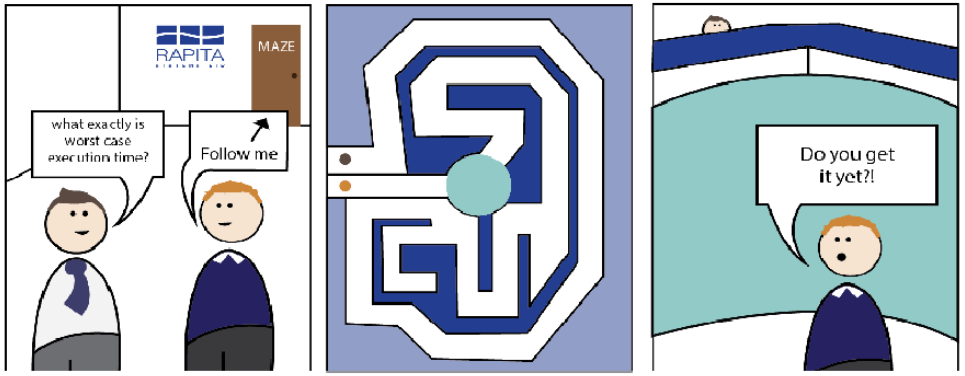
Worst Case Execution Time (WCET) using AI/ML

What is the WCET?

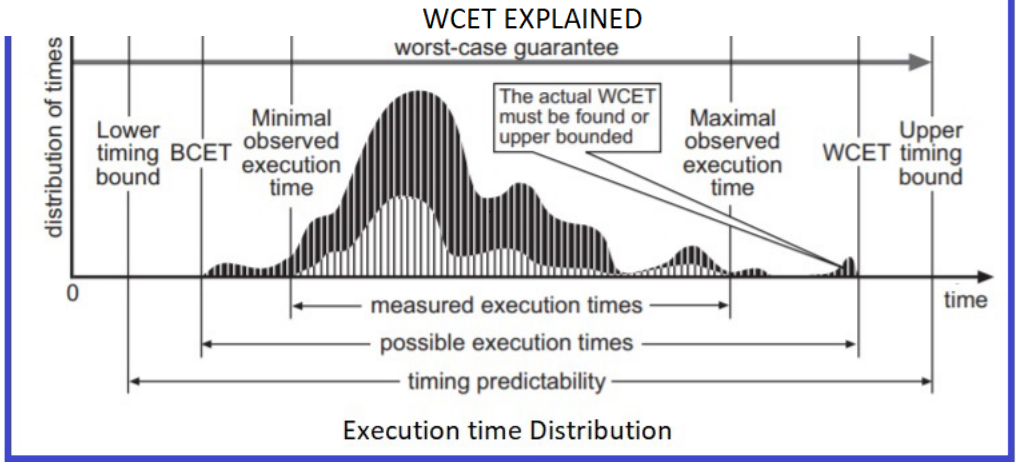
It is the MAXIMUM LENGTH of time a task could take to execute on a specific hardware platform.

WCET depends on:

- **type/quality of hardware**
- **interactions**
- **hidden state of the hardware**
- **environment**



Source: https://www.rapitasystems.com/worst-case-execution-time



Complex Digital Systems, A11D.SDS.6

Safety Verification Framework for Learning-based Aviation Systems (SVF-LAS)

Prime Performer: UWV, GWU, Honeywell

Status: Start 09/01/2021

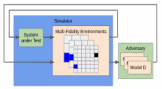
Planned End Date 08/31/2023

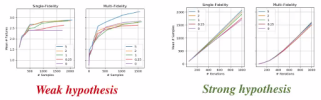
Major Accomplishments:

1) Developed and Began Tests of SVF-LAS Software
2) Published and Presented Paper: *"A Verification Framework for Certifying Learning-Based Safety-Critical Aviation Systems"*, Baheri, A., Ren, H., Johnson, B., Razzaghi, P., and Wei, P., 06/20/2022, AIAA 2022-3965

Thrust I: Design-time assurance using multi-fidelity reinforcement learning approach

- Adaptive stress testing (AST) success story in aviation: **ACAS-X** (replacement of **TCAS**)
- **AST limitation:** fails to find particularly sparse failure.
- **Goal:** Extend AST to multi-fidelity setting to verify two hypothesis:
- ***Strong hypothesis:*** could we *accelerate* the testing process?
- ***Weak hypothesis:*** could we find *more (and potentially rare)* events?
- **Takeaway:** proposed multi-fidelity AST verifies **BOTH** hypotheses*





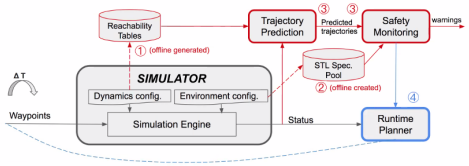
Weak hypothesis *Strong hypothesis*

low-fidelity sim high-fidelity sim

*Submitted to IROS 2022

Thrust II: Runtime Safety Monitoring

Overview: apply reachability analysis of vehicle dynamics and control to generate reachability matrix regarding the current states, inputs, and environmental uncertainty, to enable online bounded-time trajectory prediction in 3D space through runtime querying.



Thrust III: Development of hardware testbed

The experimental platform: Tarot T18 Octocopter

Schematic diagram of the electrical components

Thrust III: Use case and Test Field

- The test field is the Northern Virginia Radio Control (NVRC) Poplar Ford Park Field, which is a model airplane field located in Centreville, Virginia
- One of the use case scenario (Merging)
- The ownship aircraft flight path denoted with white points. The intruder flight path indicated by red points. The merge point is the yellow point. The blue points show the merged path.
- The DRL model will adjust ownship aircraft's speed by updating the flight plan via GCS interface.

Discussion and Wrap-up

Srini.Mandalapu@faa.gov

Backup Slides

SAS Questions from previous meeting
and
FAA responses

Complex Digital Systems, A11D.SDS.6

SAS members' Questions and Answers

Question: How can errors (from AI system or human interaction with the AI system) be identified, mitigated and managed in AI/ML based systems? 25.1302 requires operationally-relevant behavior of the installed equipment must be predictable and unambiguous. How can this be assessed considering the probabilistic nature of the new systems?

Response:

The Intended Behavior is captured in the requirements. The system will be tested against the requirements. Research is required to better understand the generalization required with limited data training, balanced against the distribution of 'real-world' behavior.

This is a challenge for AI/ML systems' implementation and assurance. Depending on where it is implemented with what criticality level, the training data requirements, validation of the models, and explainability aspects must be considered and stringent assurance objective must be developed. One consideration of implementing a wrapper (runtime monitor) is getting traction in the standards community when implementing in safety-critical systems to mitigate a unsafe condition.

Complex Digital Systems, A11D.SDS.6
SAS members' Questions and Answers

Question: Are Software Assurance/Development Assurance compliance protocols maintaining relevancy as technology advances and product complexity increases?

Response:
Research in New Methodologies is proposed. OPs concept is one of the answers to address this concern.

This has been the focus of Digital System Safety research considerations. Several of the research tasks have been considered to explore the new technologies (including AI/ML), study the safety implication, identify the gaps in current standards and guidance, and develop recommendations or new assurance criteria to address the gaps.…

E.g..1. Model-Based Systems Engineering and Model-Based Safety Analysis:
2. Use of Virtual Machines in Avionics Systems and Assurance Concerns
3. Requirements Modelling to address complexity (continuing)…

Complex Digital Systems, A11D.SDS.6

SAS members' Questions and Answers

Question: Section 2.1 - the term "non-prescriptive certification approach" is used, but not sure about the context. Does this mean don't use DO-178 or do you mean not being focused on software development anymore and instead look at architectural weakness and results?

Response:

It means taking a holistic view of the entire system using risk as a measure. It should be possible to establish a blend of evidence, some based on DO-178, and some based a convincing argument that safety is maintained through other means, e.g. safety monitors, voting logic and so on.

DO-178 is an industry consensus standards for Software Assurance. FAA does not mandate the use of DO-178C and allows other means of compliance. There is a general observation is the DO-178 is too prescriptive and it is very expensive. FAA accepts DO-178C an acceptable means, but not the only means (AC No: 20-115C). Architectural weakness shall be addressed through ARP-4754A, RTCA DO-297 at system level.

Complex Digital Systems, A11D.SDS.6

SAS members' Questions and Answers

Question: What the new human-factors issues that can be expected due to increasing autonomy associated with the transition (advisory → assistive→ responsible)? For example, shared authority between pilot and aircraft automation will require new ways of mode awareness, transitions, hand-offs and automation transparency. There will also be a need to design to maintain pilot engagement. The research should provide guidelines for design, operations and certification.

Response: Valid concern. I guess this is a human factors issue and should be directed to that group.

Human Factors is covered by other BLI's.

Complex Digital Systems, A11D.SDS.6

SAS members' Questions and Answers

Question: Section 2.0 - should there be better definitions or boundaries being addressed between software development and hardware performance? .

Response: Agreed with the concern. Current SAE G-34 standards identified this concern. The following is an extract from AS6988:Artificial Intelligence in Aeronautical Systems: Statement of Concerns, Section 4.2.2 Development Life Cycle Activities:

"The implementation phase focuses on the software and/or hardware development activities. It consists of designing, implementing, and verifying the trained NN in the target environment. A trained NN has passed acceptance criteria and represents system requirements allocated to software or hardware implementation. Special cases will be considered where it may apply, particularly in a model-based approach."

Development of AI/M-L models is often done using huge computing resources that learn to generalize based on the behavior of the AI model in a test environment. The training process may use very many parallelized processors to adjust the neural network data including the structure of the network itself (adjustment of hyper-parameters). The training network may be different than the inference network (which is optimized for performance). The differences between these models needs to be understood and validated.

Complex Digital Systems, A11D.SDS.6

SAS members' Questions and Answers

Question: Section 2.2 – how does the FAA plan to embrace lowering the costs of implementing safety related systems without compromising safety. What platforms would this be applicable or acceptable, i.e. only in low-risk operations? .

Response:
Currently we use 4 Design Assurance Levels with DO-178, but FAA is promoting a safety continuum which is based on risk levels, and possible mitigations that can be adjusted with much finer granularity. Using DALs at intermediate levels would lower costs. We need to understand how this should be done.

FAA's goal is to assure safety. All our efforts are focused on improving safety. To address the costs, FAA would allow the alternate means compliance. Our research on OPs is to streamline the assurance process. There are some efforts outside the research organization to address this concern (Abstraction Layer).

Complex Digital Systems, A11D.SDS.6

SAS members' Questions and Answers

Question: Is the R&D going to focus how software is developed or how it is evaluated once software/hardware is developed and implemented? .

Response:. As the title of this BLI, "Complex Digital Systems", indicates that our research considers systems, software, and hardware. It is tough to isolate them from the safety and assurance perspective. They all go in hand in hand.

It is up to the applicant to propose the means for validation and verification. We want to provide the applicant the flexibility to propose the means, but want to ensure they do their work in a framework that allows the FAA to evaluate and find the evidence acceptable.

Complex Digital Systems, A11D.SDS.6

SAS members' Questions and Answers

Question: I don't have a lot of specifics to offer on the AI/ML piece as I do not have the expertise in software certification. However, there's a difference between using AI/ML to help design airborne software, which can then be certified using traditional means, and using AI/ML on the aircraft itself (more difficult to have assurance.). That piece seems to be missing from the write-up in 3.0 (e.g. maybe there's a crawl-walk-run approach here?)

Response: The reason we are putting in lot of efforts in to AI/ML assurance because the traditional standards, policy and guidance are not adequate. Some of current and proposed research will address this concern.

We have not had any AI/M-L based software help with the certification of on board software. If such tools show up (for example a tool to generate software tests) then they will be covered through DO-330 (Tools supplement) In general showing that the tools cannon conceal an error in the operational software is hard, and showing that the tool cannot insert and error is harder. The applicant would need to provide qualification evidence depending on the Tool Qualification Level, as described in the Tool qualification supplement.