IDA | Research Insights

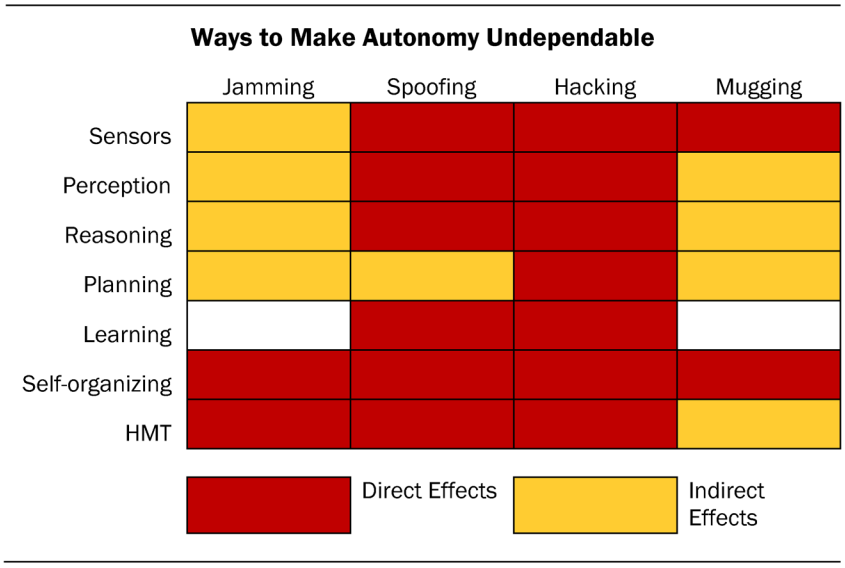
May 2019

Assured Dependability for Autonomous Systems

David Tate (dtate@ida.org)

Sequential diagnostic, operational, and acceptance testing is the normal approach to assuring system developers and users that any kind of system will perform its mission safely, effectively, and reliably. The dependability of an autonomous system, however, depends on the system's decision processes, which interpret sensor data, model the environment, consider mission goals and priorities, choose courses of action, observe outcomes, and potentially modify the system's own logic over time through post-fielding learning. **The normal testing approach cannot possibly effectively test, evaluate, verify, and validate system behavior in every decision context an autonomous system could face**. A different approach is needed to assure developers, operators, and commanders that autonomous systems will perform dependably in situations that may differ significantly from any that were tested explicitly prior to fielding.

Autonomous systems rely on successful integration of many enabling technologies to be dependable. These technologies can include computer vision, sensor fusion, knowledge representation, expert systems, inference engines, path planning, optimization, machine learning, and others. Autonomous systems that team with humans also depend on detailed concepts of operations for how the humans and the machines will interact. All of these represent ways a system's dependability could be threatened; the inputs to each enabling technology generate a novel attack surface, in addition to the usual cybersecurity attack surfaces of advanced systems.



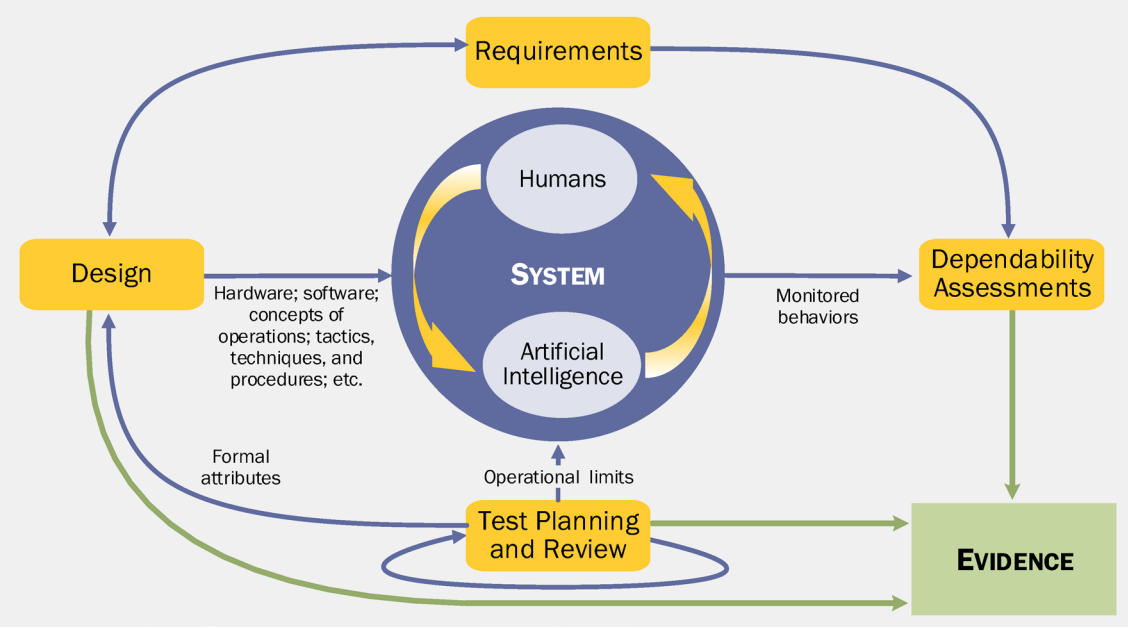
For establishing dependability, the environment might as well be an adversary. All of the attack surfaces of autonomous systems can lead to undependable behavior (see matrix at left). Attacks could be generated by an adversary or by a complex environment and could involve denial of information (jamming), misleading inputs (spoofing), unauthorized control (hacking), or threats of physical harm (mugging). **It is impossible to test a system against all possible attacks, but it *is* possible to accumulate evidence over the course of a system's development and operation, providing a time series of evidence of increasingly dependable behavior. This approach is called *evidence-based licensure* (EBL)**, and it offers the best hope of fielding systems with assured dependability on reasonable time scales.

(continued)

NS D-10669

EBL develops arguments for when a system can be expected to be dependable. Essentially, a certifying authority would license use of a system within defined operational parameters identified during testing. In some situations, only certain human-system combinations would be certified because they work together dependably. For example, car and driver teams are licensed to operate within certain restricted parameters. The car's commercial roadworthiness certification doesn't certify it for off-road military operations, and the driver's car license doesn't extend to motorcycles or big rig trucks.

Collecting the needed evidence for EBL requires a radically different approach to test design and instrumentation. (See diagram.) First, the distinctions among systems engineering, developmental testing, operational testing, acceptance testing, and post-deployment reassessment would need to be eliminated. Further, test instrumentation would need to support continual comparison of system behavior—including internal cognitive behavior—against specific goals at all levels. For example, to assess compliance with the goal to not crash into trees would require instrumentation that allows testers to distinguish between failing to see the tree, failing to recognize it as a tree, failing to understand that trees are impassible, failing to plan the correct path to avoid the tree, or choosing to hit the tree to avoid some other undesirable outcome. In other words, the system's *thinking*, not just the system's performance, must be measured and assessed. Certification bodies could then be confident that the system will behave in dependable ways, even in situations not specifically tested, because it is consistently *acting for the right reasons*.



EBL is an extension of what is already done for safety and cybersecurity assurance. Defense organizations concerned with safety and security of defense systems, such as the Joint Weapon and Laser Safety Working Group and the Joint Services-Software Safety Authorities, operate essentially as certifying bodies. Currently, their activities are not well-integrated into system development or other test and evaluation goals. Achieving the integration of all aspects of dependability testing will be a major cultural and organizational challenge for defense acquisition, but will be essential to rapid and effective fielding of nontrival autonomy in military systems.

Based on IDA P-5325, *A Framework for Evidence-Based Licensure of Adaptive Autonomous Systems*, D. M. Tate, R. A. Grier, C. A. Martin, F. L Moses, and D. A Sparrow, March 2016; IDA NS D-8982, *Acquisition Challenges of Autonomous Systems*, D. M. Tate and D. A. Sparrow, April 2018; and IDA NS D-10523, *Attack Surfaces of Autonomy*, D. M. Tate, March 2019. Research sponsored by the Air Force Research Laboratory and the Office of the Assistant Secretary of Defense for Research and Engineering.