Edge Case Research, LLC
The Ice House Building
100 43rd Street, Suite 208
Pittsburgh, PA 15201

City Infrastructure for Safe Autonomy

A WHITE PAPER TO THE CITY OF PITTSBURGH DEPARTMENT OF INNOVATION & PERFORMANCE

RFI Title: "Request for Information for Smart Streetlights"

Firm Contact:
Michael Wagner
Edge Case Research, LLC
The Ice House Building
100 43rd Street, Suite 208
Pittsburgh, PA 15201
Phone: 412-606-3842 (note, no fax)
Email: mwagner@ecr.guru

We acknowledge that responses to this RFI may be considered public information in accordance with the Commonwealth of Pennsylvania Right to Know Laws as described in Section 5 of the RFI. Edge Case Research met with Councilman Gilman on March 23, 2017 to discuss generally issues of autonomous-vehicle safety.

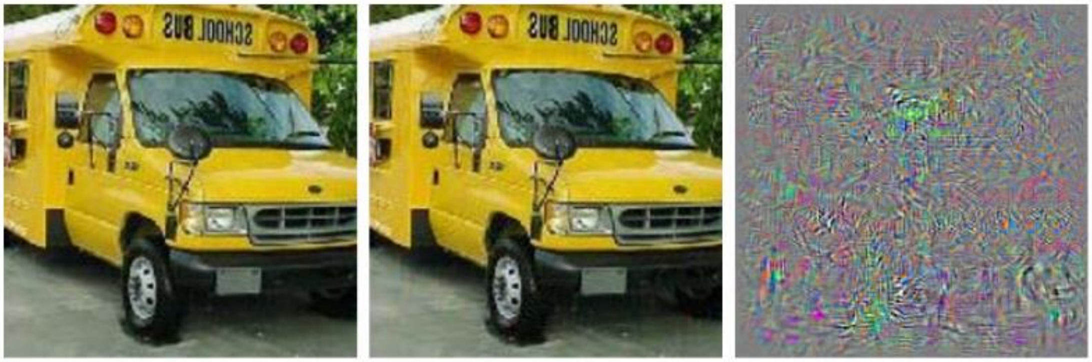
April 10, 2017

Problem Description

AUTONOMOUS VEHICLES hold the promise of dramatic reduction in driving fatalities, traffic congestion, and the stress of daily commute driving. Many city governments believe this will happen within ten years. But, how to get fully autonomous vehicles to actually be safe is no simple matter. We have outlined a number of areas that present significant challenges to creating and deploying a full-size fleet of acceptably safe, fully autonomous vehicles (Koopman and Wagner, 2017).

The question is not whether autonomous vehicles will be perfect. They won't be. The question is when we will be able to deploy a fleet of fully autonomous driving systems that are actually safe enough to leave humans completely out of the driving loop. The challenges are significant, and span a range of technical and social issues for both acceptance and deployment (Rupp and King, 2010), (Bengler et al., 2014), (Learner, 2015). A holistic solution will be needed, and out of necessity must include a broad appreciation for the range of challenges, and potential solutions, by all the relevant stakeholders and disciplines involved. Achieving a fleet of safe autonomous vehicles is not something that can be solved with a single technological silver bullet. Rather, it is a coupled set of problems that must be solved in a coordinated, cross-domain manner.

One prominent problem is that there are currently no safety standards for how to build or test autonomous vehicles. A big reason why existing standards don't apply is the use of machine learning for important features such as pedestrian detection. Using massive numbers of training images, machine learning builds an understanding independently, on its own, of how to spot pedestrians in the world. Unfortunately such processes are inscrutable; humans cannot intuitively understand what a machine-learning system has actually learned. This yields the potential for unexpected risks. Consider the figure below. The image on the left is of a school bus. A machine-learning system called a neural network has been trained to correctly detect the school bus. To the human eye, the image in the middle is the same. However, an imperceptible amount of noise (to a human) has been added to it. Although we intuitively know this small level of noise should not affect the classification, in this case it causes the neural network to miss the school bus entirely.



Left: An image that a neural network correctly classifies as a school bus. Middle: A very similar image that the neural network does *not* classify as a school bus. Right: A magnified view of the difference between the left and middle images. (Szegedy et al., 2013)

While there might be some special cases, in general the problem of "legibility" (Dosovitskiy and Brox, 2015) (Zeiler and Fergus, 2014) of machine learning in terms of being able to explain in human terms how the system behaves is unsolved. Although some might assume that a system's quoted accuracy statistics hold over any conceivable input, they do in fact only measure performance on the test data, and may be wildly different on different data sets that they encounter in the wild (e.g., Nguyen et al., 2015). Any claims of safety have to argue that the system has been trained on *all potential safety-relevant situations*, which might be an impossible claim to justify.

Our Objective

TRULY DRIVERLESS CARS are extremely rare today – most "autonomous" vehicles have a safety driver ready to take over should any unforeseen situation present itself, or a software design defect rear its ugly head. Although roboticists can show that autonomy *usually* works very well, work remains to show that autonomy *always* works well. We believe that there is incompatibility between traditional safety engineering and autonomy technologies such as machine learning could hamper the adoption of autonomous vehicles. And this is not unjustified – the software running hundreds or thousands of vehicles throughout our city should undergo close scrutiny and should prove itself trustworthy.

As we have discussed previously (Koopman and Wagner, 2016) one way to manage safety challenges of autonomous vehicles is to constrain operational concepts and engage in a *phased deployment*. In other words, preliminary deployment of a novel autonomous vehicle could include a combination of:

- safety operators who can shut down the vehicle in unexpected situations
- operating only in controlled environments that are favorable to the technology (e.g., closed roads and good weather)
- relying on detailed prior maps

The original motivation for phased deployment was to establish a bootstrapping strategy for deploying successively more sophisticated technology in a progressively more complex operational context, e.g., (Bayouth and Koopman, 1998), (Schladover et al., 2001). However, since our research was published, Edge Case Research has helped a number of clients design phased-deployment strategies, and this experience shown the potential value of considering the environment as *part* of the autonomous system rather than separate from it. We have come to appreciate the benefits that can be gained by designing the environment in which an autonomous system operates. Taken to its natural conclusion, we suggest that the built environment in cities represents an opportunity to mitigate the uncertainties of autonomy technology.

Hypothetical Examples

TO ILLUSTRATE this point, consider how the availability of reliable infrastructure technologies could improve can safety of autonomous vehicles operating in the hypothetical examples below.

Left: Police officers directing traffic in busy streets in Boston (Image: J. Leonard) and Pittsburgh.

Events or disruptions can necessitate unexpected manual interventions in traffic flow. Police officers, construction workers, and other personnel currently count on driver attention, signage, and high-visibility clothing to safely handle such situations. It is not clear how developers of autonomous vehicles would *prove* that their technology would unerringly handle such open-ended, ill-defined scenarios. Instead consider a case where city workers instead have beacons that they can use to reliably communicate status to all autonomous vehicles in the area. This would provide a redundant means of informing the vehicles about the situation, and possibly even provide public safety personnel with the ability to direct autonomous vehicles' behavior as needed.

Left: a city pedestrian crossing. Middle: unusual crosswalk configurations. Left: an impromptu crosswalk created by a school bus.

Even fixed traffic infrastructure can be difficult to manage reliably. Pedestrian crossings are challenging for a number of reasons. Pedestrians waiting to cross on the sidewalk may not be detected as an obstacle to avoid, despite the fact that legally the autonomous vehicle needs to stop to allow the pedestrian to cross. This level of scene understanding is difficult to verify to rigorous safety standards, therefore it would be useful for infrastructure to step in and assist. Potential examples include devices that pedestrians can use to reliably inform autonomous vehicles of their intention to cross.

This idea could extend to even trickier scenarios such as school bus stop signs. Return to the previous example of machine learning failing to detect a school bus. In this cases, redundant, high-integrity infrastructure could be integrated with the bus's sign-deployment mechanisms to reliably notify autonomous vehicles of the situations.

Even high visibility, standard signage can be easily obscured. Even if they are seen clearly, their purpose can be misunderstood.

We can extend the above idea even further to signage throughout the city. Such signage is not always captured by mapping activities, and the software responsible for recognizing such signs is complex and difficult to verify. Autonomous vehicles relying on signage or even maps could therefore be prone to error, which could result in accidents. Therefore, we can consider improvements to sign design, for example, the use of RFID elements integrated within the sign that define the behavior that the city requires of the autonomous vehicle. Programming instructions for the vehicle into an installed road sign could disambiguate the confusion that we all see in inexperienced Pittsburgh drivers. Although just an illustrative example, such improvements could make it easier for autonomous vehicles to detect the signage and properly understand their intent.

Verifying the Safety of Infrastructure

MANY EFFORTS are currently underway to develop infrastructure similar to that discussed above, including efforts in Pittsburgh. Such efforts are meant to address a number of concerns including improved traffic flow, lowering operating costs, and collecting valuable environmental data. However these technologies must be matured to be counted upon to deploy autonomous vehicles safety within the city. They must exhibit characteristics such as:

- *fault tolerance* to continue to provide safety features in the face of hardware and software failures that are bound to crop up in operation
- *robustness* to unexpected situations, including cybersecurity threats
- the ability to build and *verify* a "safety case" that argues the suitability of the technology

The notion of a “safety case” is a central concept, and it strong logical argument for why the city can trust the infrastructure to keep autonomous vehicles safe, and Goal Structuring Notation is an established standard for structured safety cases, being used in a number of domains for nearly twenty years (Kelly and Weaver, 2004). (Knight et al., 2015) describes a tailored approach to safety-case creation, regulatory evaluation, monitoring, upgrade and maintenance called the Comprehensive Lifecycle for Assuring System Safety (CLASS). CLASS is a system lifecycle that: (a) treats a system and its safety case as a composite, and (b) supports safety analysis, system development and certification across the entire system lifecycle from concept to decommissioning. The composite must be synchronized, i.e., the safety case must reflect the system accurately and vice versa. CLASS includes a spectrum of technologies that cover all aspects of the safety-case lifecycle and includes mechanisms for monitoring operational systems to check that operational evidence is consistent with assumptions made in the safety argument. Finally, CLASS includes a process for safety-case modification as needed by system changes during the system's operational lifetime.

Our Proposal: A Safety Case for Pittsburgh

WE PROPOSE to collaborate with the City of Pittsburgh to pull together technologies developed within its borders such as 3-D mapping, traffic control, data mining, simulation, and self-driving cars, and collaborate with community stakeholders, to create a “safety case for Pittsburgh” – an evidence-based argument for why autonomous vehicles are suitably safe to operate in the city without putting pedestrians and other drivers at risk. Edge Case Research does not produce 3-D maps, RFID devices, or beacons, but we do understand how these technologies can help avoid the mistakes that will inevitably occur in the incredibly complex software driving autonomous vehicles.

The Edge Case Research team has a decade of experience leading functional safety programs for autonomous systems for industry and the Department of Defense. Members of our team also have decades of experience *building* autonomous vehicles, so we are intimately familiar with the challenges they pose. This gives us the necessary expertise to helping self-driving car companies interested in operating in Pittsburgh to integrate with the “safety case for Pittsburgh”.

This safety case could initially serve as a roadmap for the City to guide investment into mitigating the most severe risks – for example, behavior in school zones or residential areas. We could then work with stakeholders to design, demonstrate, and verify infrastructure to begin pushing down risks.

Benefits to Pittsburgh

THE CITY OF PITTSBURGH could benefit by championing this kind of approach. By mitigating risks present in autonomous vehicles, the City could establish itself as a cost-effective place to deploy *safe* autonomous vehicles, thus securing Pittsburgh’s role in the self-driving future. Providing a safety infrastructure would not only solidify Pittsburgh as a place where well-established companies such as Uber and Argo operate, but would also provide a strong incentive for self-driving start-ups as well – Pittsburgh's safety infrastructure could make it less challenging for any player to set up operations in Pittsburgh.

By maintaining a "safety case for Pittsburgh", the city takes a proactive stance toward autonomous-vehicle safety that includes citizen oversight. This could help mitigate concerns some have about the impact that autonomous vehicles will have on safety in our city. It also mitigates defects that might, for various reasons, be present in autonomous vehicles sold. We believe this is an important point, because the latest NHTSA guidelines for the industry allow for manufacturers to self-certify the safety of their vehicles, a practice not permitted in safety-critical industries such as aerospace, rail, medical, chemical processes, or nuclear.

References

(Bayouth and Koopman, 1998) Bayouth, M. & Koopman, P., "Functional Evolution of an Automated Highway System for Incremental Deployment," Transportation Research Record, #1651, Paper #981060, pp. 80-88.

(Bengler, 2014) Bengler, K. et al., "Three decades of driver assistance systems," IEEE Intelligent Transportation Systems Magazine, Winter 2014, pp. 6-22.

(Dosovitskiy and Brox, 2015) Dosovitskiy A., Brox, T., "Inverting convolutional networks with convolutional networks," CoRR, vol. abs/1506.02753, 2015.

(Kelly and Weaver, 2004) T. Kelly and R. Weaver, "The Goal Structuring Notation – A Safety Argument Notation", Intl. Conf. on Dependable Systems and Networks.

(Koopman and Wagner 2017) Koopman, P. & Wagner, M., "Autonomous Vehicle Safety: An Interdisciplinary Challenge," IEEE Intelligent Transportation Systems Magazine, Special Issue on SSIV, 2017, in press.

(Koopman and Wagner, 2016) Koopman, P. and Wagner, M., "Challenges in Autonomous Vehicle Testing and Validation," SAE Int. J. Trans. Safety 4(1):2016, doi:10.4271/2016-01-0128.

(Knight et al., 2015) J. Knight et al., "A Safety Condition Monitoring System", Computer Safety, Reliability, and Security: SAFECOMP 2015

(Learner, 2015) Learner, P., "The hurdles facing autonomous vehicles," Automobile, Jun. 22, 2015.

(Nguyen et al., 2015) A. Nguyen et al., "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images", CVPR 2015.

(Rupp, 2010) Rupp, D. & King, A., "Autonomous Driving – A Practical Roadmap," SAE 2010-01-2335.

(Schladover et al., 2001) Shladover, S. et al., Development and performance evaluation of AVCSS deployment sequences to advance from today's driving environment to full automation, California PATH Research Report UCB-ITS-PRR_2001-18, August 2001.

(Szegedy et al., 2013) C. Szegedy et al. "Intriguing Properties of Neural Networks," arXiv preprint arXiv:1312.6199 (2013).

(Zeiler and Fergus, 2014) Zeiler, M. D., Fergus, R., "Visualizing and understanding convolutional networks." In ECCV, 2014.

About Us

EDGE CASE RESEARCH was formed in 2014 by Carnegie Mellon researchers to make complex software more robust. Our consulting services include embedded software testing and training services as well as autonomous vehicle and robotic functional safety. Our clients span a number of markets including aerospace, defense, robotics, consumer electronics, and industrial power systems. We have a team of over ten people with deep experience in robotics, embedded systems, and software safety.

We have executed multiple functional safety deployments of autonomous vehicles and robotics, and can help you do so as well. Our team has a deep background in this area, with multiple members of our team having over a decade of experience developing and testing autonomous robots and vehicles.

Edge Case Research is located in the Ice House Building in the Lawrenceville.

Our growing team is dedicated to making complex software more robust.

Michael Wagner

Michael Wagner is a co-founder of Edge Case Research and serves as its CEO. Mr. Wagner has nearly twenty years of experience developing advanced robotic systems for mining, agriculture, manufacturing, the Department of Defense, and NASA. Since 2006 his work has focused on building safer robots and researching ways for evaluating whether we are justified in trusting autonomous technologies. He was the project manager of the Automated Stress Testing for Autonomy Architectures (ASTAA) project, funded by the Test Resources Management Center's Unmanned Autonomous Systems Test group. ASTAA developed innovative stress-testing tools for unmanned-vehicle software that expose failure modes that are generally not uncovered with traditional testing. Prior to that, Mr. Wagner served as system-safety lead for the Autonomous Platform Demonstrator (APD) project, which built an advanced nine-ton, six-wheeled unmanned ground vehicle for the U.S. Army. He designed a "safety monitor" that acts as a safeguarding agent for the entire APD vehicle. He developed it with a high level of rigor, and it served as a basis for APD's safety-release for soldier experiments granted by U.S. Army Developmental Test Command. He also has experience building robots that operate in the most extreme environments on Earth. From 1999 to 2006, he led software implementation for five field expeditions at the Field Robotics Center at Carnegie Mellon. Mr. Wagner holds an

Michael Wagner
CEO and Co-Founder
Edge Case Research, LLC
mwagner@ecr.guru

M.S. in Electrical and Computer Engineering from Carnegie Mellon University.

Education:

M.S. Electrical and Computer Engineering, Carnegie Mellon University, 2002

B.S. Electrical and Computer Engineering with Physics Minor, Carnegie Mellon University, 1998

Professional Experience:

Founder, Edge Case Research, LLC (2013 – Present)

Senior / Commercialization Specialist, National Robotics Engineering Center, Carnegie Mellon (2006 – Present)

Senior / Research Programmer, Robotics Institute, Carnegie Mellon University (1999 – 2006)

Philip Koopman

Philip Koopman
Chief Scientist and Co-founder
Edge Case Research, LLC
pkoopman@ecr.guru

Professor Koopman co-founded Edge Case Research, and is a faculty member at Carnegie Mellon University with appointments in the Department of Electrical & Computer Engineering, the Institute for Software Research, and the Robotics Institute. His background includes time as a submarine officer for the US Navy, a principal in a several startups, an embedded CPU architect for Harris Semiconductor, and an embedded system architect for United Technologies Research Center. At Carnegie Mellon he was the principal investigator of the Automated Stress Testing for Autonomy Architectures (ASTAA) project, and currently serves as PI of the Robustness Inside-Out Testing (RIOT) project, both funded by the DoD's Test Resource Management Center's Unmanned and Autonomous Systems Test (UAST) group. Dr. Koopman has worked in the broad areas of wearable computers, software robustness, embedded networking, dependable embedded computer systems, and autonomous vehicle safety. Dr. Koopman was the leader of the Ballista project at Carnegie Mellon, and has 20 years of experience with applying robustness testing to real-world systems. He has learned what it takes to get embedded software right over the course of more than 150 industry design reviews, and currently teaches embedded systems to both undergraduate and graduate students. Additionally, he serves as the testifying expert on software safety in the ongoing Toyota Unintended Acceleration cases. He is the author of the book Better Embedded System Software, which distills this experience into a set of lessons learned that broadly apply across the entire embedded software industry. He holds 27 issued US patents and has well over 100 publications.

Education:

B.S. (Magna cum Laude) in Computer and Systems Engineering, Rensselaer Polytechnic Institute, 1982.

M.Eng. Computer Engineering, Rensselaer Polytechnic Institute, 1982.

Ph.D., Dept. of Elec. and Computer Engr., Carnegie Mellon University, 1989.

Professional Experience:

2013-present: Founder, Edge Case Research, LLC

1997-present: Assistant Professor, Associate Professor (2001), Electrical and Computer Engineering Dept., Carnegie Mellon University. Research and education in safety critical and secure embedded systems.

1996-1997: Visiting Senior Research Engineer, Institute for Complex Engineered Systems, Carnegie Mellon University. Launched the Ballista software robustness testing project.

1991-1995: Principal Research Engineer. United Technologies Research Ctr., Research on distributed embedded systems including aviation applications.

1989-1990: Senior Scientist. Harris Semiconductor, Embedded CPU architect.

1985-1987: Engineering Duty Officer. US Navy, Trident Command and Control Systems Maintenance Activity (TRICCSMA), Newport, RI.

1983-1985: Submarine Officer. US Navy, USS Haddock (SSN-621), U.S. Pacific Fleet. Active duty deployment.