人工知能を搭載する自動車のセキュリティ論証について

溝口 誠一郎[1]　櫻井 幸一[2]

概要：自動車の安全性を主張するためには，自動運転車の設計とその運用における様々なリスクが管理されていることを説明する必要がある．サイバーセキュリティ上のリスクも自動車の安全性に関わるリスクとして，それが適切に管理されていることを主張しなければならない．本稿では，自動車の安全論証の標準・ベストプラクティスを参考に，人工知能を搭載する自動運転車のセキュリティの説明戦略について考察する．

キーワード：自動車，機械学習/AI，安全論証

Argument of Cybersecurity for Vehicles using AI

Seiichiro Mizoguchi[1]　Kouichi Sakurai[2]

Abstract: To claim the safety of automated vehicles, it is necessary to explain that the risks during development and operation of automated vehicles are properly managed. Cybersecurity risks which affect the safety of autonomous vehicles also have to be managed. In this paper, we show the best practices of safety argument of vehicles and study about argument strategy of cybersecurity for autonomous vehicles using AI.

Keywords: Vehicle, Machine Learning/AI, Safety Argument

1. はじめに

自律走行機能を持つ自動車にとって，その安全性を説明できることは,社会から受容される上で必要な要件である．2021 年に施行された UN-R155/156/157[1][2][3]は,自律走行機能の 1 つである車線維持装置の基準 (UN-R157)と，サイバーセキュリティ (UN-R155) および自動車のソフトウェアアップデート (UN-R156) に関する基準であり，自動運転機能およびそのサイバーセキュリティとソフトウェア更新に対して，安全性への企業の取り組みを問うものとなっている．

自動車分野では，機能安全規格 (ISO 26262[4]) が安全論証の主流であったが，自動運転技術の進歩により，意図機能の安全規格 (ISO 21448[5]) も標準化された．サイバーセキュリティについては，サイバーセキュリティ規格 (ISO/SAE 21434[6]) が標準化されている．自動車分野では，現状，これらの安全規格に沿ってリスクマネジメントすることが業界として求められつつある．

2. 自律走行製品の安全論証

安全性について主張するためには，適切な論証構造を意識する必要がある．自律走行製品の安全性評価のガイドラインとして ANSI/UL4600[7]がある．UL4600 では，セーフティケースに対するアセスメントを重視した規格であり，構築されたセーフティケースにおいて，自律運転特有の機能/能力に対するリスクが考慮されているか，システムのディペンダビリティの観点でセーフティに関わるリスクが考慮されているかなどが評価される．セーフティケースの構造そのものに対する評価基準も存在し,『セーフティケースが，要求，論証および証拠に対して，一貫性をもつ規定されたフォーマットを使用しなければならない』とし,『高度に構造化された方法でセーフティケースを整理するための確立された手段（例えば、SACM[8]や GSN[9]など）を用いる』ことが推奨されている．

3. 自動車のサイバーセキュリティの論証

サイバーセキュリティについては，ISO 21434 に基づいた活動によって生成される成果物がサイバーセキュリティケースを支える証拠となり，各成果物の意味や関係によって論証構造が説明される．

具体的な活動としては，リスク分析のための，資産の識別，脅威の識別，損害の識別，リスク値算定，リスク評価，リスク対応といった基本的なサイバーセキュリティの活動を，開発時のコンセプトレイヤーとシステム開発レイヤーごとに実行し，証拠を固めていく．サイバーセキュリティに関わるリスクは，生産時あるいは車両の運行時にも起きるため，ライフサイクルを通じてサイバーセキュリティのリスクが管理されていることを示すことも，サイバーセキ

1 DNV ビジネスアシュアランスジャパン
DNV Business Assurance Japan
2 九州大学
Kyushu University

ュリティケースには含まれる．UN-R155 は，製品としてのサイバーセキュリティが担保されていることを示す前に，自動車のセーフティに関わるサイバーセキュリティリスクが組織として適切に管理されていることを示さなければならない．これが，CSMS (Cybersecurity Management System) 認証と呼ばれるものであり，ISO 21434 に従って適切に論証できる，ということは UN-R155 の CSMS の能力を持っていることを主張することにもつながる．

4. 機械学習/AI を用いるシステムのセキュリティ論証

自動運転車において，自律走行に関わる機能がセキュリティのリスクから合理的なレベルで保護されていることを論証する必要がある．例えば，UN-R157 の車線維持装置であれば，特定の運行設計領域 (ODD) において，車線を認識し，進路を計画し，ステアリングを制御するというタスクを実行する．この認識の場面において機械学習/AI が用いられることが想定される．

UL4600 では，8 章で『自律機能及び支援』に関するセーフティケースへの評価項目が規定されている．そのなかで，8.4 節が『認知』に関わる内容であり，認知機能が許容できる機能的性能を提供できることを説明できる必要がある．また，8.5 節では，システムで利用される機械学習及び"AI"の手法の説明に対する評価項目が規定されている．セキュリティの観点でこれらの能力に対するリスクを考慮する必要がある．ISO/IEC TR 24028[8]では，機械学習・人工知能を利用するシステム特有のセキュリティ課題として，Data Poisoning，Adversarial Attack，Model Stealing，Hardware-focused Threats を挙げているが，認知機能に関わる攻撃としては，Data Poisoning と Adversarial Attack が該当する．これらの攻撃（脅威）は，データ汚染の場合，8.5.3 の『機械学習の訓練及び V&V は，許容できるデータを使用しなければならない』という要件に対して影響を与えるものであり，Adversarial Attack は，8.5.4 の『機械学習に基づく機能は，データの多様性に対して，許容できる程度にロバストでなければならない』という要件に対して影響を与えるものとなる．

さて，UL4600 では，サイバーセキュリティは，9 章の『ソフトウェア及びシステムエンジニアリングプロセス』における開発プロセス管理の側面と，10 章の製品の『ディペンダビリティ』確保の観点での 10.8『サイバーセキュリティ』として，論証を整えていくことになる．この 10.8 の活動として，サイバーセキュリティ計画を作成するが，このサイバーセキュリティ計画に人工知能を搭載するシステムによって実現される機能が，学習時と運用時それぞれで，Poisoning や Adversarial Attack に晒され，8.5.3 や 8.5.4 の要件を満たせなくなるというシナリオを考慮する必要がある．また，リスクに対して対策を取るか取らないかは，リスク値の算定結果を元に判断される．リスク値は，ISO21434 では，攻撃実現可能性 (Attack Feasibility) と損害シナリオのインパクト (Impact) の掛け合わせによって算定することとなっている．攻撃実現可能性は，AI を利用するシステムのアーキテクチャとその運用環境に依存するため，世の中で報告される事例がシステムアーキテクチャを踏まえてどれだけ攻撃実現可能性に寄与しているかを分析する必要がある．

さらに，AI の説明可能性やアップデートについても考慮が必要である．安全論証では，「合理的」あるいは「許容できる範囲」という言葉が用いられる．リスクマネジメントでは，許容できるかどうかを論理的に説明できるかが評価（アセスメント）の焦点になるため，用いる AI 技術について不確かなリスクが存在している状態では，合理性の判断がしづらい．アップデートなどの機能改善についても，そのアップデートが能力不足や不具合をどのように改善し，一方で他の機能に対して性能低下（デグレード/リグレッション）が起きていないことを説明するためには，説明性の低い状態では，そのアップデートを承認することが難しくなる．AI の説明性は，性能改善と攻撃方法検討の両方に寄与するため，注視していく必要のある分野である．

自律走行製品を開発する組織にあたっては，これらのリスクが適切に管理されていることを主張する必要がある．

5. おわりに

自律走行機能の実現には，機械学習/AI の技術を活用することが必須であるが，その安全性の論証については統一的な例が見当たらない．自律機能に応じて，そのシステムアーキテクチャが複数存在することが予想されるため，それらのデザインパターンに応じた論証構造を考える必要がある．それにより，自律走行機能の開発における安全性論証のコストを抑え，自律走行製品の普及に貢献できると期待される．

謝辞　本研究は Open RDG-AI/SS 事務局の支援を受けています。

参考文献

[1] UN-R155, Cyber security and cyber security management system
[2] UN-R156, Software update and software update management system
[3] UN-R157, Automated Lane Keeping Systems
[4] ISO 26262, Road vehicles – Functional Safety
[5] ISO 21448, Safety of the intended functionality
[6] ISO/SAE 21434, Road vehicles – Cybersecurity Engineering
[7] ANSI/UL4600, Standard for Safety for the Evaluation of Autonomous Products
[8] Structured Assurance Case Metamodel, https://www.omg.org/spec/SACM/2.2/About-SACM/
[9] Goal Structuring Notation, https://scsc.uk/SCSC-141C
[10] ISO/IEC TR 24028, Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence