

**Escola Superior de tecnologia e Gestão**

**Instituto Politécnico da Guarda**

**Engenharia Informática 2020**

# Segurança de veículos autônomos, um desafio interdisciplinar

Trabalho elaborado no âmbito da unidade curricular de Inteligência Artificial

Nome: Vagner Bom Jesus

Professor: Celestino Pereira

## **Introdução**

Uma previsão típica do futuro dos veículos autônomos inclui as pessoas sendo aliviadas do estresse de dirigir diariamente, talvez até tirando um cochilo a caminho do trabalho. Espera-se que isso seja acompanhado por uma redução dramática nas mortes ao dirigir devido à substituição de motoristas humanos imperfeitos por (presumivelmente) pilotos automáticos computadorizados.

Mas, como fazer com que esses veículos totalmente autônomos sejam realmente seguros não é uma questão simples.

Vai ser descritas uma série de áreas que apresentam desafios significativos para a criação de veículos totalmente autônomos e seguros em comparação com os veículos de alguns anos atrás, com ênfase na dificuldade de validar a autonomia na escala de uma frota de veículos de tamanho real.

A questão não é se os veículos autônomos serão perfeitos (eles não serão). A questão é quando seremos capazes de implantar uma frota de sistemas de direção totalmente autônomos que sejam seguros o suficiente para deixar os humanos completamente fora do circuito de direção.  Uma solução holística será necessária e, necessariamente, deve incluir uma ampla apreciação da gama de desafios (e soluções potenciais) por todas as partes interessadas e disciplinas envolvidas.

Nosso trabalho na construção de argumentos de segurança e mecanismos de segurança em tempo de execução para veículos terrestres autônomos nos ensinou que mesmo entender o que “seguro” realmente significa para veículos autônomos não é tão simples. “Seguro” significa, pelo menos, a implementação correta de comportamentos no nível do veículo, como obedecer às leis de trânsito (que podem variar dependendo da localização) e lidar com perigos não rotineiros na estrada, como linhas de energia derrubadas e inundações. Mas também significa coisas como planejamento de missão de failover, encontrar uma maneira de validar estratégias de aprendizagem com base indutiva, fornecer resiliência em face de prováveis ​​lacunas nos requisitos de sistema implantados antecipadamente e ter uma estratégia de certificação de segurança apropriada para demonstrar que um nível suficiente de segurança foi realmente alcançado.

Assim, conseguir um veículo autônomo seguro não é algo que possa ser resolvido com uma única solução tecnológica. Em vez disso, é um conjunto acoplado de problemas que devem ser resolvidos de maneira coordenada e entre domínios.

O restante deste artigo descreve algumas áreas de problemas gerais e algumas das interações entre elas. À medida que todos ganham mais experiência com a tecnologia, sem dúvida, mais alguns problemas de alto nível e muitos problemas mais detalhados surgirão, mas este é um ponto de partida para entender o quadro geral.

## **Engenharia segura**

Vamos começar assumindo que já temos a implantação em pequena escala de veículos totalmente autônomos de Nível 4 (NHTSA 2013) na estrada que geralmente são bem-comportados.  Assim, começamos com a expectativa de que a maioria dos veículos funcionará bem na maior parte do tempo nas condições ambientais cotidianas da estrada. Agora queremos implantar em escala. O desafio passa a ser o gerenciamento de falhas que são muito raras para um único veículo, mas, mesmo assim, acontecerão com muita frequência para serem aceitáveis ​​à medida que a exposição aumenta para milhões de veículos em uma frota.

Existe um corpo de conhecimento bem desenvolvido sobre como tornar seguros os sistemas automotivos baseados em computador e uma história ainda mais longa de criação de sistemas baseados em computador essenciais para a segurança de 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.

A prática atual aceita para segurança de sistema baseado em computador de veículo é normalmente baseada em uma abordagem de segurança funcional (por exemplo, o padrão de segurança ISO 26262 específico para automóveis).

Uma questão importante é que a ISO 26262 geralmente dá a um sistema crédito para um motorista humano ser o responsável pela segurança.

Mas, com um veículo totalmente autônomo, o humano não será responsável por dirigir. Contar com a autonomia para ser totalmente responsável pela segurança do veículo sem a supervisão do motorista é uma grande mudança em comparação com os sistemas avançados de assistência ao motorista atualmente implantados, que dependem amplamente do motorista para ser 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 como zero para um sistema autônomo, o que poderia aumentar drasticamente os requisitos de segurança de uma variedade de funções automotivas em comparação com os veículos de hoje. Se isso vai funcionar, ou mesmo se a ISO 26262 pode ser usada efetivamente como está para validar veículos autônomos é uma questão interessante, mas em aberto.

Uma preocupação significativa da certificação de segurança é validar qualquer uso por veículos autônomos de comportamento de sistema auto-adaptativo.

Adaptação irrestrita, como aprendizagem em tempo real de novos comportamentos significa que um veículo pode ter um comportamento diferente durante a operação do que foi exibido durante o teste e a certificação.

As abordagens de certificação atuais são essencialmente incapazes de lidar com essa situação, porque exigem considerar todos os comportamentos possíveis do sistema desde o início, no processo de design e validação. A menos que os limites sejam colocados de alguma forma na adaptação e totalmente explorados durante o projeto do sistema, pode ser impossível garantir a segurança de tal sistema no momento do projeto, porque o sistema que está sendo testado não terá o mesmo comportamento que um sistema adaptado que é implantado.

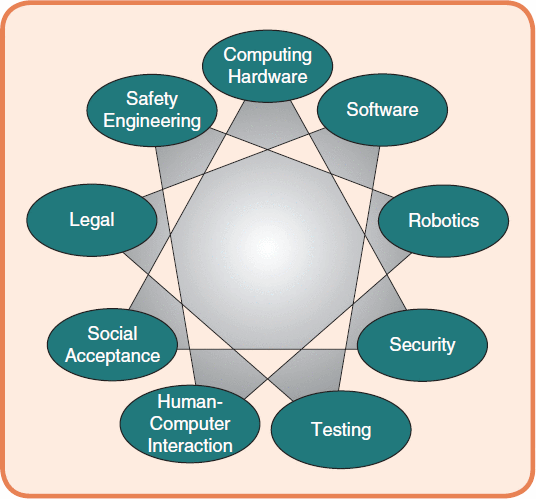
Abordagens de método formal podem ser capazes de provar propriedades sobre sistemas adaptativos, mas tais provas vêm com suposições que não são necessariamente prováveis ​​ou testáveis, e tais abordagens atualmente não escalam bem para sistemas de software de tamanho real.

Observe que por "adaptativo" queremos dizer que o sistema muda essencialmente seus comportamentos dependendo de seu histórico operacional, por exemplo, usando aprendizado de máquina on-line.

Esta é uma gama de comportamentos muito mais dinâmica do que a vista em sistemas baseados em controles mais tradicionais, como o controle de cruzeiro adaptativo, que pode ser validado usando métodos mais tradicionais.

Queremos dizer que o sistema muda essencialmente seus comportamentos, dependendo de seu histórico operacional, por exemplo, usando aprendizado de máquina on-line.

Esta é uma gama de comportamentos muito mais dinâmica do que a vista em sistemas baseados em controles mais tradicionais, como o controle de cruzeiro adaptativo, que pode ser validado usando métodos mais tradicionais.



**Figura 1**

Muitas áreas diferentes requerem uma abordagem coordenada e interdisciplinar para garantir a segurança.

Uma abordagem comumente mencionada para proteger a segurança em nível de sistema para sistemas altamente autônomos é reconectar o motorista quando há uma falha de equipamento, fornecendo uma rede de segurança humana para automação. Por exemplo, um humano pode estar tirando uma soneca e precisará de tempo para adquirir consciência situacional suficiente para assumir a responsabilidade de dirigir. Para preencher a lacuna de desatenção humana, o veículo precisará ter algum tipo de capacidade de autonomia operacional de falha para continuar até que um humano possa recuperar o controle.

Felizmente, os carros normalmente podem atingir um estado seguro em segundos (parar no acostamento), em comparação com as horas para os aviões (voar para um aeroporto de desvio). Assim, uma estratégia de segurança eficaz pode ser que os veículos mudem os modos operacionais para uma “missão de segurança” de curta duração quando um componente primário crítico 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 pode ser projetado de uma forma menos complexa do que um sistema de autonomia total de direção.

Por exemplo, um subsistema de segurança pode permanecer na faixa atual enquanto para e, assim, dispensar os sensores e sistemas de controle necessários para manobras de mudança de faixa. Além disso, um curto tempo de missão (segundos, não horas) provavelmente facilitaria os requisitos de confiabilidade e redundância no próprio subsistema de segurança. Como benefício adicional, projetar uma capacidade de missão de desligamento seguro pode relaxar os requisitos de segurança na autonomia do veículo principal. Se uma missão de segurança está sempre disponível, a autonomia primária não precisa ser falha operacional. Em vez disso, pode ser suficiente para garantir que uma missão de segurança seja invocada sempre que houver uma falha do sistema de autonomia primário, permitindo um sistema de autonomia primário menos que perfeito, desde que as falhas sejam detectadas com rapidez suficiente para invocar uma missão de segurança. Relaxar os requisitos de segurança na autonomia primária (enquanto mantém o veículo como um todo seguro) pode oferecer uma redução dramática no custo e complexidade geral do sistema. (Koopman 2016).

## **Robôs ultraconfiáveis**

Fazer com que os sistemas autônomos (que são robôs) funcionem em uma ampla variedade de situações de direção cotidianas, como foi feito nos protótipos atuais, é uma conquista verdadeiramente significativa e impressionante. No entanto, fazê-los funcionar bem o suficiente para atingir os níveis de segurança exigidos para uma frota de veículos totalmente autônomos exigirá outro conjunto significativo de conquistas.

Por exemplo, considere um possível objetivo de tornar os carros totalmente autônomos tão seguros por hora de operação quanto os aviões. Isso exigiria um nível de segurança de cerca de 1 bilhão de horas de operação por evento catastrófico.

Vamos chamar essa meta de segurança de "ultra-confiabilidade".

Vários desafios surgirão na obtenção de veículos autônomos ultra-confiáveis, começando com a melhoria da robustez do sistema para situações ambientais complicadas (por exemplo, lidar com detritos, desordem e ruído do sensor). Em geral, parece implausível projetar um sistema que possa lidar perfeitamente com todas as situações ambientais possíveis, especialmente nos estágios iniciais de implantação de uma frota. Assim, parece desejável garantir que os sistemas não sejam frágeis e, em particular, tenham alguma forma de saber quando eles não estão funcionando corretamente. Em outras palavras, esses sistemas precisam ser capazes de automonitorar sua confiança em sua própria operação adequada e ser muito bons em saber quando não sabem o que está acontecendo. Será difícil conseguir uma detecção confiável de degradação do sistema. Uma alta taxa de falsos negativos levará a veículos operando involuntariamente de forma insegura. Mas uma alta taxa de falsos positivos deixará muitos carros presos na beira da estrada devido a um alarme cibernético de angústia cibernética (esperançosamente após ter executado uma missão de segurança bem-sucedida em resposta à falha de autonomia).

Outro desafio significativo é que as técnicas de aprendizado de máquina (Domingos 2012), como classificadores que são amplamente usados ​​em veículos totalmente autônomos (por exemplo, Aeberhard 2015),

tendem a ser baseadas em abordagens de treinamento indutivo em vez de um processo de design baseado em requisitos mais tradicional. Validar o raciocínio indutivo é conhecido por ser inerentemente difícil (Hume 1748), e não parece haver uma maneira de fazer níveis ultra-confiáveis ​​de garantias de como tal sistema se comportará quando encontrar dados que não estejam no conjunto de treinamento nem conjunto de dados de teste. Os veículos autônomos operam com dados altamente dimensionais, com fluxos de dados de alta taxa de vídeo, LIDAR e radar. É certo que eles serão expostos a dados do mundo real que diferem de alguma forma dos dados de treinamento e validação.

Os resultados do aprendizado de máquina geralmente envolvem regras de decisão que geralmente são inescrutáveis ​​para revisores humanos.

Assim, é difícil raciocinar sobre a correção do comportamento do sistema de aprendizado de máquina diante de novos dados.

### Programas

Segurança de software é um tópico de pesquisa de longa data.

Com sistemas adaptativos e de aprendizado de máquina, pode ser desafiador articular os requisitos do sistema de uma forma que suporte o V ou outros processos tradicionais de engenharia de sistema.

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 sua precisão em um conjunto de validação é suficientemente alta levanta a questão de saber se o sistema funcionará realmente como precisa quando confrontado com a confusão do mundo real.

O que realmente importa é que o conjunto de validação do aprendizado de máquina deve ser abrangente o suficiente para garantir que não haja lacunas no comportamento do sistema. Em termos do modelo “V”, o conjunto de treinamento é o mais próximo que temos dos requisitos do sistema, e o conjunto de validação é o mais próximo que temos de um plano de teste. Porém, saber que os conjuntos de treinamento e validação são bons o suficiente não é tão fácil. Como podemos ter certeza de que os casos extremos e as interações comportamentais sutis que provavelmente afetam a segurança são realmente aprendidos pelo sistema?

Esperançosamente, o conjunto de treinamento e o conjunto de validação são extremamente abrangentes e cobriram todos os cenários operacionais concebíveis. Mas, como acontece com os sistemas mais tradicionais, ainda há a questão dos cenários operacionais inconcebíveis que ainda podem acontecer (aquelas famosas “incógnitas desconhecidas”). Além disso, existe a possibilidade de que algum novo cenário operacional pareça comum para uma pessoa (e, portanto, não incluído no conjunto de dados de teste), mas na verdade é excepcional de alguma forma para o algoritmo de aprendizado de máquina, potencialmente causando comportamentos imprevistos do sistema.

/\*\*/

Basear um argumento de segurança na suficiência dos dados de treinamento e validação também torna potencialmente crítico o sistema que coleta esses dados.

Afinal, a segurança de tais sistemas depende, em última análise, da precisão da coleta de dados de treinamento e validação. Isso pode, por exemplo, levar à necessidade de o sistema de coleta de dados ser desenvolvido de acordo com os padrões de software de segurança crítica, com atenção para reduzir riscos, como enviesamento não intencional ou distorção nos dados coletados.

Uma solução potencial para o problema de validação do aprendizado de máquina é definir separada e independentemente o que significa operação "segura".

Este conjunto separado de requisitos de segurança poderia ser imposto como um conjunto de requisitos comportamentais monitorados de forma independente sobre a autonomia do veículo autônomo.

Esse monitoramento pode ser usado durante a validação, teste na estrada e talvez implantação para garantir que o veículo não exiba comportamentos inseguros, mesmo se houver lacunas ou falhas nos sistemas de aprendizado de máquina.

### Hardware Informático

Mesmo com o uso de uma estratégia de missão de segurança, um desafio óbvio de hardware é criar hardware de custo ultrabaixo com comportamentos de falha seguros. Isso requer a criação de uma arquitetura combinada de hardware / software que emprega redundância de maneira adequada.

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 para cálculos. No entanto, uma abordagem mais completa para garantir redundância suficiente e tolerância a falhas provavelmente será necessária. Há uma longa história de projeto de tais sistemas para aeroespacial e outras aplicações críticas de segurança, fornecendo um rico conhecimento prévio.

Um desafio de hardware mais sutil, mas crítico, é a questão da detecção de falhas latentes. A redundância totalmente livre de falhas é normalmente assumida no início de cada missão ao realizar cálculos de confiabilidade. Qualquer falha não detectada prejudica drasticamente os benefícios da redundância. Mesmo um ou dois por cento das lacunas no autodiagnóstico têm implicações dramáticas para a confiabilidade alcançável. Por exemplo, atingir apenas 950/0 de cobertura de teste pode reduzir a confiabilidade alcançável do sistema automotivo redundante em ordens de magnitude.

A razão para isso é que as falhas não diagnosticadas podem se acumular por toda a vida útil do veículo, portanto, a probabilidade de ocorrer várias falhas independentes não diagnosticáveis ​​durante a vida de um veículo é bastante alta em comparação com a probabilidade de várias falhas ocorrendo durante uma única missão de direção para partes diagnosticadas do sistema. Assim, será importante criar chips que possam ser autotestados antes de cada ciclo de condução com um nível extremamente alto de cobertura de diagnóstico.

### Testando

As práticas e regulamentações tradicionais de segurança veicular pré-computadorizada enfatizam os testes em nível de veículo. Mais recentemente, os projetos de protótipos de veículos autônomos têm enfatizado a importância dos testes na estrada.

No entanto, é bem sabido que uma abordagem somente de teste é insuficiente para garantir a segurança até de sistemas críticos baseados em software não autônomos.

Mais do que apenas testes são necessários para a implantação em escala real de qualquer software automotivo de segurança crítica. No entanto, testes completos ainda são necessários.

O teste rigoroso de autonomia envolve vários problemas. O principal entre eles é que no modelo de desenvolvimento “V”, o teste compara um documento de design rigorosamente definido com um sistema para determinar se o sistema corresponde ao seu design e requisitos. Para sistemas probabilísticos como planejadores

espera-se que o comportamento do sistema seja diferente em cada execução de teste, mesmo para condições iniciais essencialmente idênticas. Além disso, pequenas mudanças nas condições iniciais podem resultar em grandes mudanças no comportamento do sistema. Portanto, o oráculo de teste (algo que prediz qual seria um resultado de teste correto) precisará suportar a análise abstrata do resultado do teste em vez da técnica mais tradicional de alimentar alguns valores específicos em um pedaço de software e esperar um resultado específico e exato de um computação.

Os sistemas de aprendizado indutivo são ainda mais desafiadores de testar, porque não existe um design como tal e, portanto, nenhum ponto de partida para construir um oráculo de teste. Em vez disso, conforme discutido, há um conjunto de dados de treinamento e um conjunto de dados de validação. No entanto, mesmo com conjuntos de dados abrangentes, não está claro como garantir que o sistema de aprendizado de máquina tenha treinado nas características essenciais dos dados de treinamento, em vez de correlações coincidentes.

As técnicas de aprendizado de máquina são bastante sofisticadas, mas um argumento típico para correção acaba sendo de natureza estatística. Embora alguns possam presumir que as estatísticas de precisão citadas de um sistema sustentam qualquer entrada concebível, na verdade, eles medem apenas o desempenho nos dados de teste e podem ser muito diferentes em diferentes conjuntos de dados que encontram na natureza (por exemplo, Nguyen 2015). Qualquer alegação de segurança deve argumentar que não há situações relevantes para a segurança ausentes dos conjuntos de dados de treinamento e teste. Isso pode funcionar bem para alcançar confiabilidade moderada por meio de abordagens de força bruta (por exemplo, precisão de 99,9%). No entanto, não está claro como garantir a ultra-confiança para algoritmos de aprendizado de máquina.

Considere um sistema de exemplo que executa 10.000 operações por hora para controle de veículos (cerca de 3 por segundo) e uma frota de um milhão de veículos. O teste para validar uma taxa de falha específica requer o processamento de mais casos de teste do que a taxa de falha desejada. Assim, garantir menos de uma falha catastrófica por hora para esta frota exigiria passar significativamente mais de 10 bilhões de casos de teste *representativos*. Essa estimativa exige, adicionalmente, a satisfação da suposição otimista de que as falhas são independentes e que o conjunto de dados é realmente representativo de tudo que uma frota de veículos encontrará. Também pode ser uma meta de segurança muito modesta, porque se cada veículo da frota rodar apenas uma hora por dia, isso ainda permitiria uma falha catastrófica diária do veículo.

Em outras palavras, os desenvolvedores teriam que testar por mais horas de exposição do que afirmam como a taxa de falha da frota. Coletar tantos dados de teste será claramente um desafio significativo, assim como validar essa quantidade de dados sintéticos como sendo realistas em todos os aspectos. (Kalra 2016) Se o objetivo for estilo aviação de garantir que toda a sua frota não sofra uma falha catastrófica de software durante sua vida operacional (FAA 1988), então é difícil ver como isso pode ser feito apenas por meio de testes.

Outra dimensão do teste é a injeção e gerenciamento de falhas. Com a implantação de veículos em grande escala, ocorrerão ocorrências diárias de falha do equipamento do veículo simplesmente devido ao grande número de veículos na frota. Se a controlabilidade do veículo for totalmente de responsabilidade de um sistema de autonomia, será necessário caracterizar o que acontece quando o sistema de autonomia tem que lidar com um veículo que sofre um estouro de pneu, falha de sensor, falha de 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 cada vez mais atenção e não mostra sinais de se tornar um problema fácil. Obviamente, os veículos autônomos também terão que lidar com a segurança. (SAE J3061).

Além de ataques a veículos específicos, as medidas de segurança precisarão abranger ataques e falhas no nível do sistema. Em particular, pode ser problemático confiar cegamente na segurança de outros veículos ou mesmo na infraestrutura de beira de estrada ao realizar manobras autônomas otimizadas, como tráfego de interseção de fluxo livre. Por exemplo, a criptografia de 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 informações maliciosamente incorretas? E se alguém invadiu fisicamente um computador de infraestrutura de beira de estrada e o reprogramou, ou desligou a energia de um conjunto de sistemas de suporte de infraestrutura de beira de estrada em um ataque coordenado?

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

## **Interação Humano-Computador**

À medida que os veículos autônomos substituem os motoristas humanos, a capacidade da automação de se comunicar e cooperar com as pessoas se tornará mais importante. Os riscos de desatenção do supervisor humano em sistemas com quase - mas não totalmente - autonomia total devem ser evidentes. Mas, mesmo os veículos totalmente autônomos precisarão, pelo menos, garantir que os ocupantes sintam que o comportamento do veículo é seguro se quiserem conquistar a confiança do cliente e também precisarão aprender como antecipar o comportamento de outros veículos. (Gindele 2015) Embora alguns possam dizer que a confiança do cliente não é estritamente uma questão de segurança, é uma questão vital para a adoção de tecnologia e, portanto, indiretamente de segurança, se os veículos autônomos cumprirem sua promessa de salvar vidas na estrada. Outros problemas de interação humano-computador também constituem desafios significativos.

Os veículos autônomos terão que interagir com os motoristas humanos 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. Tirar os motoristas humanos de cena provavelmente levará muitos anos enquanto a penetração no mercado de tecnologia totalmente autônoma aumenta. Mesmo quando chegar o dia em que todos os carros serão totalmente autônomos, a estrada ainda será o lar de motoristas humanos de bicicletas, patinetes, cavalos, equipamentos agrícolas e assim por diante. Muitos desses usuários humanos das estradas ficarão relutantes ou serão incapazes de seguir as regras normais de trânsito e as expectativas para veículos de passageiros. Mesmo se houver faixas exclusivas inicialmente para facilitar a implantação (Shladover 2009), com o tempo, parece provável que haverá uma tremenda pressão pública para espalhar a autonomia em cenários mistos de veículos autônomos / humanos. Portanto, parece provável que os cenários de tráfego misto terão que ser tratados eventualmente.

Em qualquer ambiente urbano, os veículos autônomos também terão que agir com os pedestres, que dificilmente seguirão as regras de trânsito em todos os momentos. O veículo precisará reagir com segurança a pedestres mal-comportados, crianças imprevisíveis e brincalhões.

Uma necessidade subjacente em muitas áreas será que a automação se comporte de uma forma que seja compreensível para os humanos. Com isso, queremos dizer que um ser humano deve ser capaz de perceber prontamente o que a automação pretende fazer e por que exibiu algum comportamento. Isso será importante para áreas como a segurança da interação do veículo humano (O veículo está parando para me deixar atravessar? Ou nem mesmo me nota? Como faço algo semelhante ao contato visual com um veículo autônomo para ter certeza de que não me atropelar?); teste de cobertura (o veículo parou porque viu uma criança na faixa de pedestres ou porque algumas folhas soprando o confundiram?); e compreensão do projeto (onde está a parte do sistema de aprendizado de máquina que sabe o que significa ver uma criança,

## **Legal**

Uma questão inicial significativa na implantação desses veículos será lidar 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 desatento (talvez até dormindo). Os registros de dados do veículo podem ser a principal fonte de informações disponíveis sobre o que aconteceu em um acidente. No entanto, os dados de um veículo com defeito não podem ser confiados cegamente. Afinal, se o veículo causou um acidente devido a um mau funcionamento, por que devemos presumir que todos os dados desse sistema com defeito são precisos? Embora se possa imaginar um sistema de registro de dados independente e satisfatório para a perícia forense de acidentes, tal sistema deve ser intencionalmente projetado de maneira adequada.

Uma questão importante de responsabilidade será determinar quem é o responsável pela operação adequada do veículo. É o ocupante que entrou em um veículo alugado mesmo com danos no sensor que deveriam ser notados por um leigo? Foi o fabricante do veículo que confiou em um conjunto de dados de treinamento de terceiros com defeito? Foi o mecânico que instalou por engano uma versão ligeiramente incompatível do software do sensor substituto? 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 evitar um acidente malicioso? Embora algumas dessas questões possam ser puramente legais, a resolução de muitas questões jurídicas exigirá uma base tecnológica adequada para a construção. Embora muita experiência esteja sendo adquirida com implantações piloto (Pai 2013), as questões jurídicas em torno dos veículos autônomos ainda estão em aberto, assim como uma série de tópicos jurídicos mais gerais. (Transport Styrelsen 2014).

## **Aceitação social**

A aceitação social de veículos autônomos será sem dúvida um processo complexo. (Anderson 2014) Um incentivo primário para a adoção é a expectativa de que os veículos autônomos, em geral, sejam motoristas mais seguros do que as pessoas. No entanto, não é realista, especialmente no início, supor que isso significará zero acidentes. Os casos um tanto diretos a serem considerados serão aqueles em que evitar uma colisão é fisicamente impossível (por exemplo, uma árvore caindo essencialmente em cima de um carro durante uma tempestade). Mas nem todas as situações serão tão fáceis de julgar. Teremos que abordar se o padrão para segurança autônoma deve ser se é melhor do que um excelente motorista humano, ou apenas um *típico*motorista humano, e exatamente como um motorista “típico” pode ser caracterizado. Situações especialmente complicadas serão aquelas em que um motorista humano comum teria uma boa chance de evitar um acidente (pelo menos na visão de um motorista leigo sentado em um júri), mas o veículo autônomo bateu.

O estabelecimento de uma base atuarial para fins de seguro é frequentemente discutido como um obstáculo significativo para veículos autônomos. Mas, no final, talvez isso possa ser resolvido pela aplicação adequada de reservas monetárias. Por exemplo, se um fornecedor de veículos autônomo atua como ressegurador, ele pode definir uma taxa de resseguro arbitrariamente baixa, subsidiando a adoção de tecnologia e, em seguida, ajustar as taxas e o design do veículo conforme as taxas de perda reais se tornam aparentes.

## **Conclusão**

No final, terá que haver uma estratégia de certificação de segurança de algum tipo para veículos totalmente autônomos. Essa estratégia deve abordar as questões interdisciplinares de engenharia de segurança, confiabilidade de hardware, validação de software, robótica, segurança, testes, interação humano-computador, aceitação social e uma estrutura jurídica viável. Em cada uma dessas áreas, haverá casos extremos e trocas sutis a serem resolvidas, e provavelmente trocas de acoplamento cruzado significativas entre as áreas. Algumas dessas compensações já estão sendo exploradas com implantações de protótipos do mundo real, enquanto outras só se tornarão problemas urgentes quando a frota implantada aumentar. Este artigo, em particular, aponta o desafio de validar sistemas baseados em aprendizado de máquina para os níveis ultra-confiáveis ​​necessários para frotas de veículos autônomos, e como esse desafio se relaciona a uma série de outras áreas. Uma tarefa de longo prazo diante de nós é atualizar as práticas aceitas para criar um projeto ponta a ponta e um processo de validação que trate de todas essas questões de segurança de uma forma que seja aceitável em termos de custo, risco e considerações éticas.