Comments to HAV Federal Automated Vehicles Policy

Jon Hagar, Grand Software Testing, LLC, Colorado USA

IEEE and ISO Lead Editor/Author for ISO 29119 Software Testing, and IEEE 1012 V&V standards

Object Modeling Group (OMG) co-chair: UML Testing Profile (UTP 2.0) for Models

Book Author: Software Test Attacks to Break Mobile and Embedded Devices, and IoT Dev-Testing

Contact information: jon.d.hagar@gmail.com or 303-903-5536.

Comments dated: Oct 2016

Comments made regarding: version Sept 2016

File Attached

Executive Summary

The following comments are made on Federal Automated Vehicles Policy for highly automated vehicles (HAVs). Comments are provided on specific suggestions and then general comments to the overall policy. The focus of my comments is from a Validation, V&V, and test practice viewpoint, which includes: systems, hardware, and software considerations. Suggestions are made for the national level but have application to international areas, since the HAV market will eventually be global in nature. Information on reference materials and follow up recommendations are included at the end of this paper.

This work was conduct by Jon D Hagar, CEO/consultant for Grand Software Testing. Questions can be directed to jon.d.hagar@gmail.com or 303-903-5536.

Introduction

The policy represents a good early step for USDOT. The document contains clear statements on safety (Clause 3 and throughout), public concerns and industry goals. The USDOT effort to seek industry and public comments is to be commended as to what will be a society-changing technology. While not addressing the societal issues such as job market displacement and major shifts in the business sectors, the policy is focused on how government, the public, and industry should safely proceed with HAV.

The primary focus of the remaining feedback in this paper is on the USDOT HAV policy focused on Validation, Test/V&V. Below are specific comments by: page, a policy quote, and suggestion/consideration. Following these specific items are general comments and inputs for reviewer considerations.

Specific comments by page, quote, issue, and suggestions

Page 14:
"Finally, as shown in Figure I, tests should be developed and conducted that can evaluate (through a combination of simulation, test track or roadways) and validate that the HAV system can operate safely with respect to the defined ODD and has the capability to fall back to a minimal risk condition when needed."

Comment: The statement of "combination of simulation, test track or roadways" needs refinement and clarification. The definition of simulation must include (or be expanded to include) test lab environments with tools such as emulation, hardware in the loop labs, modeling, analysis, structural test techniques, low level functional testing, integration testing, system conops testing, etc. driven by experience, risk-based, and other test strategies addressed in test planning (ref ISO 29119 and others). The space, aircraft, nuclear, and medical industry all have history and extensive levels of validation, verification (V&V), testing, analysis, and inspection in lab settings. The driving idea is that in labs and simulated environments many more tests can be achieved compared to test track and roadways. The definition of test environments may belong in a lower level policy or USDOT standard, but as currently stated the words may be open to lesser levels of testing. While test track and roadways are critical, the use of many different engineering and test techniques can be achieved during earlier levels of Validation.

Note: the policy should refer to definitions for Validation, Verification, Quality Assurance, and Testing since there are different uses of these terms in industry. For consistency, I recommend following ISO and IEEE (see reference section below).

Page 16

"For HAV systems deployed on public roadways for testing or production purposes, the Agency envisions that manufacturers and other entities will likely update the vehicle's software through over-the-air updates or other means."

Comment: The subject of software updates can be a little contentious. Currently, many embedded systems in cars must be updated at a dealership, and usually customers are not fully aware of the change. In a networked (IoT) environment the ability of manufacturers to "push" an update may be resisted by some consumers. The flip side, that if an opt in/out approach is taken by manufacturers and customers critical features may not be implemented. This would create a software configuration management (SCM) problem which could be a serious problem causing issues with the vehicle or worse cause a life-threatening situation to happen. So, if opt in/out happens then the consumer will figure out ways to defeat updates, with the same SCM problems resulting. Further, the short quote does not address current fundamental problems such as fragmentation of configuration (reference Android market), which requires more testing and analysis. This quote and area should be the subject of more scrutiny and research.

Page 31

Validation Methods

"Given that the scope, technology, and capabilities vary widely for different automation functions, manufacturers and other entities should develop tests and validation methods to ensure a high level of safety in the operation of their HAVs. . . ."

Comment: This entire Clause (all paragraphs) needs to be the subject of research and discussion by experts. Current industry standards such as IEEE 1012 and ISO 29119 should be used as starting points for the expansion and clarification of this clause. IEEE 1012 has a long history of usage for system, hardware, and software V&V. A variety of industries have used it. ISO 29119 has a shorter use history, but is based on international and IEEE standards, which do have historic usage. The use by USDOT of industry standards will better support international considerations.
Note: any used standard should serve only as a starting point with large degrees of tailoring by industry experts for use by USDOT. USDOT should expand beyond Federal Aviation Administration (FAA) practice, which, while robust, predates aspects of historic standards and the FAA does not address level 5 HAV (maybe level 4 too). In consulting with many organizations, historic test ideas f have not been practiced and state of the art ideas (not in standards yet) were also underused. Experts at groups like the U.S. National Institute of Standards and Technology (NIST, universities, and industry projects should be engaged by USDOT.

Page 77

"Tools a. Tool I: Variable Test Procedures to Ensure Behavioral Competence and Avoid the Gaming of Tests
For several reasons, variations in test environments are sometimes necessary to accomplish the purposes of the Vehicle Safety Act."

Comment: Historically, many critical industries have used a concept called Independent Verification and Validation (IV&V). IV&V should be considered as a "tool" for the above quote. IV&V has worldwide recognition (e.g. Malaysian Government Policy) when there are concerns for "gaming" of tests. Additionally, as the quote indicates, USDOT policy and standards can be put in place to minimize "gaming". However, as seen in the VW emission scandal shows, IV&V may be good "policy" to have in place for HAV level 4 and 5 systems (or lower). IV&V would be in addition to other checks and balances.

Page 78

"b. Tool II: Functional and System Safety "
Comment: This title and Clause words seems to indicate that only the qualities of functionality and safety are critical to HAV. ISO and other standards recognize many other non-functional qualities of hardware, software, and systems, which are vital. For example, qualities performance and reliability are different than safety but have been shown to impact the overall system, resulting in loss of functionality and/or safety. I recommend this Clause be expanded to recognize most (if not all) of the qualities of software/systems, particularly for higher levels of HAV. The engineering, analysis, development, and validation of the qualities should be included in the policy to achieve the stated ones of "functional and safety").

Page 79

"Tool III: Regular Reviews for Making Agency Testing Protocols Iterative and Forward-Looking "

Comment: This should be an important research area. Detailed conops, use-case test modeling, and test procedures with open architecture of data driven keywords could be very useful. Such frameworks should be open to the public and researchers. Further, as data from HAV systems is gathered, data analytics should employed to define fielded errors and missed testing protocols. This can be done with a test support technique known as "error taxonomies," which results in patterns of test tours and attacks (reference work by James Whittaker). The data is likely to grow large, so USDOT or support agencies may need to use analytic concepts such as deep learning, AI, and statistics to "mine" the HAV big data. Note: this may run into issues of individual and company privacy as well as industry intellectual property, which the US government would need to overcome.

Page 80

"Tool IV: Additional Recordkeeping/Reporting"

Comment: The current and most complete standard for complete test lifecycle documentation is ISO/IEEE 29119- 3. This couldbe the starting point for USDOT, though likely subsets and tailoring of part 3, which is comprehensive, should be done by industry experts.

Page 95-96

"The Agency focused on the FAA because its challenges seem closest to those that National Highway Traffic Safety (NHTSA) faces in dealing with HAVs. FAA uses an agency pre-market approval process 116 to regulate the safety of complex, software-driven products like autopilot systems on commercial aircraft."

Comment: This is true, but only with HAV SAE levels 3 or 4. Nobody has standards or common practices for man-rated level 5 systems. Similar systems exist in space industries that function without human control for long periods of time due to factors such as distance or speed. Further, the nuclear industry, which has humans in the loop, has had to deal with catastrophic risk factors (1,000s of people might be hurt by a single software-system fault). It is recommended that for level 4 and certainly 5 USDOT "raise the bar" in Validation/V&V/testing and IV&V.

Additionally, cited later is the "Boeing 787 Dreamliner", but this "overall system" has several software bugs that were detected after it went live. The plane still suffers apparent "interface integration" issues between elements of its systems even after "200,000 hours of FAA staff time". "Self certification" may be the way to go, but much care and research is need. HAV development is going further than aircraft development has gone. Thus, there have already been recommendations for IV&V, more levels of Validation, and engineering for the higher levels of HAV.

Several pages

The inclusion of policy updates, evolution, and being "agile" is why this comment paper is generated. Clause 11, Ethical Considerations, should evolve and be the subject of continuing research, particularly in regards to Clause 2 on page 28. Normal driving test cases will need more cases, conops (concept of operations see https://en.wikipedia.org/wiki/Concept_of_operations), cases, and refinement. The expansion of these sections is needed for Validation/test and certification efforts, which are referenced throughout the policy.

General comments

1. The definition of "Validation" should include Verification Validation (V&V), test, review, inspection, analysis, demonstration, modeling, etc.

2. USDOT policy is good start, though much work remains over the coming years as these systems evolve.

3. Measurements at a high level are defined (death and accident rates), and these are the ultimate "top lines," but many other measures should be considered and defined by experts for lower levels e.g., reliability, MTBF, error taxonomy, critical sensor measurements, etc. Smart and Internet of things (IoT) systems, which HAV will be classified as a part of, will be "driven" and evolve using data. Identification, recording, reporting, and analysis of data will be critical and cannot be done solely by companies. On page 18, the policy states "Data sharing is a rapidly evolving area that…." and then "**e.** Tool V: Enhanced Data Collection Tools "on page 80. I agree that measurement, analysis, and action based on data will be critical to the general evolution of the USDOT policy and any lower level statements. Likely teams or researchers should be established in parallel to analyze and use the data. The *independence* of such data analysis will be important given industry concerns with competition and costs. Such independence is a clear role of support by USDOT.

4. Evolution recommendations for the policy

    a. Lower levels of policy statements and/or standards are needed for HAV from USDOT

    b. Reference and/or compliance to national and international standards

    c. Work with standards groups such as IEEE and ISO

    d. Include model-based engineering and testing

    e. Expand test strategies and planning elements in the policy (or a new standard)

5. Clause 2 Privacy and Clause 4 Vehicle Cybersecurity are critical areas and should be subject to funded research. As a practitioner/author in this area, I have been approached by researchers and we agree there are more questions than answers in the privacy and security of HAVs.

Particularly in regards to Clause 9 post-crash behavior, as this may deal with when privacy stops for the public.

6. Product certification as defined in Clause 8 is a start for certifications, but much more detail is needed. Example questions include:

    a. What will be the legal implications of an HAV's certification?

    b. Who issues a certification and what exactly are they certifying: e. g. one or more of the HAV's systems or the overall vehicle

    c. How will the policy assure proper HAV level classification of a vehicle for a certification?

    d. Will the consumer understand what they are getting in each level of certification? Note: Table 1 on page 34 is a start but more detail in lower levels is needed there.

7. The words "best practice" are used several places in the policy, but it is not clear to many experts what a "best practice" really is. Best practice is not universally defined nor do standards define "best" or "state of the art practice." Rather than using these words, USDOT should define in a policy or standard what they require for a "practice."

8. The definition of Operational Design Domain (ODD) and specifications of test environment with the associate conops, needs to be a near-term research and definition effort. Further, how is ODD information and operational "fail safe" status is communicated to user needs to be better defined

9. The policy calls out a "fall back strategy" that manufacturers need to put in place. This should include developers defining the levels of: fail safe, safing software systems, and battle override modes. These concepts are well known in the space industry and in books such as "Safeware," (N Levenson, 1989) with emphasis on developing "safe" and not trying to "test it in" to design.

10. Details on "Tools," as the policy calls them, needs to include:

    Test Environments

    Levels of testing (white box to black box).

    Life cycles (traditional, iterative, agile, etc.).

    V&V/test approaches; planning, strategy, methods and techniques all need to be in a standard for reference by the industry.

Reference and Expert recommendations for follow-on updates, working groups, and/or research

Experts support should be provided to the US government by members of different V&V/testing communities, because there are differing views on V&V/testing which should be considered in a modern policy. Communities to include should include experts from:

Industry organizations: ISO, IEEE, SAE, and OMG

Test/QA: Association of Software Testing (AST), ASQ

Academic researchers: major universities

Selected experts from these areas would support working groups and policy updates (as already noted in the policy).

Research reference recommendations which should be added on page 69 Clause B, include:

1) Standards to include: IEEE 1012, ISO 29119, ISO 12207, ISO 15288, OMG UTP 2.0, and ISO 26262
2) Definition of life cycle test documentation for audit, certifications, and regulatory usage.

   Note: Current IEEE and ISO recommendation is ISO/IEEE/IEC 29119 part 3.
3) Improve the HAV policy and supporting standards beyond current Food and Drug Administration (FDA) and FAA regulations.
4) Address industry-defined engineering models using conops, detailed test scenarios, test attacks, test tours, test levels, etc. (see item 1 above)
5) Reference materials such as test books, experts, conferences, and web sites provided to USDOT on request.

-----------------------End------------------