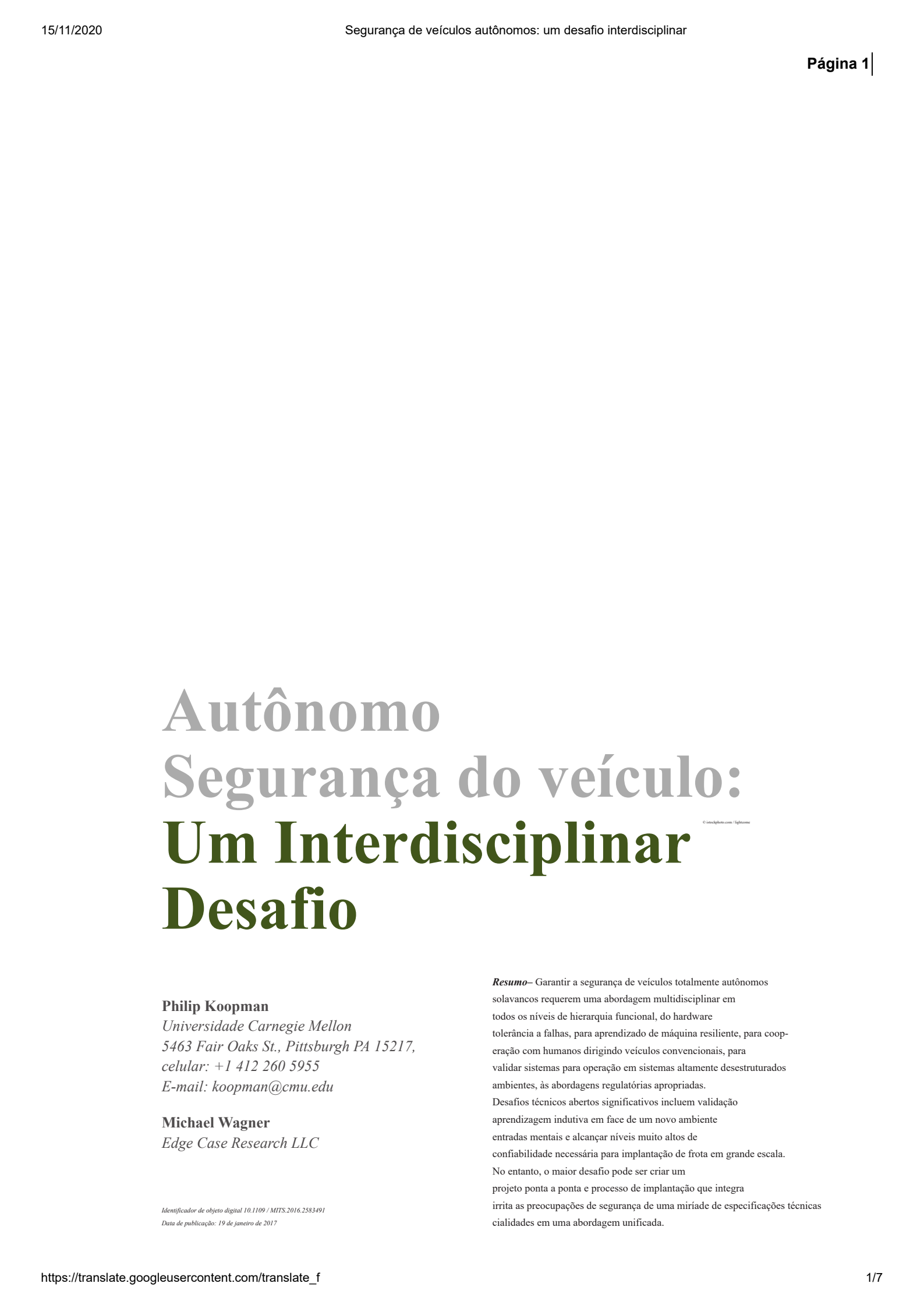
Autônomo Segurança do veículo: Um Interdisciplinar Desafio



© istockphoto.com / lightcome

Philip Koopman
Universidade Carnegie Mellon
5463 Fair Oaks St., Pittsburgh PA 15217,
celular: +1 412 260 5955
E-mail: koopman@cmu.edu

Michael Wagner
Edge Case Research LLC

Identificador de objeto digital 10.1109 / MITS.2016.2583491
Data de publicação: 19 de janeiro de 2017

***Resumo*–** Garantir a segurança de veículos totalmente autônomos solavancos requerem uma abordagem multidisciplinar em todos os níveis de hierarquia funcional, do hardware tolerância a falhas, para aprendizado de máquina resiliente, para cooperação com humanos dirigindo veículos convencionais, para validar sistemas para operação em sistemas altamente desestruturados ambientes, às abordagens regulatórias apropriadas. Desafios técnicos abertos significativos incluem validação aprendizagem indutiva em face de um novo ambiente entradas mentais e alcançar níveis muito altos de confiabilidade necessária para implantação de frota em grande escala. No entanto, o maior desafio pode ser criar um projeto ponta a ponta e processo de implantação que integra irrita as preocupações de segurança de uma miríade de especificações técnicas cialidades em uma abordagem unificada.

REVISTA IEEE INTELLIGENT TRANSPORTATION SYSTEMS • 90 • Spring 2017 1939-1390 / 17 © 2017IEEE

Uso licenciado autorizado limitado a: b-on: Instituto Politécnico da Guarda. Baixado em 01 de novembro de 2020 às 01:54:15 UTC do IEEE Xplore. Restrições aplicadas.

Página 2

Introdução

UMA pre-visão típica do futuro de veículos autônomos
cle, inclui pessoas sendo aliviadas do estresse de
dirigir diariamente, talvez até tirar uma soneca
no caminho do trabalho. Espera-se que seja acompanhado
negado por uma redução dramática nas mortes ao dirigir devido a
substituindo motoristas humanos imperfeitos por (presumivelmente) melhor
pilotos automáticos computadorizados. Governos municipais aparentemente
acredito que isso acontecerá dentro de 10 anos (Boston Consult-
ing Group 2015). Mas, como obter esses veículos totalmente autônomos
obstáculos para estar realmente seguro não é uma questão simples (Luettel 2012,
Gomes 2014). Delineamos uma série de áreas que apresentam
desafios significativos para a criação de aceitáveis seguras e totalmente automáticas
veículos tonômicos em comparação com os veículos de até mesmo alguns
anos atrás, com ênfase na dificuldade de validação
autonomia na escala de uma frota de veículos de grande porte.

A questão não é se os veículos autônomos irão
ser perfeito (eles não serão). A questão é quando podemos
para implantar uma frota de sistemas de direção totalmente autônomos que
são realmente seguros o suficiente para deixar os humanos completamente fora de
o ciclo de condução. Os desafios são significativos e abrangem
uma série de questões técnicas e sociais para aceitação
e implantação (Rupp 2010, Bengler 2014, Learner 2015).
Uma solução holística será necessária, e deve necessariamente
incluem uma ampla apreciação pela gama de desafios
(e soluções potenciais) por todas as partes interessadas relevantes
e disciplinas envolvidas.

Nosso trabalho na construção de argumentos de segurança e tempo de execução segu-
mecanismos para veículos terrestres autônomos nos ensinou
que até mesmo entender o que "seguro" realmente significa para autono-
veículos mous não é tão simples. “Seguro” significa pelo menos corretamente
implementar comportamentos no nível do veículo, como obedecer ao tráfego
leis fic (que podem variar dependendo da localização) e
enfrentando perigos não rotineiros da estrada, como energia reduzida
linhas e inundações. Mas também significa coisas como failover
planejamento de missão, encontrando uma maneira de validar com base indutiva
estratégias de aprendizagem, proporcionando resiliência em face de prováveis
lacunas nos requisitos de sistema implantados antecipadamente e ter um
estratégia de certificação de segurança apropriada para demonstrar que um
nível suficiente de segurança foi realmente alcançado.

Assim, alcançar um veículo autônomo seguro não é algo
coisa que pode ser resolvida com uma única prata tecnológica
bala. Em vez disso, é como um conjunto acoplado de problemas que devem ser
resolvido de maneira coordenada e entre domínios. Lá-
O mainder deste artigo descreve alguns problemas gerais ar-
eas e algumas das interações entre eles. Como todos
ganha mais experiência com a tecnologia, sem dúvida alguns
mais problemas de alto nível e muitos problemas mais detalhados
surgirá, mas este é um ponto de partida para a compreensão
A foto maior.

Engenharia segura

Vamos começar supondo que já temos em pequena escala
implantação de Nível 4 totalmente autônomo (NHTSA 2013)
veículos na estrada que geralmente são bem comportados. Portanto,
começamos com a expectativa de que a maioria dos veículos funcionará
bem na maioria das vezes no ambiente diário na estrada
condições. Agora queremos implantar em escala. O desafio
torna-se gerenciamento de falhas que são muito raros para
qualquer veículo, mas vai acontecer com muita frequência
ser aceitável à medida que a exposição aumenta a milhões de veículos
em uma frota.

Existe um corpo de conhecimento bem desenvolvido sobre como
para tornar os sistemas automotivos baseados em computador seguros, e um
história ainda mais longa de criação de computadores essenciais para a segurança
sistemas baseados em trens, componentes de processos químicos,
aeronaves, e assim por diante. Vamos primeiro explorar os desafios deste
ponto de vista da engenharia de segurança e, em seguida, revisitar as coisas
do ponto de vista de outras disciplinas.

Prática atual aceita para veículos baseados em computador
a segurança do sistema é normalmente baseada em um aplicativo de segurança funcional
proach (por exemplo, o padrão de segurança ISO 26262 específico para automóveis
dard). Uma questão importante é que a ISO 26262 geralmente fornece um sistema
crédito para um motorista humano, em última análise, sendo responsável por
segurança. Mas, com um veículo totalmente autônomo, o humano
não será responsável por dirigir. Contando com autono-
meu ser totalmente responsável pela segurança do veículo sem
a supervisão do motorista é uma grande mudança em comparação com a atual
implantou sistemas avançados de assistência ao motorista que em grande parte
confie no motorista como responsável pela segurança do veículo.
Uma abordagem para lidar com a falta de supervisão humana do motorista
é definir o aspecto de "controlabilidade" da ISO 26262 para zero para um
sistema autônomo, o que poderia aumentar dramaticamente
os requisitos de segurança de uma variedade de funções automotivas
ções em comparação com os veículos de hoje. Se isso vai funcionar,
ou mesmo se ISO 26262 pode ser usado efetivamente como está para validar
veículos autônomos é uma questão interessante, mas em aberto.
(Koopman 2016).

Uma preocupação significativa de certificação de segurança é validar
qualquer uso por veículos autônomos de sistema auto-adaptativo
comportamento. (de Lemos 2013) Adaptação irrestrita como
como aprendizagem em tempo real de novos comportamentos (Rupp 2010) significa
que um veículo pode ter um comportamento diferente
durante a operação do que foi exibido durante o teste e
certificação. As abordagens de certificação atuais são essenciais
parcialmente incapaz de lidar com essa situação, porque eles exigiam
considerando todos os comportamentos possíveis do sistema antecipadamente no
processo de projeto e validação. A menos que os limites sejam de alguma forma
colocar em adaptação e totalmente explorado durante o projeto do sistema,
pode ser impossível garantir a segurança de tal sistema
em tempo de design, porque o sistema que está sendo testado não terá
o mesmo comportamento de um sistema adaptado que é implantado.
Abordagens de métodos formais podem ser capazes de se provar adequadas
laços sobre sistemas adaptativos, mas tais provas vêm com
sumptions que não são necessariamente prováveis ou testáveis,
e tais abordagens atualmente não escalam bem para o tamanho real
sistemas de software. Observe que por "adaptativo" queremos dizer que o
sistema essencialmente muda seus comportamentos, dependendo de seu

REVISTA IEEE INTELLIGENT TRANSPORTATION SYSTEMS • 91 • primavera 2017

Uso licenciado autorizado limitado a: b-on: Instituto Politécnico da Guarda. Baixado em 01 de novembro de 2020 às 01:54:15 UTC do IEEE Xplore. Restrições aplicadas.

Página 3

FIg 1 Muitas áreas diferentes exigem uma coordenação interdisciplinar abordagem para garantir a segurança.

histórico operacional, por exemplo, usando máquina on-line Aprendendo. Esta é uma gama de comportamentos muito mais dinâmica do que visto em sistemas baseados em controles mais tradicionais, como como controle de cruzeiro adaptativo, que pode ser validado usando métodos mais tradicionais. (Kianfar 2013).

Uma abordagem comumente mencionada para hedge de nível de sistema A segurança para sistemas altamente autônomos é reatar o motorista quando houver uma falha de equipamento, proporcionando uma rede de segurança humana para automação. Por exemplo, um hu- o homem pode estar tirando uma soneca e vai precisar de tempo para ganhar consciência situacional suficiente para assumir a responsabilidade para dirigir. Para preencher a lacuna de desatenção humana, a ve- hicle precisará ter algum tipo de falha operacional autonomia capacidade de realizar até que um humano possa recuperar o controle.

Felizmente, os carros normalmente podem atingir um estado seguro em segundos (puxar para o lado da estrada), em comparação com horas para aviões (voar para um aeroporto de desvio). Assim, um eficaz estratégia de segurança pode ser para veículos mudarem operacional modos para uma "missão de segurança" de curta duração quando um crítico o componente primário cal falha. A estratégia aqui é que um sistema de autonomia que só precisa ser inteligente o suficiente para puxar um veículo para o lado da estrada ao longo de alguns segundos podem ser projetados de uma forma menos complexa do que um sistema de autonomia total de condução.

Por exemplo, um subsistema de segurança pode permanecer na corrente alugue a faixa ao parar e, assim, dispensar com sensores e sistemas de controle necessários para mudança de faixa manobras. Além disso, um curto tempo de missão (segundos, não horas) provavelmente facilitaria a confiabilidade e a redundância requisitos no próprio subsistema de proteção. Como um adicionado benefício, projetar uma capacidade de missão de desligamento seguro pode relaxar os requisitos de segurança no veículo primário o meu. Se uma missão de segurança está sempre disponível, au- tonomia não precisa ser falha operacional. Em vez disso, pode ser suficiente para garantir que uma missão de segurança seja invocada quando- sempre que houver uma falha do sistema de autonomia primário, permitindo um sistema de autonomia primária menos que perfeito contanto que as falhas sejam detectadas com rapidez suficiente para invocar uma missão de segurança. Relaxando os requisitos de segurança em pri- autonomia de Maria (mantendo o veículo como um todo seguro) poderia potencialmente oferecer uma redução dramática no sistema geral custo e complexidade. (Koopman 2016).

Robôs ultraconfiáveis

Fazendo sistemas autônomos (que são robôs) funcionarem em uma ampla variedade de situações de direção cotidianas, como sempre feito nos protótipos atuais é verdadeiramente significativo e im- conquista pressiva. No entanto, fazendo com que funcionem bem o suficiente para atingir os níveis de segurança exigidos para uma frota de veículos totalmente autônomos terão outro conjunto significativo de realizações. Por exemplo, considere um possível objetivo de tornando os carros totalmente autônomos seguros por hora de operação como aeronave. Isso exigiria um nível de segurança de cerca de 1 bilhão milhões de horas de operação por evento catastrófico. (FAA 1988) Vamos chamar essa meta de segurança de "ultra-confiabilidade".

Uma série de desafios para alcançar o ultra-confiável veículos autônomos surgirão, começando com a melhoria robustez do sistema para situações ambientais complicadas (por exemplo, lidar com detritos, desordem e ruído do sensor). No geral, parece implausível projetar um sistema que pode lidar com todas as situações ambientais possíveis perfeitamente, especialmente nos estágios iniciais de implantação de uma frota. Assim, parece desejável para garantir que os sistemas não sejam frágeis, e, em particular, tem alguma maneira de saber quando eles estão não está funcionando adequadamente. Em outras palavras, esses sistemas precisam ser capaz de monitorar sua confiança em sua própria propriedade operação e ser muito bom em saber quando eles não sabe o que está acontecendo. Atingindo detecção confiável de sistema sua degradação será difícil. Uma alta taxa de falsos negativos levará a veículos operando involuntariamente em um local inseguro maneira. Mas uma alta taxa de falsos positivos deixará muitos carros preso na beira da estrada devido a um falso alarme cibernético angústia (espero que depois de ter realizado um saf- missão em resposta à falha de autonomia).

Outro desafio significativo é que o aprendizado de máquina técnicas (Domingos 2012), como classificadores que são amplamente utilizado em veículos totalmente autônomos (por exemplo, Aeberhard 2015) tendem a ser baseados em abordagens de treinamento indutivo em vez de mais do que um projeto mais tradicional baseado em requisitos cess. Validar o raciocínio indutivo é conhecido há muito tempo ser inerentemente difícil (Hume 1748), e não há parecem ser uma forma de tornar os níveis de garantia ultra-confiáveis tees sobre como tal sistema se comportará quando encontrar dados que não estão no conjunto de treinamento nem no conjunto de dados de teste. Autônomo ve- solavancos operam com dados altamente dimensionais, com alta taxa fluxos de dados de vídeo, LIDAR e radar. É certo que

Uso licenciado autorizado limitado a: b-on: Instituto Politécnico da Guarda. Baixado em 01 de novembro de 2020 às 01:54:15 UTC do IEEE Xplore. Restrições aplicadas.

Página 4

eles serão expostos a dados do mundo real que diferem de alguma forma de dados de treinamento e validação. Resultados de aprendizado de máquina muitas vezes envolvem regras de decisão que geralmente são inescrutáveis para requisitos poderiam ser impostos como um conjunto de requisitos comportamentais monitorados no autônomo autonomia do veículo. (Kane 2015) Esse monitoramento pode ser

revisores humanos. (Dosovitskiy 2012) Assim, é difícil
razão sobre a correção do sistema de aprendizado de máquina
seu comportamento em face de novos dados.

Programas

Segurança de software é um tópico de pesquisa de longa data. (Leveson 1986) Abordagens atuais de segurança de software, como a ISO 26262 O processo "V" normalmente assume que alta qualidade re-peculiaridades são refinadas em uma implementação. Este ulti-matamente produz uma cadeia de evidências que acopla o teste final resultados de volta aos requisitos de sistema relevantes para a segurança.

Com sistemas adaptativos e de aprendizado de máquina, pode ser desafiador articular os requisitos do sistema de uma forma que suporta o V ou outro engenheiro de sistema tradicional processos de gestão. (Koopman 2016) Por exemplo, considere um classificador de pedestres que foi criado com base em um conjunto de dados de treinamento. Dizer que o sistema é seguro porque é ac-a precisão em um conjunto de validação é suficientemente alta implora a questão para saber se o sistema realmente funcionará como precisa quando confrontado com a confusão do mundo real.

O que realmente importa é que o aprendizado de máquina valida conjunto de informações deve ser abrangente o suficiente para garantir que não há lacunas no comportamento do sistema. Em termos de "V" modelo, o conjunto de treinamento é o mais próximo que temos do sistema de re-peculiaridades, e o conjunto de validação é o mais próximo que temos de um plano de teste. Mas, sabendo que o treinamento e validação conjuntos são bons o suficiente não é tão fácil. Como podemos ter certeza de que os casos extremos e as interações comportamentais sutis que são susceptíveis de afetar a segurança são realmente aprendidas pelo sistema?

Esperançosamente, o conjunto de treinamento e o conjunto de validação são ex-tremendamente abrangente, e cobrimos todos os possíveis cenários operacionais. Mas, como acontece com os sistemas mais tradicionais, ainda há a questão do cenário operacional inconcebível ios que ainda pode acontecer (aqueles famosos "desconhecidos conhecidos "). Além disso, existe a possibilidade de que alguns novos cenário operacional pareceria comum para uma pessoa (e, portanto, não incluído no conjunto de dados de teste), mas na verdade é ex-cepcional de alguma forma para o algoritmo de aprendizado de máquina, potencialmente causando comportamentos inesperados do sistema.

Baseando um argumento de segurança na suficiência do treinamento e os dados de validação também tornam potencialmente o sistema que coleta esses dados essenciais para a segurança. Afinal, segurança para tal sistemas, em última análise, dependem da precisão do treinamento e coleta de dados de validação. Isso pode, por exemplo, levar à necessidade de o sistema de coleta de dados ser desenvolvido operado de acordo com os padrões de software críticos de segurança, com atenção à redução de riscos, como viés não intencional ou distorção nos dados coletados.

Uma solução potencial para o problema de validação de ma-aprendizagem é definir separadamente e de forma independente o que significa operação "segura". Este conjunto separado de segurança usado durante a validação, testes on-road e talvez de-procedimento para garantir que o veículo não apresente comportamentos, mesmo se houver lacunas ou falhas na máquina sistemas de aprendizagem.

Hardware Informático

Mesmo com o uso de uma estratégia de missão segura, um óbvio desafio de hardware é criar hardware de custo ultrabaixo com comportamentos de falha seguros. Isso requer a criação de um componente arquitetura de hardware / software combinada que emprega redun-dançar corretamente. (Hammett 2001) Como um exemplo de progresso nesta área, os fabricantes de chips introduziram chips de computação com núcleos duplos que podem fornecer pelo menos redundância parcial cy para cálculos. No entanto, uma abordagem mais completa para garantir redundância suficiente e tolerância a falhas provavelmente será necessário. Há uma longa história de criação de tais sistemas para aeroespacial e outras aplicações críticas de segurança, fornecendo um rico conhecimento de base. (Randell 1978).

Um desafio de hardware mais sutil, mas crítico, é o questão da detecção de falhas latentes. Completamente livre de falhas redun-dança é normalmente assumida no início de cada missão ao realizar cálculos de confiabilidade. Qualquer não detectado a falha prejudica dramaticamente os benefícios da redundância. Mesmo um por cento ou dois das lacunas no autodiagnóstico têm dramáticas implicações para a confiabilidade alcançável. Por exemplo, Achiev-a cobertura de teste de apenas 95% pode reduzir a confiabilidade alcançável a redundância do sistema automotivo por ordens de magnitude. (Honeywell 1995) A razão para isso é que não diagnosticada falhas podem se acumular por toda a vida útil do veículo, portanto, a probabilidade de experimentar múltiplos independentes falhas não diagnosticáveis dentadas durante a vida de um veículo é bastante alto em comparação com a probabilidade de várias falhas ocorrendo durante uma única missão de condução para diagnóstico partes do sistema. Assim, será importante criar chips que pode ser testado antes de cada ciclo de direção com um nível extremamente alto de cobertura de diagnóstico.

Testando

Práticas tradicionais de segurança de veículos, pré-computador e os regulamentos enfatizaram os testes em nível de veículo. Mais recentemente, projetos de protótipos de veículos autônomos têm enfatizou a importância dos testes em estrada. (por exemplo, Urmson 2008, Levinson 2011, Broggi 2013, Ziegler 2014, Aeberhard 2015, SAE J3018) No entanto, é bem conhecido que uma abordagem apenas de teste é insuficiente para garantir o segurança de até mesmo não autônomo baseado em software crítico sistemas. (Butler 1993) Mais do que apenas testes são necessários para implantação em escala total de qualquer automotivo de segurança crítica Programas. No entanto, testes completos ainda são necessários.

Testes rigorosos de autonomia são executados em vários processa. O principal entre eles é aquele no desenvolvimento "V"

Uso licenciado autorizado limitado a: b-on: Instituto Politécnico da Guarda. Baixado em 01 de novembro de 2020 às 01:54:15 UTC do IEEE Xplore. Restrições aplicadas.

Página 5

modelo, o teste compara um documento de design rigorosamente definido contra um sistema para determinar se o sistema combina seu design e requisitos. Para sistemas probabilísticos como planejadores (Geraerts 2002), o comportamento dos sistemas espera-se que o tempo seja diferente em cada teste executado, mesmo para essen-condições iniciais aparentemente idênticas. Além disso, pequenas mudanças nas condições iniciais pode resultar em grandes mudanças no sistema comportamento. Portanto, o oráculo de teste (algo que prevê qual seria um resultado de teste correto) precisará a frota não experimentará uma falha catastrófica de software durante sua vida operacional (FAA 1988), então é difícil veja como isso pode ser feito por meio de testes sozinho.

Outra dimensão do teste é a injeção de falha e falha gestão de ure. Com a implantação de veículos em grande escala, vêm ocorrências diárias de falha do equipamento do veículo simplesmente devido ao grande número de veículos na frota. Se veículo controlabilidade é totalmente responsabilidade de uma autonomia sistema, será necessário caracterizar o que acontece

suporte a análise abstrata de resultados de teste em vez de mais
técnica tradicional de alimentar alguns valores específicos em
um pedaço de software e esperando um resultado específico e exato
a partir de um cálculo. (Feather 2001).

Os sistemas de aprendizagem indutiva são ainda mais desafiadores
para testar, porque não há design como tal e, portanto,
nenhum ponto de partida para construir um oráculo de teste. Em vez disso, como dis-
cussed, há um conjunto de dados de treinamento e um conjunto de validação
dados. No entanto, mesmo com conjuntos abrangentes de dados, é
não está claro como garantir que o sistema de aprendizado de máquina
treinou nas características essenciais do treinamento
dados em vez de correlações coincidentes.

As técnicas de aprendizado de máquina são bastante sofisticadas,
mas um argumento típico para correção acaba sendo estatística
tical por natureza. Embora alguns possam supor que um sistema
as estatísticas de precisão citadas prevalecem sobre qualquer entrada concebível,
na verdade, eles apenas medem o desempenho nos dados de teste,
e pode ser totalmente diferente em diferentes conjuntos de dados que eles
encontro na natureza (por exemplo, Nguyen 2015). Qualquer reclamação de
segurança tem que argumentar que não há nenhuma situação relevante para a segurança
ções ausentes nos conjuntos de dados de treinamento e teste. Isto
pode claramente funcionar bem para alcançar confiabilidade moderada via
aproximações de força bruta (por exemplo, precisão de 99,9%). no entanto
não está claro como garantir ultra-confiabilidade para a máquina
algoritmos de aprendizagem.

Considere um sistema de exemplo que executa 10.000 op-
rações por hora para controle do veículo (cerca de 3 por segundo),
e uma frota de um milhão de veículos. Testando para validar um
taxa de falha particular requer processamento de mais casos de teste
do que a taxa de falha desejada. Assim, garantindo menos de um
falha catastrófica por hora para esta frota exigiria
passando significativamente mais de 10 bilhões de *representantes*
casos de teste. Além disso, esta estimativa exige a satisfação
a suposição otimista de que as falhas são independentes
e que o conjunto de dados é realmente representativo de todos
coisa que uma frota de veículos encontrará. Também pode ser
uma meta de segurança modesta, porque se cada veículo da frota for
conduzido apenas uma hora por dia, que ainda permitiria uma diária
falha catastrófica do veículo.

Em outras palavras, os desenvolvedores teriam que testar mais
horas de exposição do que afirmam ser a taxa de falha da frota.
Coletar tantos dados de teste será claramente um fator significativo
desafio, pois seria validar essa quantidade de
dados como sendo realistas em todos os aspectos. (Kalra 2016) Se o
objetivo é um estilo de aviação de garantir que todo o seu
quando o sistema de autonomia tem que lidar com um veículo ex-
periencing um pneu estourado, falha do sensor, falha do atuador,
e até mesmo uma falha de algoritmo de autonomia em todo o
espectro de condições operacionais.

Segurança

A segurança da computação automotiva tem recebido
aumentou a atenção e não mostra sinais de se tornar um
problema. Claramente, os veículos autônomos terão que lidar
com segurança também. (SAE J3061).

Além de ataques a veículos específicos, medidas de segurança
certezas precisarão abranger ataques e falhas no nível do sistema
ures. Em particular, pode ser problemático confiar cegamente em
a segurança de outros veículos ou mesmo infraestruturas rodoviárias
tura ao realizar manobras autônomas otimizadas
como tráfego de interseção de fluxo livre. Por exemplo, en-
criptografar comunicações de veículo a veículo pode ajudar com
a segurança das mensagens de coordenação entre veículos. Mas
e se o veículo com o qual você está se comunicando com segurança
foi subvertido e está fornecendo maliciosamente incorreto
em formação? E se alguém invadiu fisicamente um
computador de infraestrutura de beira de estrada e reprogramá-lo, ou
mata a energia de um conjunto de sistemas de suporte de infraestrutura à beira da estrada
Tems em um ataque coordenado?

No mínimo, parece prudente garantir que cada
veículo autônomo tem a capacidade de perceber quando está
alimentando informações externas incorretas ou maliciosas, detectar
que um ataque está ocorrendo e realizar uma missão de segurança
se não puder continuar a operação total em face do ataque.

Interação Humano-Computador

À medida que os veículos autônomos suplantam os motoristas humanos,
capacidade de comunicação e cooperação com as pessoas
se tornará mais importante. Os riscos da supervisão humana
desatenção em sistemas com quase - mas não totalmente - cheio
a autonomia deve ser evidente. Mas mesmo totalmente autônomo
veículos precisarão pelo menos se certificar de que os ocupantes
sentir que o comportamento do veículo é seguro se eles quiserem construir
confiança do cliente, e precisará aprender como antecipar o
haviors de outros veículos também. (Gindele 2015) Enquanto alguns
pode dizer que a confiança do cliente não é, estritamente falando, uma segurança
questão, é uma questão vital para a adoção de tecnologia e, portanto,
segurança direta se os veículos autônomos cumprirem sua promessa de
salvar vidas na estrada. Outra interação humano-computador
questões também constituem desafios significativos.

Uso licenciado autorizado limitado a: b-on: Instituto Politécnico da Guarda. Baixado em 01 de novembro de 2020 às 01:54:15 UTC do IEEE Xplore. Restrições aplicadas.

Página 6

Veículos autônomos terão que interagir com o hu-
motoristas de outros veículos. Um carro que é muito educado ou muito
rude vai atrapalhar o fluxo do tráfego no mínimo, e talvez
indiretamente causar problemas de segurança mais significativos. Corte
motoristas humanos fora de cena podem levar muitos anos
enquanto a penetração no mercado de tecnologia totalmente autônoma
aumenta. Mesmo quando chega o dia em que todos os carros estão totalmente
autônoma, a estrada ainda será o lar de motoristas humanos
de bicicletas, patinetes, cavalos, equipamentos agrícolas e
em breve. Muitos desses usuários humanos da estrada ficarão relutantes
ou incapaz de seguir as regras e expectativas normais de trânsito
para veículos de passageiros. Mesmo se houver faixas exclusivas
inicialmente para facilitar a implantação (Shladover 2009), ao longo do tempo
parece provável que haverá uma tremenda pressão pública
para espalhar a autonomia para uma mistura autônoma / humana
cenários de veículos. Assim, parece provável que o tráfego misto
Dispositivos de gravação precisam ser repensados para fornecer
dados para este fim.

Uma questão importante de responsabilidade será determinar quem é o principal
responsável pela operação adequada do veículo. É o ocupante
que entrou em um veículo alugado, embora tivesse sensor
idade que deveria ser perceptível para um leigo? É o
fabricante de veículos que confiou em um trem de terceiros com defeito
conjunto de dados ing? É o mecânico que instalou por engano um
versão ligeiramente incompatível do software de substituição do sensor
porcelana? É o serviço de atualização de mapeamento que era muito lento para
registrar uma lavagem da ponte? É o fornecedor do sistema operacional
que não implantou um patch de segurança rápido o suficiente para prevenir
um acidente malicioso? Embora alguns desses problemas possam ser
puramente legal, a resolução de muitas questões jurídicas exigirá um anúncio
igualar a base tecnológica sobre a qual construir. Enquanto muito
experiência está sendo adquirida com implantações piloto (pai

os cenários terão que ser enfrentados eventualmente.

Em qualquer ambiente urbano, os veículos autônomos irão também tem que agir com os pedestres, que provavelmente não seguirão regras de baixo tráfego em todos os momentos. O veículo precisará reagir com segurança para pedestres mal-comportados, crianças imprevisíveis, e brincalhões.

Uma necessidade subjacente em muitas áreas será para o automação para se comportar de uma forma que seja compreensível para humanos. Com isso, queremos dizer que um ser humano deve ser capaz para perceber prontamente o que a automação pretende fazer e por que exibiu algum comportamento. Isso será importante para áreas como segurança de interação de veículos humanos (é o vehicle parando para me deixar atravessar? Ou nem percebe mim? Como faço algo semelhante ao contato visual com um veículo autônomo para garantir que não me atropele?); cobertura de teste (o veículo parou porque viu uma criança na faixa de pedestres, ou porque algumas folhas soprando confundiram isto?); e compreensão do design (onde está a parte do masistema de aprendizagem chine que sabe o que significa ver um criança, e cobre todas as crianças ou apenas as do conjunto de treinamento que tudo apontava para a fantasia procurando veículo de coleta de dados?).

Legal

Um problema inicial significativo na implantação desses veículos será estar lidando com questões legais de responsabilidade. (Marchant 2012) Quando um veículo totalmente autônomo está envolvido em um acidente, pode muito bem ser que o passageiro do veículo esteja justificadamente indiferente atento (talvez até dormindo). Os registros de dados do veículo podem ser a principal fonte de informação disponível sobre o que aconteceu encurralado em um acidente. No entanto, os dados de um veículo que tem com defeito não pode ser cegamente confiável. Afinal, se o veículo causou um acidente devido a um mau funcionamento, por que deveria presumimos que quaisquer dados desse sistema com defeito é preciso? Embora se possa imaginar uma independência satisfatória sistema de gravação de dados dentados para análise forense de acidentes, como um sistema tem que ser intencionalmente projetado de maneira adequada. Não seria nenhuma surpresa se dados de eventos atualmente implantados

2013), as questões jurídicas em torno dos veículos autônomos ainda estão muito abertos, assim como vários outros tópicos jurídicos relacionados. (Transport Styrelsen 2014).

Aceitação social

A aceitação social dos veículos autônomos, sem dúvida ser um processo complexo. (Anderson 2014) Um incentivo primário tiva à adoção é a expectativa de que os veículos autônomos serão, em geral, motoristas mais seguros do que pessoas. no entanto é irrealista, especialmente no início, supor que isso irá significa zero percalços. Os casos um tanto simples enfrentar serão aqueles em que evitar uma colisão é fisicamente impossível (por exemplo, uma árvore caindo essencialmente no topo de um carro durante uma tempestade). Mas nem todas as situações serão tão fáceis julgar. Teremos que abordar se o padrão para segurança autônoma deve ser se é melhor do que um ex um motorista humano inteligente, ou apenas um motorista humano *típico* , e exatamente como um driver "típico" pode ser caracterizado. Situações especialmente complicadas serão aquelas em que um ordinenhum motorista humano teria uma boa chance de evitar- em um acidente (pelo menos na visão de um motorista leigo sentado em um júri), mas o veículo autônomo caiu.

Estabelecer uma base atuarial para fins de seguro é frequentemente discutido como um obstáculo significativo para autônomos veículos. Mas no final, talvez isso possa ser resolvido por aplicação eficiente de reservas monetárias. Por exemplo, se um vendedor autônomo de veículos atua como ressegurador, então pode definir uma taxa de resseguro arbitrariamente baixa, subsidiando a tecnologia adoção de tecnologia e, em seguida, ajuste as taxas e projeto do veículo à medida que as taxas de perda reais se tornam aparentes.

Conclusão

No final, terá que haver uma estratégia de certificação de segurança egy de algum tipo para veículos totalmente autônomos. Esta estratégia egy deve abordar as preocupações interdisciplinares de segurança segurança, confiabilidade de hardware, validação de software, robótica, segurança, testes, interação humano-computador, aceitação social e um quadro jurídico viável. Em cada um de

Uso licenciado autorizado limitado a: b-on: Instituto Politécnico da Guarda. Baixado em 01 de novembro de 2020 às 01:54:15 UTC do IEEE Xplore. Restrições aplicadas.

Página 7

nessas áreas, haverá casos extremos e compensações sutis a ser resolvido, e provavelmente uma negociação significativa de acoplamento cruzado eoffs entre áreas. Algumas dessas compensações já são sendo explorado com implantações de protótipos do mundo real, enquanto outros só se tornarão problemas urgentes quando o a frota implantada aumenta. Este artigo em particular pontos o desafio de validar o aprendizado de máquina sistemas para os níveis ultra-confiáveis exigidos para auton frotas de veículos omous, e como esse desafio se relaciona a um número de outras áreas. Uma tarefa de longo prazo diante de nós é atualizar práticas aceitas para criar um design ponta a ponta e processo de validação que aborda todas essas questões de segurança de uma forma que seja aceitável em termos de custo, risco e ética considerações cal.

sobre os autores

Philip Koopman tem um Ph.D. em Computer Engineering da Carnegie Mellon University. Como Elétrico e Membro do corpo docente de Engenharia da Computação ber na Carnegie Mellon University ele é especialista em segurança de software e de-

[8] P. Domingos, "Algumas coisas úteis para saber sobre aprendizado de máquina," *CACM* , vol. 55, não. 10, pp. 78–87, outubro de 2012.
[9] A. Dosovitskiy e T. Brox, "Inverting convolutional networks with redes convolucionais " , *CoRR* , vol. abs / 1506.02753, 2015.
[10] "FAA, System Design and Analysis," AC 25.1309-1A, 21 de junho de 1988.
[11] M. Feather e B. Smith, "Teste de automação do oráculo para V&V de um autono- planejador da nave espacial mous ", AAAI Tech. Relatório SS-01-04, 2001.
[12] R. Geraerts e MH Overmars, "A comparative study of probabilistic planejadores de roteiro ", em *Proc. Workshop Algorithmic Foundations Robotics* , 2002, pp. 43-57.
[13] T. Gindele, S. Brechtel e R. Dillmann, "Learning driver behavior modelos de observações de tráfego para tomada de decisão e planejamento, " *IEEE Intell. Transp. Syst. Mag.* , pp. 69-79, primavera de 2015.
[14] L. Gomes, "Hidden obstáculos for Google's self-driving cars", *MIT Technol. Rev.* , 28 de agosto de 2014.
[15] Hammett. "Design por extrapolação: Uma avaliação de tolerante a falhas aviônica ", em *Proc. IEEE 20th Conf. Digital Avionics Systems* , 2001.
[16] Honeywell. "Relatório de área de atividade de gerenciamento de avarias para AHS gerenciamento de saúde ", relatório AHS Precursor Task E, DoT FHA Publicação, FHWA-RD-95-047, novembro de 1995.
[17] D. Hume, *An Inquiry Concerning Human Understanding . Nova York* : Collier, 1910.
[18] *Veículos rodoviários: segurança funcional,* padrão ISO 26262, 2011.
[19] N. Kalra e SM Paddock. (2016). *Dirigindo para a segurança: quantas milhas de direção seria necessário para demonstrar a confiabilidade do veículo autônomo ity? Santa Monica, CA* : RAND [Online]. Disponível: http: //www.rand .org / pubs / research_reports / RR1478.html
[20] K. Datta e Koopman, "Um estudo de caso sobre monitoramento de tempo de execução de um sistema de veículo de pesquisa autônomo (ARV) " , *RV* , 2015.
[21] R. Kianfar, P. Falcone e J. Fredriksson, "Safety verification of au- sistemas de direção automatizados ", *IEEE Intell. Transp. Syst. Mag.* , pp. 73-86, Inverno 2013.
[22] Koopman e Wagner, "Desafios em testes de veículos autônomos e

projeto de sistema pendente. Ele tem afilia com o Instituto de Robótica e o Instituto de Pesquisa de software. Ele é um membro sênior do IEEE e ACM. E-mail: koopman@cmu.edu. Correio: Prof. Philip Koopman, CMU / ECE HH A-308, 5000 Forbes Ave., Pittsburgh, PA 15213.

Michael Wagner tem um mestrado em Engenharia Elétrica e de Computação - da Carnegie Mellon University. Ele é o CEO e cofundador da Edge Case Research, LLC, especializada es em testes de robustez de software e software de alta qualidade para autônomo veículos, robôs e sistemas embarcados. Ele também tem um afiliação ao Centro Nacional de Engenharia Robótica. E-mail: mwagner@edge-case-research.com Correio Postal: Michael Wagner, Edge Case Research LLC, 100 43rd Street, Suite 208, Pittsburgh, PA 15201.

Referências

[1] M. Aeberhard, et al., “Experiência, resultados e lições aprendidas com direção automatizada nas rodovias da Alemanha ”, *IEEE Intell. Transp. Syst. Mag.* , pp. 42-57, primavera de 2015.

[2] J. Anderson, et al., *Autonomous Vehicle Technology: A Guide for Policy-fabricantes . Santa Monica CA* : RAND, 2014.

[3] K. Bengler, et al., "Três décadas de sistemas de assistência ao motorista", *IEEE Intell. Transp. Syst. Mag.* , pp. 6–22, inverno de 2014.

[4] Boston Consulting Group. “ *Veículos autônomos em um contexto urbano press briefing* , ”World Economic Forum, novembro de 2015.

[5] Broggi, et al., "Testes extensivos de tecnologias de direção autônoma", *IEEE Trans. Intell. Transp. Syst.* , vol. 14, não. 3, pp. 1403–1415, setembro de 2013.

[6] F. Butler, “A inviabilidade da quantificação experimental da crítica vital confiabilidade física do software ”, *IEEE Trans. SW Engr.* , vol. 19, não. 1, pp. 3-12, Janeiro de 1993.

[7] R. de Lemos, et al., “Engenharia de software para sistemas auto-adaptativos: A segundo roteiro de pesquisa ” , *LNCS* , vol. 7475, pp. 1-32, 2013.

validação ”, em *Proc. SAE World Congress* , abril de 2016.

[23] P. Learner, "The hurdles against autônomo *Vehicles* , " *Automobile* , 22 de junho de 2015.

[24] Leveson, “Software safety: Why, what, how,” *ACM Comput. Surv.* , pp. 125-163, junho de 1986.

[25] Levinson et al., “Rumo à direção totalmente autônoma: sistemas e algoritmos ”, em *Proc. IEEE Intelligent Vehicles Symp.* , 5 a 9 de junho de 2011, pp. 163–168.

[26] T. Luettel, M. Himmelsbach e H. -J. Weunsche, “Autonomous veículo terrestre: conceitos e um caminho para o futuro ”, Proc. IEEE, pp. 1831–1839, maio de 2012.

[27] G. Marchant e R. Lindor, "A próxima colisão entre autono-veículos pesados e o sistema de responsabilidade ”, *Santa Clara Law Rev.* , vol. 52, não. 4, pp. 1321–1340, 2012.

[28] A. Nguyen, J. Yosinski e J. Clune, "Deep neural networks are eas-enganado: previsões de alta confiança para imagens irreconhecíveis, ”em *Proc. IEEE Computer Vision and Pattern Recognition* , 2015.

[29] NHTSA. (2013, maio). Declaração preliminar de política sobre auto-veículos acoplados [Online]. Disponível: http://www.nhtsa.gov/staticfiles/rulemaking / pdf / Automated_Vehicles_Policy.pdf

[30] M. Parent, et al., “Questões jurídicas e certificação do totalmente automatizado veículos: Melhores práticas e lições aprendidas ”, CityMobil2 Rep., junho 11, 2013.

[31] Randell e L. Treleaven, “Problemas de confiabilidade no sistema de computador de-sinal ”, *ACM Comput. Surv.* , pp. 123-165, junho de 1978.

[32] D. Rupp e A. King, "Autonomous driving: A prático roadmap," SAE 2010-01-2335.

[33] “Diretrizes para teste seguro na estrada do protótipo SAE Nível 3, 4 e 5 Automated Driving Systems (ADS), ”SAE J3018, março de 2015.

[34] “Prática recomendada para veículos de superfície: guia de segurança cibernética para sistemas ciber-físicos de veículos ”, SAE J3061, janeiro de 2016.

[35] S. Shladover, “Cooperativa (em vez de autônoma) veículo-rodovia sistemas de automação ”, *IEEE Intell. Transp. Syst. Mag.* , pp. 10-19, Spring 2009.

[36] D. Silver, J. Bagnell e A. Stentz, "Active learning from demonstra-para uma navegação autônoma robusta ”, em *Proc. IEEE Conf. Robótica e automação* , maio de 2012.

[37] Transport Styrelsen. “Estudo piloto de direção autônoma”, sueco Agência de Transporte, Tech. Rep. TSG 2014-1316, 2014.

[38] C. Urmson, et al., "Autonomous driving in urban environment: Boss e o desafio urbano ”, *J. Field Robot.* , pp. 425-466, 2008.

[39] J. Ziegler, et al., "Fazendo bertha drive: Uma jornada autônoma em uma rota histórica ”, *IEEE Intell. Transp. Syst. Mag.* , pp. 8–20, Summer 2014.

Uso licenciado autorizado limitado a: b-on: Instituto Politécnico da Guarda. Baixado em 01 de novembro de 2020 às 01:54:15 UTC do IEEE Xplore. Restrições aplicadas.