EXECUTIVE SUMMARY

After a fatal Uber crash in Arizona last year took the life of a pedestrian, followed a few days later by a Tesla crash that killed a driver in California, it’s fair to ask: Shouldn’t highly automated vehicles be required to pass minimum standards for safety, reliability and performance before these cutting-edge prototypes take on public roads?

Researchers at the University of Michigan’s Mcity believe such testing is crucial to the development of highly automated vehicles, and they have developed a testing concept with the potential to emerge as a voluntary standard for safety testing. When fully developed, the Mcity ABC Test could provide developers and the general public with the necessary level of trust that is vital to making automated vehicles the life-changing reality we know they will become.

Like any other software-driven systems, automated vehicles can be improved continuously to make them safer, more dependable and more efficient than human drivers. But perfecting self-driving cars and trucks requires extensive testing – and only so many tests can be run in the lab or on test tracks. At some point, automated vehicles must be subjected to the ultimate trial of traveling on public highways. This is especially
true in the first phase of automated vehicle deployment, when self-driving cars and trucks will interact primarily with human-piloted traffic instead of other automated vehicles.

This is why Mcity is developing a test-track-based procedure that promises complete and comprehensive testing of autonomous vehicles. Mcity researchers call this protocol the “Mcity ABC Test,” a highly promising three-part procedure that includes:

- Accelerated evaluation covering the driving scenarios responsible for the most common motor vehicle crashes
- Behavior competence testing to ensure the automated vehicles demonstrate they can handle the majority of driving scenarios; and
- Corner cases, or test situations that push the limits of these highly advanced automated vehicles.

BACKGROUND

The big stumbling block: Lack of trust

Automated vehicle technologies continue to be developed at an astonishing pace. Despite the many, many advances that are taking place, public opinion about automated vehicles plummeted in the aftermath of the fatal crashes involving a Level-4 Uber prototype vehicle and a Level-2 “Autopilot” Tesla vehicle.

The Uber crash of March 18, 2018, was the first fatal crash involving a Level-4 automated vehicle prototype. Two things in the preliminary report of the National Transportation Safety Board (NTSB) [1] stood out: First, that the automated emergency braking system – a “lower-level intelligence” feature available on the test vehicle – was disabled; and, second, that the on-board sensors successfully detected the doomed pedestrian six seconds before the woman was hit and killed, but the car’s perception algorithm struggled to identify her. At first, 49-year-old Elaine Herzberg was classified as an unknown object, then as a vehicle and finally as a bicycle. Finally, 1.3 seconds before Ms. Herzberg was hit, the vehicle system decided to brake, too late to avoid tragedy. Investigators later determined that the car’s safety operator was watching videos on a smartphone a few seconds before the accident.

Statistically, human errors are a contributing factor in 94 percent of the approximately 100 fatal crashes that happen every day in the United States. But only the Uber crash captured media attention that day last March.
The fatal Tesla crash just five days later was very different. The Tesla Autopilot system involved in that accident was a Level-2 automated driving function, where the human driver is supposed to pay attention all the time. Unfortunately, naming the Tesla system “Autopilot” misleads many drivers into believing the system is more capable than it really is. The preliminary NTSB report for this crash [2] found that the vehicle did provide visual and auditory alerts but that the hands of the driver weren’t always on the steering wheel. Despite the lack of driver attention, the Tesla system didn’t shut down.

In addition, Tesla designers didn’t geo-fence the operation of the car’s Autopilot function, unlike others. General Motors Co.’s SuperCruise system, as an example, is not available outside a geo-fenced area. Finally, the Tesla system was performing only a low-level “lane following” function instead of being integrated with the vehicle’s navigation function. As a result, the car didn’t go into either the left or the right lane, and instead headed directly into the crash attenuator of the concrete barrier that divided the lanes of traffic.

When not one but two automated vehicles each kill someone there naturally is a lot of media coverage, an NTSB investigation and analysis by researchers. Both of these fatal crashes were tragic and historical and rocked the development practices for highly automated vehicles. These fatal accidents also spiked the general public’s doubt about the safety and maturity of self-driving cars and trucks. Several recent surveys [3][4] found that half of respondents said they don’t trust driverless cars. A JD Power survey [5] found that 40 percent of those surveyed said they “would not ride in an (automated vehicle) regardless of what progress is made.” This understandable lack of trust toward automated vehicles presents a major problem for the development and deployment of these vehicles.

DEFINING THE CHALLENGE

How to regain trust?

Mcity researchers believe highly automated vehicles can have a major role in providing safe, convenient, efficient and sustainable transportation in several forms. Looking back at the doubts that surrounded the Wright Bros.’ “flying machine” more than 100 years ago, we see the same skepticism pointed at automated vehicles now. History demonstrates that trust isn’t given – it’s earned over time. Accordingly, automated vehicles must be developed in a responsible way to gain the trust necessary to fully realize their potential as a life-saving, life-changing technology for society. To win that trust, Mcity researchers
recommend that automated vehicle developers do two things.

**FIRST: Communicate and follow a clearly defined operational design domain**

The vision of “anytime, anywhere, any weather, any speed, by anyone” automated vehicles that has been promoted to the public is greatly at odds with the limited operational design domain being pursued by vehicle developers. Many of the Level-4 vehicles on the road today operate at low speed (for example, 12 mph for the Mcity driverless shuttles, 25 mph for May Mobility vehicles in Detroit). They operate in geofenced areas that are carefully selected, instrumented, pre-mapped and tested. These vehicles cease their operations in bad weather, are professionally maintained and always have a safety conductor on board.

Most automated vehicle developers and operators know and strictly follow the operational design domain concept – but this concept isn’t being clearly communicated to the news media and the public. Statements from some companies claiming things such as, “Our Level-5 autonomous driving feature is just two years away,” are counter-productive and a major factor creating doubt and mistrust around highly automated vehicles. Mcity researchers believe that in the next few decades, automated vehicles will have limited deployment, as they do now, remaining at Level-4 technology. Level-5 automation is a long-term opportunity.

Another benefit of using a clearly stated, responsible design domain is that the development teams can work within well-defined and bounded development and validation targets that reflect a realistic engineering challenge.

At this point, we must address why we at Mcity believe that “proving an automated vehicle 100 percent safe” is an impossible task. While some consumer surveys have offered that as a realistic choice, and many survey respondents selected that as their top choice, if “no crashes” is what people mean by being 100 percent safe, meeting that standard will be impossible. A crash can be primarily caused by the action of another road user, which is beyond the control of the vehicle. But that doesn’t mean developers can’t and shouldn’t design tests where automated vehicles respond to bad and aggressive moves from others.

**SECOND: Develop and follow comprehensive and carefully-designed safety tests**

Before 2018, no highly automated vehicles were offered on the market. Instead, the HAVs running on public roads were maintained and operated under pilot programs by Waymo, Uber, Cruise Automation, Argo.ai and others. The U.S. Dept. of Transportation decided
against imposing rules and regulations too soon, but published voluntary guidance in 2016 and 2017. In the 2018 “Automated Vehicle 3.0” document [6], the department mentioned that “performance-based safety standards” may be developed for future vehicles. Less than one week later, the National Highway Traffic Safety Administration (NHTSA) announced an advance notice of proposed rulemaking[7] for a pilot research program for Level-4 and Level-5 vehicles that also mentioned “future performance-based standards and testing to ensure safety.” This implies that some form of minimum-performance safety testing will be required before these vehicles are included in the public roads pilot program– similar to requiring human driver’s pass the test for a license. We think this is an excellent idea to balance public safety while leaving room for innovation.

What would an independent minimum-performance safety test look like? This is the main question we aim to answer with the Mcity ABC Test.

**PROPOSED SOLUTION**

**Goal and design attributes of the proposed Mcity ABC Test process**

The Mcity ABC Test is our concept of an independent safety assessment for highly automated vehicles. The ABC test uses a closed test-course process before automated vehicles could be deployed on public roads.

**Goal**

The primary goal of the Mcity ABC Test is to assess the safety performance of highly automated vehicles. It’s important for these vehicles to demonstrate “roadmanship” [8]. That means displaying behaviors that are similar to those of human drivers, being neither too aggressive nor too conservative compared with other drivers, but also showing a level of assertiveness in congested driving conditions so that the vehicle doesn’t take forever to get somewhere. However, roadmanship is a vaguely described concept and is secondary to requiring a vehicle to demonstrate safe behaviors. The primary goal and first metric for HAV testing should be safety. “Roadmanship” metrics need time to be better defined and can be added in future tests.

**Design features: Test-track based, comprehensive scenarios, random but fair, difficult miles**

The Mcity ABC Test is test-track based and doesn’t involve numerical simulations. Developers use different simulation software packages that often aren’t interchangeable
or compatible, which makes it difficult for a third-party validation process to incorporate. In addition, the software used in these systems are important intellectual properties that are carefully guarded by companies.

Testing for highly automated vehicles also needs to be comprehensive. In the past, Advanced Driver Assistant Systems, including features such as Automated Emergency Braking, Adaptive Cruise Control and Lane Keeping Systems, have been evaluated using processes from several organizations, such as the Insurance Institute for Highway Safety (IIHS) [9], European New Car Assessment Programme (Euro NCAP) [10]-[12] and the Japanese New Car Assessment Program (JNCAP) [13]. The tests were selected based on the design domains of these automated vehicle functions.

For example, automated braking commonly is tested under three scenarios: stationary obstacle (parked car); lead vehicle driving at a slower constant speed; and lead vehicle decelerating. These scenarios are an effort to assess the performance of the braking function by sampling a number of cases within the envisioned operational design domain. That won’t work with highly automated vehicles, where a human driver isn’t always present, or if present, may be distracted. To effectively judge these more advanced vehicles the test cases must include many more scenarios, such as lane changes, left turns and other commonly experienced daily driving situations that can result in major vehicle crashes.

Another example is the test approach taken to evaluate driver assistance systems. Not only are the test scenarios pre-defined, such as cut-in, cut out and front vehicle decelerating, but the exact parameters, such as vehicle speed and deceleration rate, also are pre-defined. These kinds of well-defined deterministic tests come with a major drawback: vehicle developers can focus on tuning their systems to pass the tests but the results can indicate very little about the vehicle’s safety performance under different parameters.

This is not a problem if structural damage and passenger injuries continue to increase with the relative speed of the crash. But the complex nature of highly automated vehicle control systems can cause HAVs to respond in many ways to the combined effect of sensor fusion, perception algorithm, computation load, and other factors. It’s not clear which scenarios are simple and unchanging and which ones are complex and variable. Because highly automated vehicles are designed to run with no human driver as a back-up, these vehicles simply must be tested with a more comprehensive approach. The exact test parameters shouldn’t be pre-defined, and instead must be randomly selected within defined limits.
When test cases are randomly selected, it’s important to ensure that the overall test for different vehicles have the same controlled levels of difficulty. Imagine that all HAVs receive 30 test cases for a particular test scenario — vehicle cut-ins. The HAV is driving at a required speed, when a vehicle driving at a slower speed cuts in from an adjacent lane, at controlled parameters — relative speed, time-to-collision, etc. The difficulty level of cases depend on the parameters and can be categorized as 5-point questions, 4-point questions and so on. The final score of an HAV under testing can be computed from the number of test cases passed. It is important that all HAVs under test receive the same number of 5-point questions, 4-point questions, etc., to ensure the tests involve random situations and factors while remaining valid and fair.

The final factor needed in a comprehensive safety test for highly automated vehicles would be “difficult miles.” Many vehicle developers track the number of miles accumulated by a vehicle on public roads as a primary metric. In addition, the frequency of human operator intervention is used as a primary metric for safety. At Mcity, we think both of these metrics are fine but that they can both easily be manipulated.

The 2018 self-driving safety report filed by GM [14] noted that the number of scenarios, such as left-turns, lanes blocked by construction and more, that occurred in the San Francisco area vs. occurrences in the Phoenix area can differ by a factor of 40, on a per-mile basis. Test vehicles deployed in cities with easy miles may accumulate miles faster and demonstrate a lower number of human interventions, but the amount of learning produced would be less significant. Safety assessments on a test-track shouldn’t waste time on easy miles, because those tests need to be concluded quickly, in weeks, not years, and should focus on difficult miles.

**The Three pillars of the Mcity ABC Test**

The proposed Mcity ABC test consists of three parts: Accelerated evaluation, Behavior competence and Corner cases. Taken together, they are random, valid, fair and comprehensive.

**Accelerated Evaluation**

In daily driving, there are three common scenarios: car following, lane change and left turns. Seven of the top 10 crash situations identified by the National Highway Transportation Safety Administration (NHTSA) involve these three scenarios so they need to be highlighted in testing (see Figure 1.) The accelerated evaluation process being developed by Mcity is based on the importance sampling concept, a methodology used in financial institute stress tests, communication systems failure and other applications.
with rare failures. This methodology was first applied to ADAS testing at the University of Michigan [15][16].

In this process, the first step is to collect naturalistic driving data that reflects what the test vehicles will face on public roads in normal conditions. The behavior of the human drivers are then “skewed” (based on importance sampling) to boost aggressive/risky behaviors and focus on difficult miles. This process is possible because the U-M and, in particular, the U-M Transportation Research Institute, has 20 years of experience leading field operational tests and collected tens of millions of miles of naturalistic driving data.

Figure 1. Three common driving scenarios that correspond to seven of the top ten vehicle-to-vehicle priority light-vehicle pre-crash scenarios [19]

**Behavior competence**

In the behavior competence testing, vehicles are put through a set of comprehensive scenarios to demonstrate their safety performance. Several sets of scenarios have been developed [20], [21], including a list from Waymo. [22]. Mcity researchers from both the U-M and the Mcity Leadership Circle companies reviewed these existing sets of scenarios, together with literature from Torc robotics [23], Voyage [24], China Ministry of Industry and Information Technology [25], Beijing [26], Shanghai [27], Guangzhou [28], and compiled a list of 50 scenarios.

Another factor to be considered is weather and lighting conditions under which the tests are conducted. If the design domain of a specific vehicle calls for safe operations under low-light conditions and on rainy days, then the tests should be conducted under those conditions. While testing under low-light conditions requires simply waiting for the sun to set, none of the test facilities now available can provide controlled weather and lighting
tests for a large number of driving scenarios. Should a test track be built to have controlled all-weather testing capability? While ideally this should be the case, it is cost prohibitive. Our suggestion is to test these scenarios under day-time and night-time, in combination with testing the perception system, such as cameras, radars and lidars in rain and snow. Based on that list of 50 scenarios, a shorter list of 35 scenarios was selected by the Mcity members that correspond to major motor vehicle crashes and meet the physical limitations of the Mcity Test Facility, as listed in Appendix B, with possible locations to conduct them at the Mcity Test Facility.

Finally, we also reviewed the 35 scenarios to find those that are appropriate for low-speed shuttles that navigate by largely following a GPS-defined path, with a focus on avoiding crashes in the small and confined track defined path area. We found that only 16 of the 35 scenarios were needed for low-speed shuttles like the NAVYA-built shuttles Mcity is operating on U-M’s North Campus. These 16 scenarios are shown in Appendix C.

**Corner Cases**

There is no official definition of corner cases, other than the fact most experts agree these cases should be deterministic and should be at the very edges of the operating domain design. A corner case, for example, would involve an automated vehicle traveling at its highest speed with the most obstructed view, and facing a pedestrian running at the highest speed and so on. Corner cases also can be designed to explore the known weakness of the sub-systems of the vehicle, such as asking a vehicle relying on cameras to recognize a black car on a dark night.

Few companies will share what they know about the most challenging cases for their technologies, however. The only list we know of is the Waymo Safety Report, which includes the following corner case scenarios: Flashing yellow arrow signals; wrong-way drivers; nimble cyclists’ left turn in busy traffic; pedestrians and motorcycles “splitting the lane;” and joggers zig-zagging across the street. Some of these scenarios are beyond the capabilities of current testing systems. If they were to be conducted in a test track, then human test objects would need to be used, for example, which poses a major challenge in running such tests.
CONCLUSION

What’s next?
Mcity researchers are working on executing the Mcity ABC Test, starting with the 16 behavior competence scenarios for low-speed, path-following Level-4 vehicles, and the three accelerated evaluation behaviors. At the same time, Mcity will pursue funding support needed to prove the Mcity ABC Test concept.

Researchers are developing the ABC test concept using the Mcity Test Facility, U-M’s controlled, real-world environment for testing connected and automated vehicles. The ABC test could be scaled for larger test facilities as well, such as the American Center for Mobility in Ypsilanti Township, Michigan. U-M is a founding partner of ACM.

About Mcity
Mcity at the University of Michigan is leading the transition to connected and automated vehicles. Home to world-renowned researchers, a one-of-a-kind test facility, and on-road deployments, Mcity brings together industry, government, and academia to improve transportation safety, sustainability, and accessibility for the benefit of society.
REFERENCES

15. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/120657/zhaoding_1.pdf
18. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/84378/102747.pdf
24. https://oas.voyage.auto/scenarios/
27. www.sheitc.gov.cn/cyfz/676771.htm
APPENDIX A: Complete list of 50 behavior competence scenarios

Ego-car test (25)

- Move out of the travel lane and park
- Perform lane changes
- Follow lane lines
- Detect and respond to faded or missing roadway markings or signage
- Drive on roads without lane lines
- Adjust position for cones
- Launching from a steep uphill
- Perform a U-turn
- Make appropriate reversing maneuvers
- Road with up/down elevation
- Lane change through a narrow gateway
- Drive over bridge
- Drive through tunnel
- Detect and respond to speed limit changes and speed advisories
- Detect and respond to traffic signals and stop/yield signs
- Detect and respond to access restrictions (one-way, no turn, ramps, etc.)
- Detect and respond to temporary traffic control devices for detours/change of pattern
- Follow local and state driving laws
- Navigate around unexpected road closures
- Navigate railroad crossings
- Detect and respond to vehicle control loss
- Detect and respond to unanticipated weather or lighting conditions outside of vehicle’s capability
- Detect and respond to unanticipated lighting conditions
- Navigate cul-de-sacs
- Navigate narrow streets

Interact with a vehicle (18)

- Perform high-speed and low-speed merge
- Detect passing and no passing zones and perform passing maneuvers
- Adjust position for vehicles encroaching in lane
- Detect and respond to encroaching oncoming vehicles
- Perform car following (including stop and go)
- Detect and respond to stopped vehicles and stationary obstacles
- Detect and respond to lane changes (cut-ins)
- Navigate intersections and perform (left and right) turns
- Navigate roundabouts
- Moving to a minimum risk condition when exiting the travel lane is not possible
- Navigate a parking lot respond to reversing vehicles and locate spaces
- Detect and respond to emergency vehicles, including at intersections
• Detect and respond to school buses
• Detect and respond to non-collision safety situations (e.g. Vehicle doors ajar)
• Respond to unexpected maneuvers by other vehicles
• Respond to vehicle breaking rules at traffic lights
• Navigate environments with occluded view
• Detect and respond to golf carts

Interact with VRU and other non-vehicle player (7)
• Detect and respond to work zones and people directing traffic
• Make appropriate right-of-way decisions, including at crosswalks
• Follow police/first responder/worker/citizen controlling traffic
• Detect and respond to pedestrians in road (not at intersection or crosswalk)
• Keep safe distance from vehicles, pedestrians, bicyclists on side of the road
• Detect and respond to animals
• Detect and respond to motorcyclists
APPENDIX B: 35 behavior competence scenarios—short term focus of Mcity ABC Test

1. Move out of the travel lane and park
2. Perform lane changes
3. Follow lane lines
4. Detect and respond to faded or missing roadway markings or signage
5. Drive on roads without lane lines
6. Adjust position for cones
7. Perform a U-turn
8. Lane change through a narrow gateway
9. Drive through tunnel
10. Detect and respond to speed limit changes and speed advisories
11. Detect and respond to traffic signals and stop/yield signs
12. Detect and respond to temporary traffic control devices for detours/change of pattern
13. Navigate around unexpected road closures
14. Navigate railroad crossings
15. Navigate narrow streets
16. Perform high-speed and low-speed merge
17. Detect passing and no passing zones and perform passing maneuvers
18. Adjust position for vehicles encroaching in lane
19. Detect and respond to encroaching oncoming vehicles
20. Perform car following (including stop and go)
21. Detect and respond to stopped vehicles and stationary obstacles
22. Detect and respond to lane changes (cut-ins)
23. Navigate intersections and perform (left and right) turns
24. Navigate roundabouts
25. Navigate a parking lot respond to reversing vehicles and locate spaces
26. Detect and respond to emergency vehicles, including at intersections
27. Detect and respond to school buses
28. Detect and respond to non-collision safety situations (e.g. Vehicle doors ajar)
29. Respond to vehicle breaking rules at traffic lights
30. Navigate environments with occluded view
31. Detect and respond to golf carts
32. Detect and respond to work zones and people directing traffic
33. Make appropriate right-of-way decisions at crosswalks (pedestrian + bicycle)
34. Detect and respond to pedestrians in road (not at intersection or crosswalk)
35. Keep safe distance from pedestrians and bicyclists on side of the road
APPENDIX C: 16 behavior competence scenarios—suitable for low-speed path-following shuttles

1. Perform low-speed merge
2. Adjust position for vehicles encroaching in lane
3. Detect and respond to encroaching oncoming vehicles
4. Perform car following (including stop and go)
5. Detect and respond to stopped vehicles and stationary obstacles
6. Detect and respond to lane changes (cut-ins)
7. Navigate intersections and perform (left and right) turns
8. Navigate roundabouts
9. Navigate a parking lot respond to reversing vehicles and locate spaces
10. Detect and respond to non-collision safety situations (e.g. Vehicle doors ajar)
11. Respond to vehicles breaking rules at traffic lights
12. Navigate environments with occluded view
13. Detect and respond to golf carts
14. Make appropriate right-of-way decisions at crosswalks (pedestrian + bicycle)
15. Detect and respond to pedestrians in road (not at intersection or crosswalk)
16. Keep safe distance from pedestrians and bicyclists on side of the road