Standards for AI/ML and Emerging Technologies

Natasha Neogi
System Wide Safety Project
November 28, 2023

Organization Perspective

- Standards development activities require a keen and deep understanding of the problem being solved by the standard as well as the technologies being deployed in any reference implementation of the solution.
  - It is important to understand the mechanisms and limits of the fundamental, underlying science of implementation and verification technologies used to realize and assure systems.
  - We need to understand the limits of what current process and metrics can provide with respect to new technologies.
- US leadership is important in this endeavor, and it is vital that we have a measured approach that yields sound results
  - First versus Best
- Start with simple, well-defined, non-safety critical applications
  - Recognizing normal and anomalous patterns in large datasets for research purposes (collaborative), Querying large databases for research purposes (ASRS/ASAP), etc.
- Progress to functions which have (1) clearly defined requirements, (2) means of checking the answer/output, and (3) means of intervention and mitigation of incorrect answers/outputs

Key Questions for Using AI/ML (I)

- What constitutes sufficient evidence that an AI/ML system's behavior meets its requirements?
  - Sufficient representation and size of training dataset, accuracy vs. generalizability, what constitutes an actionable specification
- Can an actionable specification for a function be extracted from a dataset?
  - Functional Requirements, Safety Requirements, Environmental Assumptions, Domain Specific Constraints, etc.
- What are the set of characteristics and parameters of an AI/ML system that allows you to bound its behavior (e.g., capabilities, limits, etc.)?
  - Data and information quality, architecture, associated metrics, etc.
- What are the limits of current processes and metrics currently used in developing and evaluating both traditional and AI/ML systems?
  - How do you use testing (i.e., creating logical based oracles, etc.), simulation (i.e., model validity), (formal) analysis (i.e., scalability), runtime verification frameworks, etc. in assurance

Key Questions for Using AI/ML (II)

- When is it appropriate to use an AI/ML implementation for a function?
  - Clear (and testable) set of requirements, outputs easily checked for correctness, corrective action can easily be taken, etc.
- What is the current human contribution to safety in the function being replaced by an ML/AI implementation (i.e., full extent of the capabilities and limitations of the human role)?
  - Consider critical information dependencies across tasks executed collaboratively by diverse agents
- Can the open world problem be solved (and standardized) without humans to handle edge cases while maintaining the current level of safety seen in the NAS?
  - Handling epistemic uncertainty, applicability of real-world data across different environmental assumptions, etc.
- What is the minimum set of information required to reconstruct and audit AI/ML performance in the case of an accident?
  - State, Environmental, and Input Information, Decision Making Logic, Configuration Management, Version Control, etc.
  - Avoid attempting to extrapolate understanding and assumptions across agent types

Key Questions for Using AI/ML (III)

- What are key domain specific considerations that may dominate the safety of AI/ML implementations and how will we address them?
  - Lack of safe default mode/state, inability of pilot to intervene, etc.
- How can change be managed in AI/ML systems in order to preserve assurance?
  - Configuration Management, Version Control, database management,
  - Full recertification, continuous authorization to operate, etc.
- How should information assurance be handled for AI/ML systems in order to yield (composable) safe systems?
  - Data fusion; information synthesis; data collection, curation, and assurance; etc.
- When is it appropriate to use AI/ML in the development and/or accident/incident analysis process?
  - Tool qualification (DO-330), ASAP/ASRS database querying for research, prognostics/diagnostics, scheduling, maintenance, etc.

Call to Action

- A measured approach to standards development for AI/ML components should target those functions for which there are actionable specifications and traditional implementation and assurance techniques
  - Develop criteria for what constitutes sufficient evidence for AI/ML safety
- Identify current and ongoing standards efforts that may be applicable (e.g., DO-330, etc.)
  - Harmonize with other standards bodies when appropriate (e.g., domain and DAL) to avoid duplicative efforts (ASTM, UL, SAE, etc.)
  - Standards efforts should be targeted at areas in which gaps are found.
- Identify the role of community-based practices in creating a level-set for the Aviation AI/ML ecosystem

Conclusions

Standardization efforts should proceed methodically and with a justifiable basis, thereby enabling safe adoption of AI/ML technologies in Aviation.

Premature efforts to standardize may damage paths to transition for AI/ML technologies, engender technical debt, or set back the entire aviation industry.