预期功能安全国际标准 ISO21448 及中国实践
白皮书

全国汽车标准化技术委员会（SAC/TC114）
道路车辆功能安全标准研究制定工作组
ISO/TC22/SC32/WG8 中国专家组

2020.12.31

前　言

近年来，随着电动化、智能化、网联化技术发展和应用，车辆功能安全和预期功能安全技术和标准在国际上日益受到各方广泛关注，国际标准化组织不断完善功能安全(ISO 26262)和预期功能安全(ISO 21448)标准的同时，联合国（UN/WP.29)、欧盟及美国等相关组织和国家也陆续将功能安全和预期功能安全理念及管理体系引入相关技术法规，特别是自动驾驶汽车安全相关法规。我国国家层面相继出台的多项政策和规划已将功能安全和预期功能安全技术及标准研究上升至国家战略层面。

为加快推动功能安全和预期功能安全（SOTIF）技术和标准在国内应用和实施，全国汽车标准化技术委员会汽车电子与电磁兼容分技术委员会（SAC/TC114/SC29）下设的道路车辆功能安全标准研究制定工作组，制定了“中国功能安全(Functional Safety)和预期功能安全(SOTIF) 技术和标准研究中长期规划(2020-2025)”、“中国功能安全(Functional Safety)和预期功能安全(SOTIF)技术及标准体系”。

该规划和标准体系以国家标准 GB/T 34590《道路车辆 功能安全》、GB/T《道路车辆 预期功能安全》为指导和研究主线，基于中国国情，开展适用于我国新能源汽车、自动驾驶汽车、传统汽车整车和关键电控系统功能安全、预期功能安全（SOTIF）的技术和标准研究，从设计开发源头，避免或降低因车辆电控系统故障、预期功能不足、性能存在局限而导致的安全事故，保障车辆运行安全。

基于中国汽车技术及产业发展实际情况，道路车辆功能安全标准研究制定工作组提出了建立“两纵三横”布局的总体研究方案，如图 1 所

示，即以国家标准 GB/T 34590《道路车辆 功能安全》(Functional Safety)和 GB/T《道路车辆 预期功能安全》 (SOTIF)给出的方法论为“两纵”，构建技术供给体系，提升产业核心能力；以新能源汽车（纯电动、混动、燃料电池汽车）、自动驾驶汽车、传统汽车为“三横”，布局整车功能安全和预期功能安全技术和标准创新链，强化整车及关键系统集成技术创新。

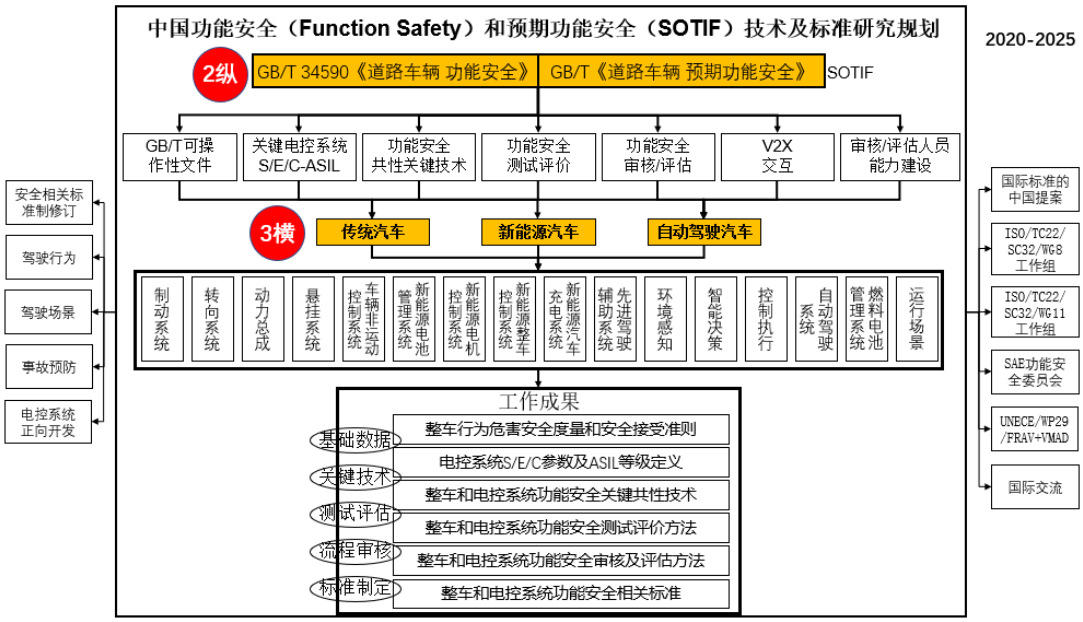

图 1 中国功能安全(Functional Safety)和预期功能安全(SOTIF) 技术和标准研究规划

本白皮书概要介绍预期功能安全（SOTIF）国际标准 ISO 21448 的制定背景、基本概念、方法，回顾国际标准 ISO 21448 的制定进程，介绍中国参与国际标准制定及提案情况，阐述“量化思想的中国提案”在国内落地的研究进展及后续研究规划，为我国汽车行业专家全面了解、正确认识、深入研究、科学应用、不断完善预期功能安全理念、技术及标准体系提供参考。

目标是建立以我国目标市场为主体的自动驾驶汽车预期功能安全(SOTIF)设计开发、验证确认、测试评价、发布的闭环技术和标准体系，

从整车、系统到感知、决策、执行各个环节，将自动驾驶系统设计不足、性能局限在已知/未知危害场景下导致的整车风险控制在合理可接受的范围内，避免竞相推高无止境的累积测试里程及竞相建立大量无效的场景库、降低开发成本、提高开发效率，从源头保障车辆运行安全。

目 录

一、 自动驾驶预期功能安全(SOTIF)的研究背景

多起因自动驾驶汽车引发的致命交通事故表明，依靠传统的以质量保障（关注失效风险的预防、探测和消除，例如：国家标准 GB/T 34590《道路车辆 功能安全》（修改采用 ISO 26262）关注并解决的是因电控系统故障而导致的整车行为危害）为中心的车辆安全体系，已经不能完全满足自动驾驶车辆的安全保障需求，全球汽车工业领域亟需建立全新的自动驾驶安全评判准则体系，以指导正向设计开发和测试评价工作。

为此，国际标准化组织下设的功能安全工作组（ISO/TC22/SC32/WG8）于 2018 年正式启动了全球首个自动驾驶安全国际标准 ISO 21448《道路车辆 预期 功 能 安 全》（Road Vehicles-Safety of The Intended Functionality）的制定工作，旨在为全球自动驾驶车辆的安全开发和测试评价提供技术指导。

ISO 21448《道路车辆 预期功能安全》立足对自动驾驶安全影响更广泛的非故障安全领域，重点关注自动驾驶汽车的行为安全，解决因自身设计不足或性能局限在遇到一定的触发条件（如环境干扰或人员误用）时导致的整车行为危害。

自动驾驶系统中涉及环境干扰和人员误用等外部触发条件的安全开发最为复杂，且受目标市场的影响较大，如何建立一套科学、合理并且广泛适用于各目标市场的安全评价体系是国际标准 ISO 21448 亟需解决的一项重要课题。

针对备受关注的自动驾驶安全，特别是预期功能安全问题，全国汽

车标准化技术委员会秘书处一方面组织骨干专家形成中国专家组深入参与到 ISO 21448 的研究制定工作，提出了大量有影响力的中国提案；另一方面针对行业痛点问题，在国内组织启动了针对性研究工作，取得了阶段性进展，后续章节将给出具体介绍。

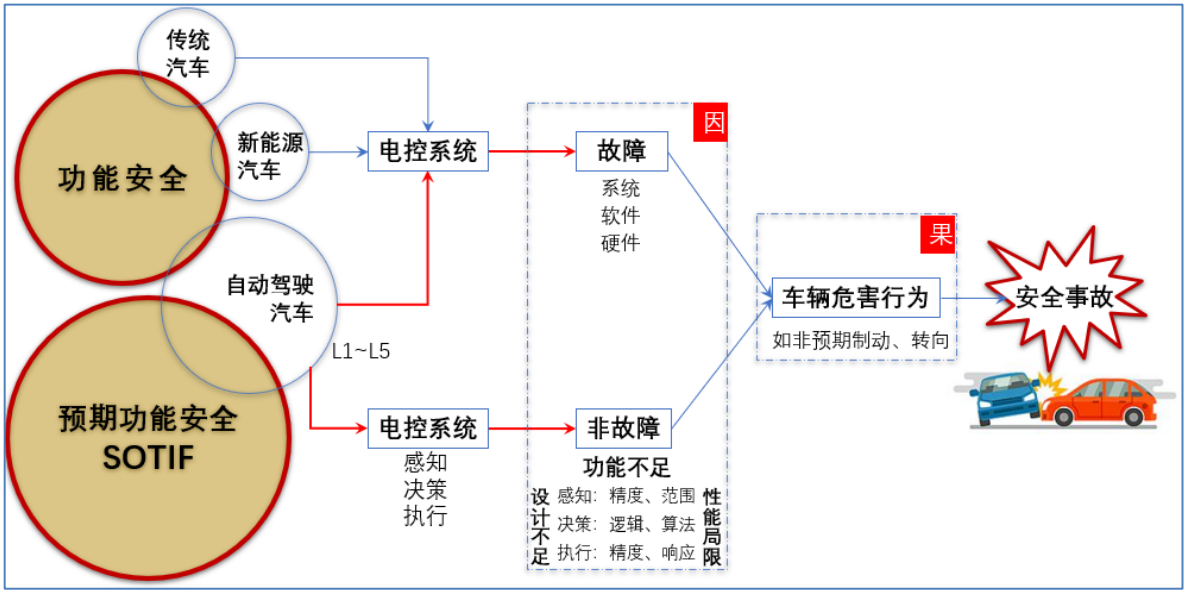

图 2 功能安全和预期功能安全(SOTIF)

图 2 给出了功能安全和预期功能安全(SOTIF)的关注点，总体来讲，功能安全(Functional Safety)和预期功能安全(SOTIF)技术为确保自动驾驶车辆在故障、非故障情况下的安全运行提供了根本保障。

二、 什么是预期功能安全(SOTIF)

预期功能安全(Safety of The Intended Functionality)，重点关注“预期的功能”的安全性，即：满足预期设计要求的功能所具有的安全水平。

因自动驾驶车辆运行场景条件的复杂性和未知性，自动驾驶功能即使满足设计要求，仍可能存在大量的安全运行风险。如何避免预期的功能所引发的安全风险，即为预期功能安全。

预期功能安全的定义：不存在因设计不足或性能局限引起的危害而导致不合理的风险，也就是将设计不足、性能局限导致的风险控制在合理可接受的范围内。

这些设计不足、性能局限在遇到一定的场景触发条件（如环境干扰或人员误用）时，将导致整车行为危害，如图 3 所示。

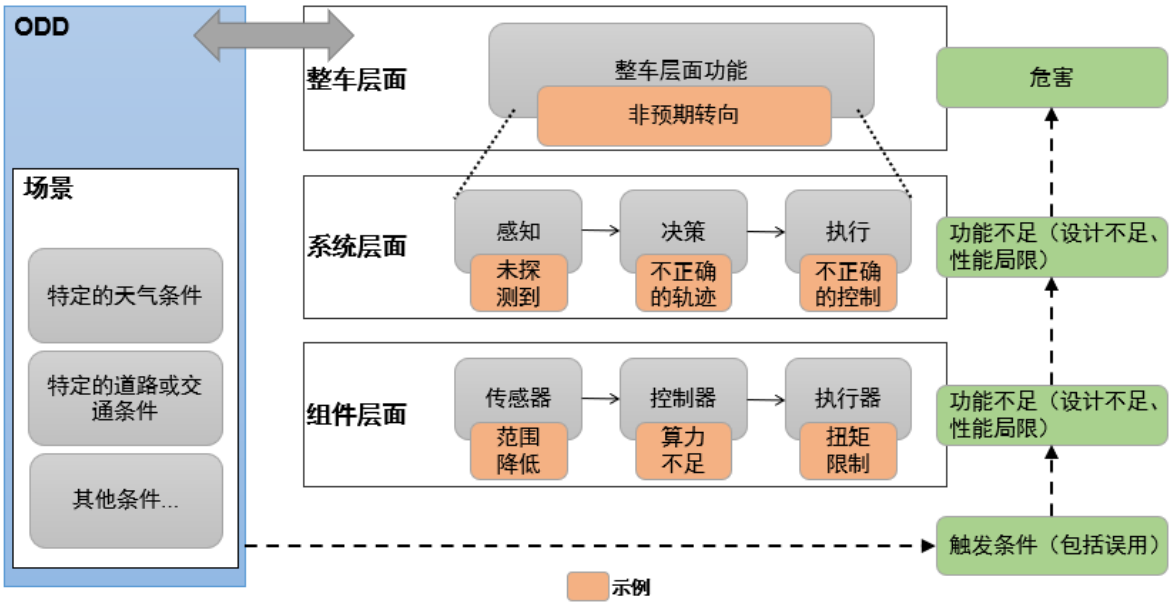

图 3 功能不足、触发条件、整车行为危害的关联

与传统车辆重点关注系统失效预防、探测与减轻不同，自动驾驶车

辆因替代了人类驾驶员的部分或全部驾驶操作行为，更需要关注运行过程中自身功能和性能的行为安全，由于使用场景的复杂性和随机性，自动驾驶系统安全相关的很多问题在设计阶段无法预见。

如图 4 所示，从安全性和已知性角度，将车辆运行场景分为已知安全场景、已知不安全场景、未知不安全场景和未知安全场景 4 个区域,。在开发之初，区域 2 和区域 3 的比例较高，SOTIF 技术通过对已知场景及用例的评估，发现系统设计不足，将区域 2 转化为区域 1，并证明区域 2 的残余风险足够低；针对区域 3，SOTIF 技术基于真实场景及用例测试、随机输入测试等，发现系统设计不足，将区域 3 转化为区域 2，同时基于统计数据和测试结果，间接证明区域 3 的风险控制到合理可接受的水平。由此实现对已知和未知风险的合理控制，完成自动驾驶车辆系统的安全提升和发布。

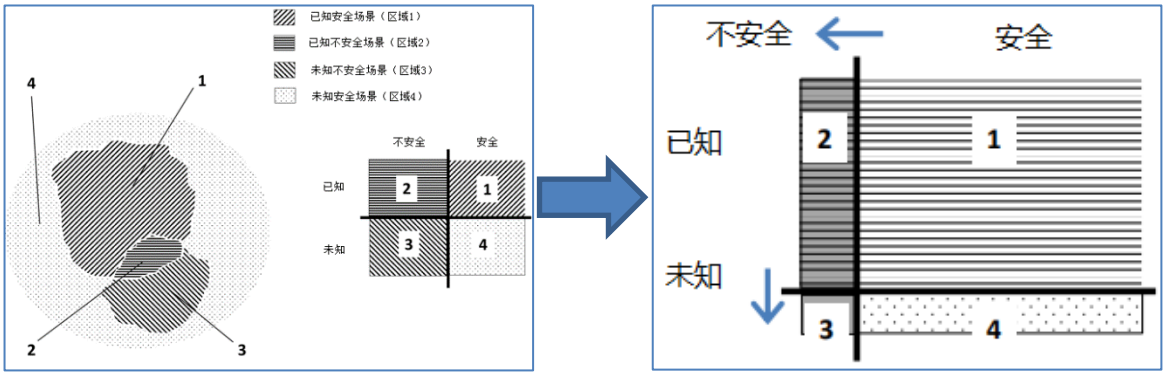

图 4 自动驾驶运行场景分类及 SOTIF 开发、验证和发布演进

自动驾驶系统安全风险的一个主要来源是未知不安全场景区域，对其无法定义需求，也难以量化评价，这成为了全球自动驾驶安全开发领域的痛点。

三、 ISO 21448 制定过程和计划

自 2016 年 2 月，国际标准化组织 ISO 下设的功能安全工作组（ISO/TC22/SC32/WG8）启动了 ISO 21448 的制定工作，参与成员来自法国、德国、美国、英国、中国、以色列、比利时、意大利、瑞典、日本、荷兰、韩国、芬兰、卢森堡、瑞士、爱尔兰、立陶宛、奥地利等 18 个国家的专家。中汽中心标准所（全国汽车标准化技术委员会秘书处）组织国内专家组成中国代表团全程参与该标准的研究制定工作。

经 WG8 功能安全工作组组内起草、协商一致，于 2019 年 1 月，以 PAS（可公开提供规范）形式发布了 ISO/PAS 21448。

在完成 ISO/PAS 21448 草案、尚未发布前，于 2018 年 6 月正式启动了 ISO 21448 的制定工作。

2019 年 12 月，形成了 ISO 21448 CD 版草案，征集各国建议。中国提出了 132 项建议，其中 96 项获得通过。

2020 年 11 月，形成了 ISO 21448 DIS 版草案，征集各国建议。

ISO 21448 计划于 2022 年 3 月发布。

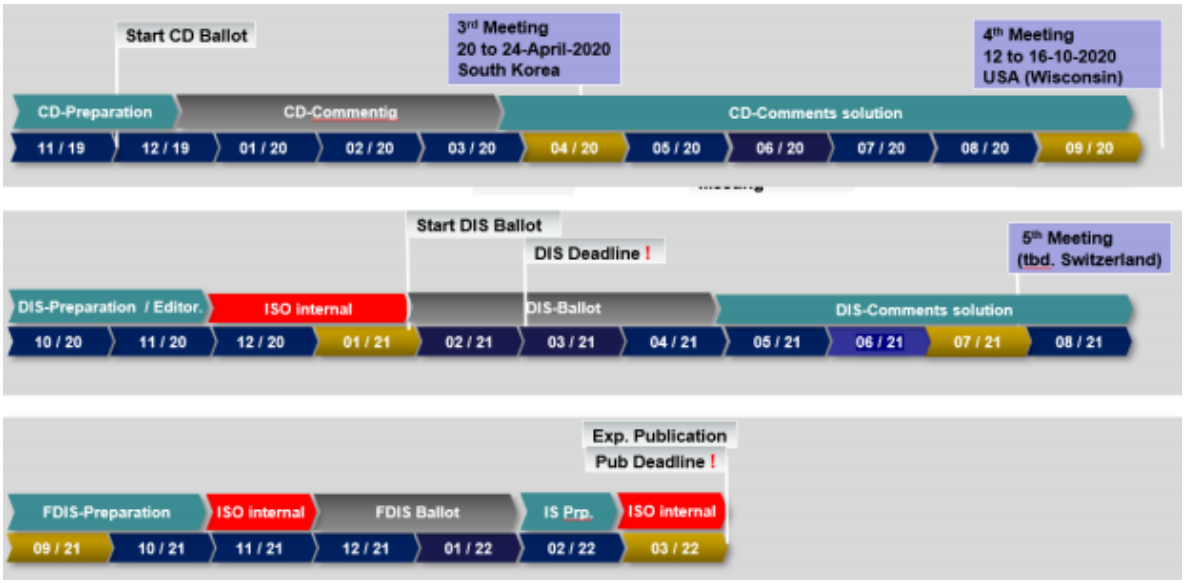

图 5 ISO 21448 制定计划

四、 多项“量化思想的中国提案”作为主线贯穿 ISO 21448

国际标准化组织（ISO/TC22/SC32/WG8）自 2016 年启动 ISO 21448 制定工作，中汽中心标准所（全国汽车标准化技术委员会秘书处）组建中国代表团全程参与，从中国国情出发，围绕自动驾驶预期功能的设计开发、基于已知/未知危害场景的测试、验证/确认、最终发布等关键主题，提出了多项中国提案。

其中，“量化思想的中国提案”，作为主线，贯穿 ISO 21448 中自动驾驶预期功能的定义和设计、危害的识别和评估、功能不足和触发条件的识别和评估、验证和确认（V&V）策略、已知/未知危害场景的评估、SOTIF 发布准则等主要内容：

中国典型提案 1：量化思想写入 ISO 标准制定纲领

2017 年 7 月，在德国博登湖召开的 ISO/TC22/SC32/WG8 全体会议讨论制定 ISO 21448 的标准纲领，中国代表团从我国国情需求出发，提出当前自动驾驶安全性受目标市场影响较大，潜在的安全风险高，亟需建立明确的量化安全接受准则，以规范和引导行业。经过充分讨论，最终该提案写入标准纲领的第 2 条和 6 条。

图 6 ISO 21448 的标准纲领

中国典型提案 2：量化思想的 SOTIF 双层安全接受准则

由于自动驾驶的高度复杂性以及其安全风险的一个主要来源是未知不安全场景区域，为验证并确认其安全性，当前行业主要采用大量的里程累积测试方法，但不断推高的里程数字仍然无法杜绝安全风险，究竟如何科学评价 SOTIF 安全水平成为行业痛点。中国代表团从预期功能安全的目的出发，即将自动驾驶因设计不足、性能局限而导致的风险控制在合理可接受的范围内，提出了量化思想的 SOTIF 双层安全接受准则：

●第一层安全接受准则：自动驾驶危害行为事件接受准则。

自动驾驶运行过程由一系列驾驶行为组成，如果相关行为不当，将可能产生危害风险，最终导致事故的发生。建立自动驾驶过程中危害行为事件的评价准则，包含定量准则和定性准则。其中，可控性指标和 SOTIF 信心度指标是定量准则的重要组成部分。

●第二层安全接受准则：自动驾驶总体安全风险接受准则。

在自动驾驶全部累积行驶里程中，可能发生不止一次的危害行为，特别是里程越高，危害行为的数量及影响可能越大，为了将总体安全风险控制在合理可接受的水平，建立总体安全确认目标，即自动驾驶总体安全风险接受准则，以评价真实道路累积全部行驶里程过程中的安全风险。

该提案反映在如下方面：

① ISO 21448 DIS 版第 3 章术语，3.1 acceptance criterion--接受准则：

3.1
acceptance criterion
criterion representing the absence of an unreasonable level of risk

Note 1 to entry: The acceptance criterion can be of qualitative as well as quantitative nature, e.g. physical parameters that define when a specific behaviour is considered as hazardous behaviour, maximum number of incidents per hour, ALARP, etc.

EXAMPLE 1 From traffic statistics a reasonable level of risk of one accident per X km is derived.

EXAMPLE 2 The comparison with an equivalent vehicle level effect that is proven in use to be controllable by the driver can support the definition of an acceptance criterion. For instance, the trajectory perturbation due to an unwanted lane keeping assist function intervention might be compared to a lateral wind gust to define an acceptable level of authority for the function.

其中，physical parameter 是危害行为的量化指标（第一层量化准则），而 maximum number of incidents 是总体风险接受准则（第二层量化准则）。

② ISO 21448 DIS 版第 6 章 Identification and evaluation of hazards--危害的识别和评估。

6 Identification and evaluation of hazards

6.1 Objectives

The purpose of this clause is to achieve the following objectives:

a) The hazards, defined at the vehicle level, shall be identified.

b) The risk that arises from the hazardous behaviour of the intended functionality in the relevant scenarios shall be systematically identified and evaluated. The parameters that define the hazardous behaviour shall be specified.

NOTE The identification and evaluation of risk includes the consideration of reasonably foreseeable direct and indirect misuse.

c) The acceptance criteria from which the validation targets are derived to evaluate the residual risk shall be specified.

其中，b)中提出应识别危害行为的量化指标（第一层准则），c)中提出应定义第二层接受准则，以得到确认目标，从而评估残余风险。

③ ISO 21448 DIS 版第 6 章 6.4 Risk evaluation--风险评估

6.4 Risk evaluation

The risk evaluation aims to evaluate the risk due to hazardous behaviours in given scenarios; this will help to specify the acceptance criteria of SOTIF-related risk.

风险评估目的是评估给定运行场景下的危害行为风险，以定义 SOTIF 相关风险是否可被接受。

④ ISO 21448 DIS 版第 6 章 6.5 Specification of acceptance criteria--接受准则的定义

6.5 Specification of acceptance criteria

If the risk of harm cannot be sufficiently reduced by functional modifications of the intended functionality then acceptance criteria are defined for the risks associated with the hazardous scenarios. If, by functional modifications of the intended functionality and a re-evaluation of the hazards, an S=0 or C=0 can be achieved then the risk of harm is acceptable and it is not necessary to specify further acceptance criteria for the remaining triggering conditions.

Acceptance criteria considers:

- applicable governmental and industry regulations; and
- if a function is new or already established in the market; and
- whether the risk is acceptable to the people who may be exposed to the risk (e.g. a vehicle owner, the operator, a pedestrian or a passenger in an automated public transport system).

如果对预期功能进行迭代改进后仍不能充分降低安全风险，需要定接受准则并考虑相应的法规、该功能在目标市场的情况、人员暴露在风险下的可接受性。

⑤ ISO 21448 DIS 版第 12 章，Criteria for SOTIF release--SOTIF 发布准则

12 Criteria for SOTIF release

12.1 Objectives

The purpose of this clause is to achieve the following objectives:

a) The SOTIF release process shall be based on the SOTIF activities and the review of their work products for completeness and correctness; and
b) The achievement of the SOTIF shall be evaluated and a clear recommendation for approval or rejection of SOTIF release shall be given.

12.2 General

To achieve the objectives of this clause, the following information can be considered:

- Specification and design (according to 5.4.1);
- Hazards (according to 6.6.1)
- Hazardous events evaluation (according to6.6.2)
- Acceptance criteria (according to 6.6.3)

以安全接受准则作为自动驾驶预期功能安全(SOTIF)发布的评判依据之一。

中国典型提案 3：SOTIF 的量化开发

为提升 SOTIF 标准的可实施性，避免仅有流程和方法，难以实施落地的问题，中国代表团提出贯穿标准的量化开发提案。该提案反映在 ISO 21448 中“5-功能的定义和设计”、“6-危害的识别和评估”、“7-潜在功能不足和触发条件的识别和评估”、“9-验证和确认策略的定义”、“10/11-已知/未知危害场景的评估”、“12- SOTIF 发布准则”等关键章节中。举例如下：

① ISO 21448 DIS 版第 5.2 章，Performance targets—性能目标：

> The performance targets (e.g. detection range of target vehicles) of the installed sensors (e.g. radars, cameras), controllers, actuators or other inputs and components (e.g. maps – see Annex D.3) used by the intended functionality;
>
> NOTE Performance targets of an automated driving system, for example, include the detection and response to critical objects and events (e.g. pedestrians, vehicles, bicycles, motorcycles, and traffic signs) within the Operational Design Domain (ODD).
>
> EXAMPLE Performance of pedestrian detection on a highway (case where a pedestrian places a warning signal for the disabled vehicle)

自动驾驶设计过程中，应定义感知系统、控制系统、执行系统的性能目标（Performance targets），为后续开发和评测提供量化输入。

② ISO 21448 DIS 版第 6.4 章，Measurable parameters—危害行为的可测量参数指标：

> The severity and controllability of the hazardous event is considered to determine whether the resulting harm is acceptable in a given scenario. The severity and controllability evaluation considers the functional specification (according to the system specification and design described in Clause 5). Harm is considered acceptable if the controllability is as "controllable in general" or the severity is "no resulting harm". In all other cases a hazardous event is considered SOTIF related. The corresponding hazardous behaviour is described using measurable parameters like speed deviations and minimum distances to other objects. The controllability evaluation should include no reaction or a delayed reaction by the involved persons to control the hazard e.g. resulting from reasonably foreseeable indirect misuse.

SOTIF HARA 危害分析和风险评估过程中，应为识别出的自动驾驶危害行为定义可测量参数指标（Measurable parameters），以作为后续开发、验证、确认的量化依据。

③ ISO 21448 DIS 版第 7 章，Identification and evaluation of potential functional insufficiencies and triggering conditions--识别和评估潜在的功能不足和触发条件

7 Identification and evaluation of potential functional insufficiencies and triggering conditions

7.1 Objectives

The purpose of this clause is to achieve the following objectives:

a) Potential insufficiencies of specification, performance limitations and triggering conditions (including reasonably foreseeable direct misuse) shall be identified.
b) The response of the system to the identified triggering conditions that could initiate a hazardous behaviour shall be evaluated for its acceptability with respect to the SOTIF.

NOTE This activity considers the potential insufficiencies of specification of the intended functionality at the vehicle level as well as the potential insufficiencies of specification or potential performance limitations of E/E elements of the system.

7.2 General

To achieve the objectives of this clause, the following information can be considered:

— Documentation detailing the specification and design, in accordance with Clause 5.4.1;

— Hazards at the vehicle level, in accordance with Clause 6.6.1;

— Risk evaluation of hazardous behaviours, in accordance with Clause 6.6.2;

— Acceptance criteria, in accordance with Clause 6.6.3; and

— Known potential functional insufficiencies of the system and its components and known potential triggering conditions (including reasonably foreseeable direct misuse) that could lead to a hazardous behaviour based on external information or lessons learned (e.g. Clause 13.5.1).

以安全接受准则作为识别和评估自动驾驶预期功能不足及对应的触发条件的组合，是否能够引起自动驾驶危害行为的评判依据。

④ ISO 21448 DIS 版第 9.3 章节，Specification of integration and testing—集成和测试的定义

9.3 Specification of integration and testing

A verification and validation strategy is defined to provide argumentation that the objectives are achieved and how the validation targets are met. The verification and validation strategy can cover the whole intended functionality in the vehicle including both E/E elements and elements of other technologies considered relevant to the achievement of the SOTIF.

The validation targets are defined to provide evidence that the acceptance criteria are met. The validation targets can be determined in many ways depending on the chosen validation methods.

为评估自动驾驶预期功能在已知/未知危害场景下的风险水平，制定验证和确认(V&V)策略（包括确认目标）、方法应以是否满足安全接受准则为评判依据。

⑤ ISO 21448 DIS 版第 10.7 章节，Acceptability of residual risk due to known hazardous scenarios--已知危害场景下残余风险的可接受性

10.7 Acceptability of residual risk due to known hazardous scenarios

The validation targets defined in Clause 9 provide the strategy how the acceptance criteria are met. These validation targets are also relevant in case of residual risk for known hazardous scenarios.

为评估自动驾驶预期功能在已知危害场景下的残余风险的可接受性，应以验证目标是否满足安全接受准则为评判依据。

中国典型提案 4：基于场景优先度子集（Subsets）的自动驾驶测试方法

自动驾驶的安全评价需要基于目标市场场景，对于无事故里程数，如果场景差异较大，其展现的安全水平也不相同。目前自动驾驶实际道路测试耗时久、成本高、针对性不强，为了提升自动驾驶测试的时效性，更好地为量产开发服务，对已知场景进行分析和管理，如图 7 所示，建立关键场景因素子集，并将场景构成因素按照暴露频次、严重度、敏感性进行评级，并据此生成优先度顺序，进而建立优先场景库，在同等投入下，提升自动驾驶里程测试的时效性。

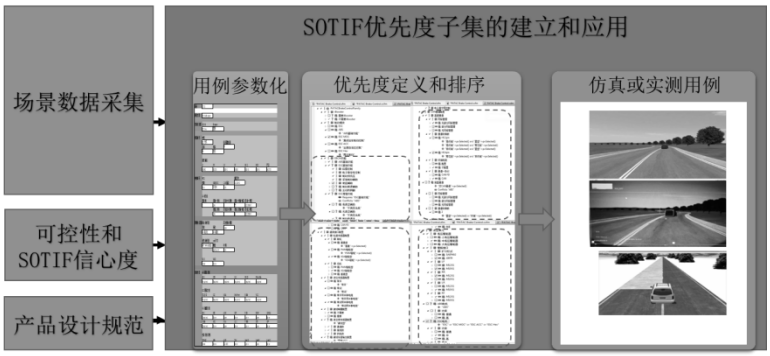

图 7 SOTIF 场景优先度子集的建立和应用

基于场景优先度子集开展自动驾驶测试，可大幅提升自动驾驶测试的效率，实现用更少的里程达到更大的场景覆盖效果。

基于优先度子集开展仿真测试，可基于关键因素衍生出更多的用例（含未知场景），以更快发现相关未知危害场景，在缩短累积里程测试的同时，避免竞相建立大量无效的场景库。

该提案反映在 ISO 21448 DIS 版的如下章节中：

① 7.3 Analysis of potential functional insufficiencies and triggering

conditions—潜在功能不足及触发条件的分析

NOTE 5 Proper abstraction (e.g. generation and use of equivalence classes or subsets) of all relevant use case parameters can be helpful to cope with large amounts of use case combinations.

在分析和发展潜在功能不足及触发条件时，优先度子集（Subsets）可以作为一种有效方法，以应对海量的场景用例组合。

② ISO 21448 第 9.3 章节，Specification of integration and testing--集成和测试的定义

9.3 Specification of integration and testing

A verification and validation strategy is defined to provide argumentation that the objectives are achieved and how the validation targets are met. The verification and validation strategy can cover the whole intended functionality in the vehicle including both E/E elements and elements of other technologies considered relevant to the achievement of the SOTIF.

The validation targets are defined to provide evidence that the acceptance criteria are met. The validation targets can be determined in many ways depending on the chosen validation methods.

NOTE 1 Acceptance criteria address the risk resulting from known and unknown hazardous scenarios. This is considered in the derivation of the validation targets.

NOTE 2 Annexes C.2 and C.6 give examples for defining and evaluating acceptance criteria and validation targets.

EXAMPLE 1 Consider a search for previously unknown triggering conditions that are relevant to the application. Validation targets would be defined with a statistical confidence that the empirical data supports the hypothesis that the triggering conditions that remain unknown do not impose unreasonable risk.

EXAMPLE 2 The validation target using simulation testing can be defined using pre-defined false positive and false negative rates for a function being tested.

If only a subset of scenarios is relevant for a specific hazard, then the exposure to the subset can be considered when determining the target values and the validation duration.

NOTE 3 Annex B.2, Table B.5 provides an example how to generate a subset of scenarios. When evaluating the likelihood that a triggering condition will violate the quantitative target, the exposure, controllability and severity of the resulting behaviour can be considered. This can result in a reduction in the effort required to demonstrate the exposure to the triggering condition.

基于特定危害与场景子集的相关性，可以针对性的定义并合理优化测试场景、目标及里程（时长），起到事半功倍的效果。

③ ISO 21448 附录 B，Guidance on Scenario and system analyses--场景和

Table B.5 — Factor subset example (e.g. considered for radar based function validation)

Category	Factor	Subset
Climate	Rainy	Subset 1
Road feature	Tunnel	
Time of day	any / don't care	
Objects off-roadway	Sign (too high position)	
...	...	
...	...	Subset n

针对特定的 SOTIF 开发工作，如开展对基于雷达的功能的确认，可考虑因素的功能敏感性、因素的出现频次、结果的严重程度，挑选建立因素子集（Subsets）用于后续确认工作。

中国典型提案 5：自动驾驶公共道路测试用车辆的安全

作为全球最具发展竞争力的汽车市场之一，中国已成为各个自动驾驶车辆测试的主战场，但对于处在开发过程中的自动驾驶汽车，由于其安全水平还未达到量产状态，安全风险很高。ISO 21448 提出应对未知风险领域的有效手段包括实际道路测试，为此，也需要应对由此带来的风险。中国专家组提出了自动驾驶公共道路测试用车辆安全的提案，虽然部分国家企业因自身企业责任可能提高而有所顾虑，但最终经过激烈讨论，获得全体会议通过。这也是整个 ISO 21448 标准中唯一对处于开发过程中车辆的安全要求。

该提案反映在如下章节中：

ISO 21448 第 11.3 章，Evaluation of residual risk due to unknown scenarios—未知场景残余风险的评估

NOTE 1 For tests in public areas additional safety measures might be necessary to prevent or mitigate the potential risk to the public due to test vehicles (e.g. emergency stop mechanism).

五、 基于“量化思想的中国提案”的落地研究

中国专家组围绕自动驾驶预期功能的设计开发、基于已知/未知危害场景的测试、验证/确认、最终发布等关键主题提出的多项提案，为完善自动驾驶预期功能安全(SOTIF)方法论提供了有力补充，进而推动 ISO 21448 进入 DIS 阶段。

作为主线贯穿 ISO 21448 的“量化思想的中国提案”，为 SOTIF 方法论在各国目标市场的落地应用提供了根本指引。

为了推动自动驾驶预期功能安全(SOTIF)方法论及“量化思想的中国提案”在中国目标市场得到切实有效的落地应用，建立以我国目标市场为主体的自动驾驶汽车预期功能安全设计开发、验证确认、测试评价体系，将自动驾驶系统设计不足、性能局限在已知/未知危害场景下导致的整车风险控制在合理可接受的范围内，避免竞相推高无止境的累积测试里程、降低开发成本、提高开发效率，从源头保障车辆运行安全，中汽中心标准所（全国汽车标准化技术委员会秘书处）联合行业企业陆续开展可操作、可应用、可实施的自动驾驶 SOTIF 研究工作。

本章节主要针对中国提案之一的“**量化思想的双层安全接受准则**”，在国内开展的工作做出具体介绍，可参考《汽车技术》中的《自动驾驶预期功能安全(SOTIF)接受准则的建立》文章，其他提案研究内容将在后续给出介绍。

1. 落实中国提案：量化思想的双层安全接受准则。

●第一层安全接受准则：自动驾驶危害行为事件接受准则。

●第二层安全接受准则：自动驾驶总体安全风险接受准则。

图 8 给出了自动驾驶预期功能安全双层安全接受准则的示例，自动驾驶过程中存在多次主动制动行为，但可能存在因设计不足或性能局限导致的制动危害，例如，感知系统误识别等原因造成的过大制动行为（事件）。以制动行为为例，导致违背第一层安全接受准则（如可控性和 SOTIF 信心度的安全度量指标）的过大制动危害行为（事件）将被记录下来，在完成全部的里程累积测试后（如 20 万 km），将违背第一层准则的危害行为（事件）数量与第二层安全接受准则进行比较，如果总数不超过一定数量（例如 2 次），则认为满足了总体的预期功能安全接受标准，以此作为自动驾驶预期功能安全的最终发布准则。

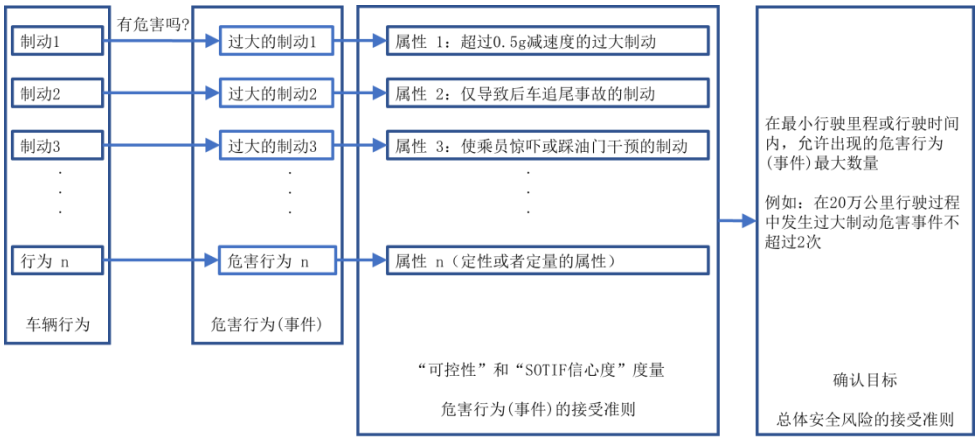

图 8 双层安全接受准则示例

2. 第一层安全接受准则：自动驾驶危害行为事件接受准则。

第一层安全接受准则，也就是危害行为事件接受准则，即针对自动

驾驶过程中危害行为事件的评价，可以包含定量准则和定性准则。其中，可控性指标和 SOTIF 信心度指标是定量准则的重要组成部分。

2.1 可控性准则及其评价

按照国家标准 GB/T 34590《道路车辆 功能安全》中给出的可控性定义，即为确定一个给定危害的可控性等级，需要预估具有代表性的驾驶员或其他涉及人员为避免伤害发生而能对场景施加影响的可能性。这种可能性预估包括：如果给定的危害将要发生，具有代表性的驾驶员能够保持或者重新控制车辆的可能性，或者在危害发生范围内的个体能够通过他们的行动避免危害的可能性。这种考量基于的假设为：危害场景中的个体为保持或者重新控制当前情况采取必要的控制行为，以及所涉及的驾驶员采取有代表性的驾驶行为。可控性预估可能受到很多因素的影响，包括该目标市场的驾驶员概况，如个体年龄、手眼配合、驾驶经验、文化背景等。因此，可控性表征驾驶员、乘员或其他涉险人员对车辆电控系统危害风险控制的难易程度，是衡量车辆行为是否构成危害的关键指标。

该研究针对可控性的衡量对象，即车辆行为危害，包括侧向、纵向、垂向运动相关危害，开展安全分析，结合测试结果及行业经验，对整车危害进行分类。通过分析车辆侧向、纵向、垂向运动功能特点，定义危害发生的典型场景，并组织大量代表中国目标市场的普通驾驶员开展实车危害行为的评估测试，定义出表征中国典型驾驶员对车辆侧向、纵向、垂向运动行为控制能力的客观度量指标。通过调整测试条件及被

测人员响应的及时性，兼顾传统汽车、新能源汽车和自动驾驶汽车相关控制系统，从整车侧向、纵向、垂向 3 个维度建立相关危害的可控性度量指标，如图 9 所示，为判断车辆行为是否构成危害提供了合理的量化准则，相关成果为自动驾驶系统的正向设计开发与测试评价，以及强制性国家标准和推荐性国家标准的落地实施提供了有效支撑，例如 GB 17675《汽车转向系 基本要求》、GB 21670《乘用车制动系统技术要求及试验方法》、GB/T《乘用车转向系统功能安全要求及试验方法》、GB/T《道路车辆 预期功能安全》等。

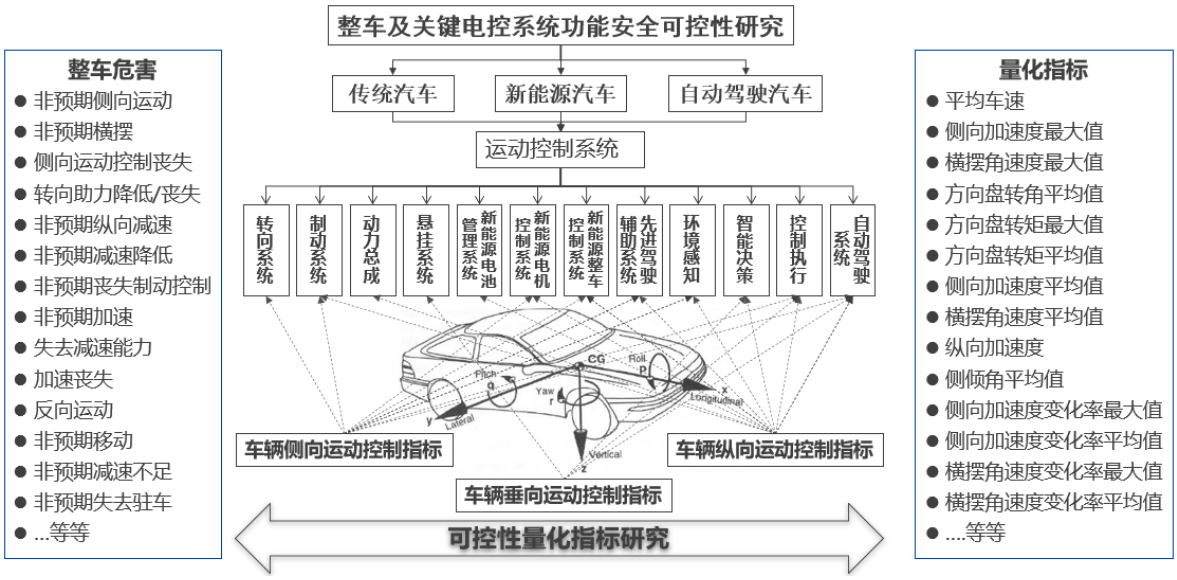

图 9 可控性安全度量指标体系

2.2 SOTIF 信心度准则及评价

可控性衡量的是车辆行为的安全边界，例如车辆制动减速度达到 0.5*g* 时，可能发生追尾事故。但对于自动驾驶汽车，如果发生了一次 0.3*g* 减速度的制动，就可能造成乘员的紧张甚至恐慌。如果乘员对自动驾驶汽车预期行为感到安全担忧，将导致功能开启率低和误干预等一系

列问题，这对自动驾驶的发展非常不利。为此，在现有安全和舒适评价维度的基础上，需要建立针对自动驾驶预期功能行为的“SOTIF 信心度”评价指标体系，如图 10 所示，在已有可控性安全边界的基础上，引入乘员对车辆行为的安全感受评价，以更加全面地评判自动驾驶安全性。

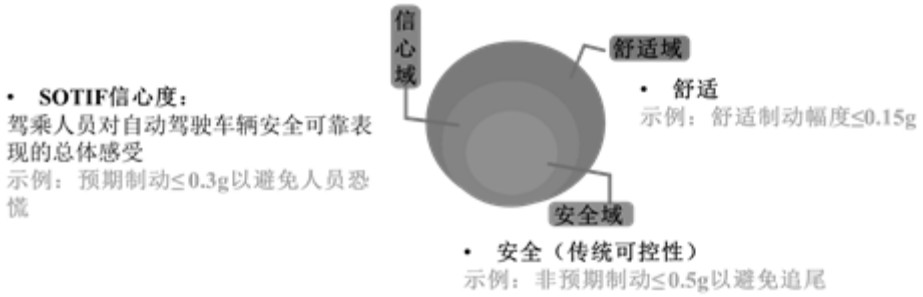

图 10 自动驾驶安全、舒适和“SOTIF 信心度”指标体系

SOTIF 信心度指标受人、车、环境多因素影响，通过分析其影响因素，挑选典型场景，开展实车主、客观对标测试，针对不同车辆行为的安全主观评价结果，开展数据学习，找出可以代表乘员信心度的客观指标值。

2.3 可控性准则、SOTIF 信心度准则试验研究

2016 年起，中汽中心联合泛亚、博世华域、捷太格特、海拉等行业企业，先后开展 6 轮研究测试，确定了可控性准则和 SOTIF 信心度准则的试验方案，为当前开展的大规模试验研究奠定了坚实基础。

表 1 为典型的整车侧向行为危害、安全目标、安全度量、验证确认方法及部分结果举例：

表 1 整车侧向行为危害、安全目标、安全度量、验证确认方法

序号	整车危害	安全目标	安全度量		验证确认方法（V&V）
			可控性准则	SOTIF 信心度准则	
1	非预期的侧向运动	车辆非预期的侧向运动应满足非预期侧向运动的安全度量	侧向加速度变化小于XXX；转向盘手力矩小于 XXX	TBD	TBD
2	非预期地失去侧向运动控制	应确保驾驶员对车辆侧向运动的控制能力，相应转向操纵力应满足非预期失去转向控制的安全度量	转向盘手力矩小于 XXX	TBD	TBD
3	失去助力情况下的转向沉重	转向操纵力应满足转向沉重的安全度量	转向盘手力矩小于 XXX	TBD	TBD
注：安全度量应基于目标市场来确定。					

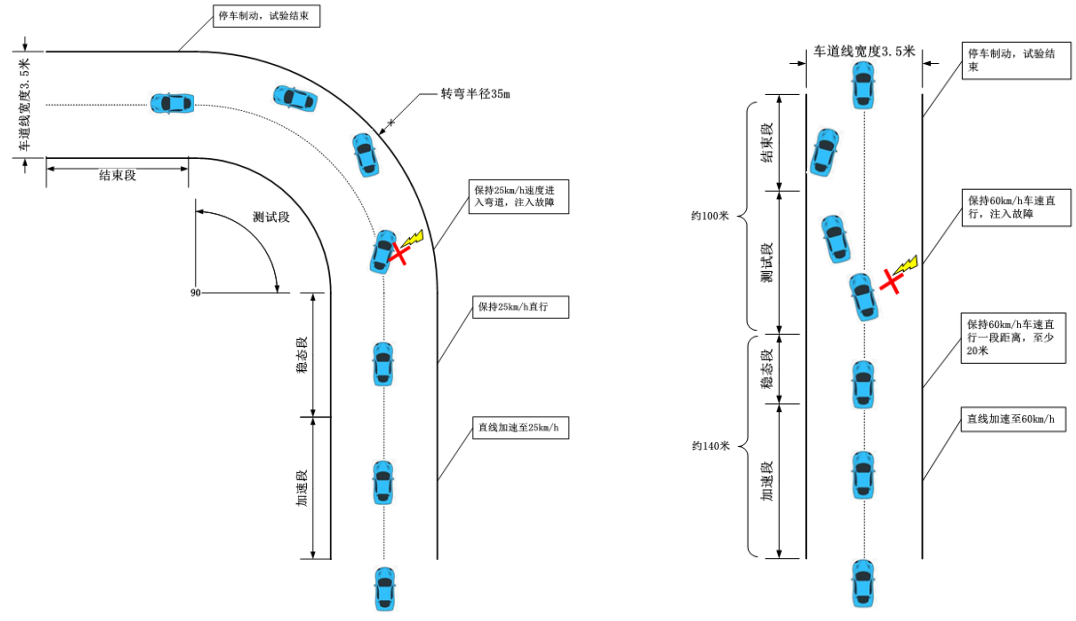

图 11 真实道路场景搭建及测试

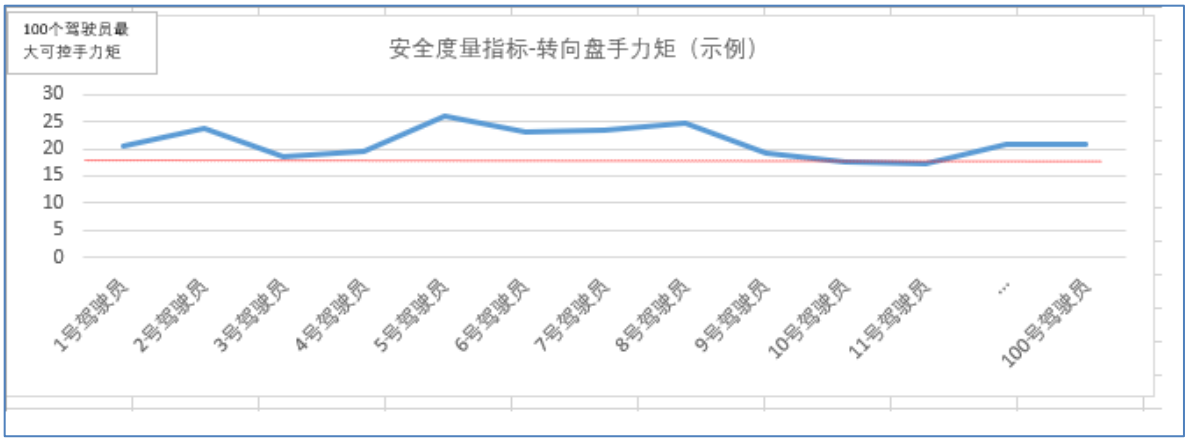

图 12 安全度量指标示例

表 2 为典型的整车纵向行为危害、安全目标、安全度量、验证确认方法，目前正在研究中：

表 2 整车纵向行为危害、安全目标、安全度量、验证确认方法

序号	整车危害	安全目标	安全度量		验证确认方法（V&V）
			可控性准则	SOTIF 信心度准则	
1	非预期的减速	TBD	TBD	TBD	TBD
2	非预期的丢失减速能力	TBD	TBD	TBD	TBD
3	非预期的纵向运动	TBD	TBD	TBD	TBD
4	非预期的制动踏板下沉影响人员对车辆的操作	TBD	TBD	TBD	TBD
注：安全度量应基于目标市场来确定。					

2019 年起通过社会公开招募的方式，选取中国目标市场具有代表性的驾驶员（100 个样本量以上），在典型驾驶场景下，当系统发生故障、预期功能不足、性能存在局限情况下，确定 2 项整车最高层面的安全接受准则及其量化指标：可控性和信心度。该安全接受准则对于 L3 及以上高级别自动驾驶功能如车辆接管、误用等具有重要指导意义。

所选取的样本量将覆盖图 13 中国目标市场的 7 个主要区域：华北、华东、华中、华南、西南、东北、西北。

针对车辆侧向行为危害，已完成天津、上海、盐城 3 站的试验研究。

车辆纵向行为危害的研究方案正在讨论中。

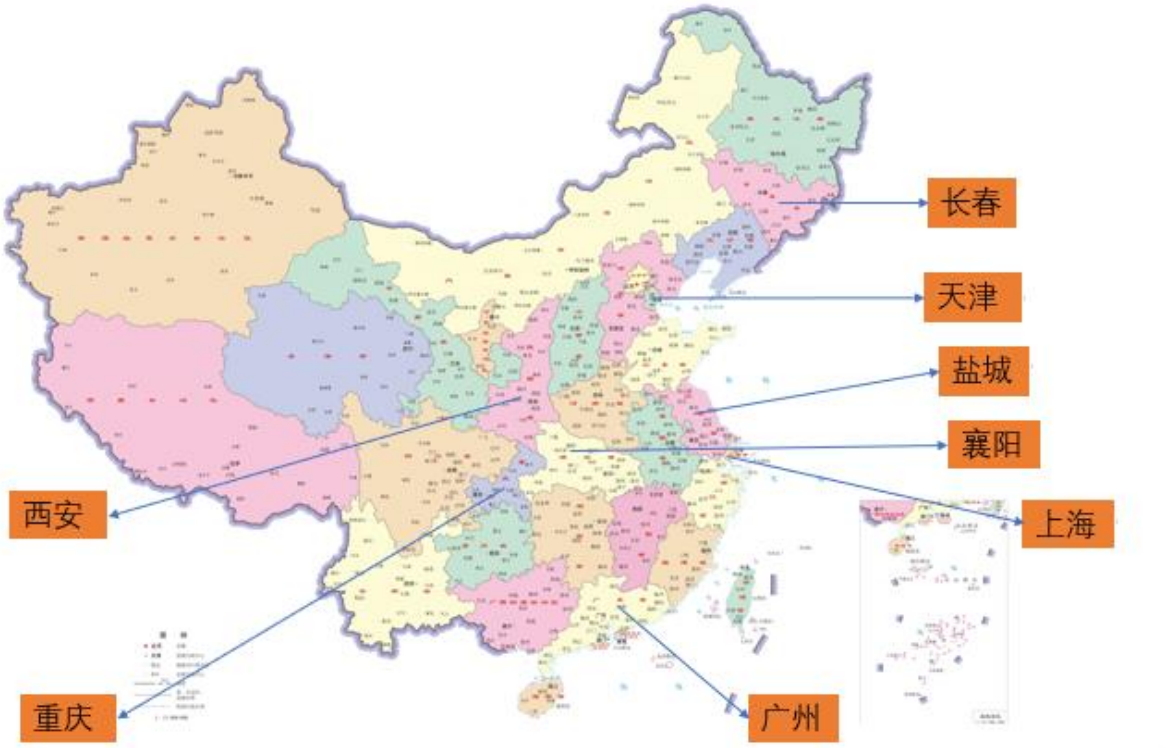

图 13 功能安全和预期功能安全（SOTIF）接受准则（Acceptance Criteria）的样本量覆盖区域

3. 第二层安全接受准则：自动驾驶总体安全风险接受准则。

第二层安全接受准则，也就是总体安全风险接受准则，针对自动驾驶汽车在真实道路累积全部行驶里程过程中的总体安全风险评价。同人类驾驶员一样，面对各种场景，自动驾驶系统也无法做到杜绝危害行为事件的发生。如果对比没有自动驾驶功能的人类驾驶安全表现，引入自动驾驶功能后，安全风险没有提高，则认为自动驾驶汽车的安全表现是可以被接受的。

3.1 自动驾驶危害行为事件的泊松分布规律

泊松分布适合描述单位时间（或空间）内随机事件发生的次数。根据 ISO 21448，真实场景中自动驾驶功能导致的危害行为事件数量也可以用泊松分布规律来描述。泊松分布的概率函数为：

$$P(\mathrm{X}=\mathrm{k})=\frac{\lambda^{\mathrm{k}}}{\mathrm{k!}}\mathrm{e}^{-\lambda},\mathrm{k}=0,1,2,\ldots \quad (1)$$

式中，λ 为单位里程（或单位时间）内危害行为事件的平均发生次数；k 为危害行为事件发生次数。

通过转化，可得危害行为事件发生的平均里程或时间间隔（即无事故里程或时长）为：

$$\tau=-\ln(1-\alpha)/\lambda \quad (2)$$

式中，α 为置信度水平。

例如，当无危害行为事件里程数达到 100 万 km 时，具有 99%置信度水平认为该系统在同等驾驶场景中危害事故率能达到 4.6×10^{-6} 次/km。

3.2 自动驾驶总体安全风险接受准则

对自动驾驶系统总体安全水平的评估，应考虑其是否带来了不合理的安全风险，即与同等驾驶场景下人类驾驶员的安全驾驶能力指标（如平均无事故里程）相比，引入自动驾驶系统后，相关指标不应变差。因此，可以认为如果自动驾驶系统没有带来明显的不合理风险，则其总体安全风险是可以接受的。

总体安全风险接受准则的定义和确认需要基于目标市场情况，假设

驾驶员安全水平较高的乘用车驾驶员平均每年行驶 2 万 km，每 10 年发生 1 次交通事故。以此作为目标，选择 95%置信度，则$\tau \approx 60 \times 10^4$ km。即为了证明在 95%置信度下认为自动驾驶车辆事故率能达到上述驾驶员的驾驶安全水平，需要累积测试 60 万 km 无危害事故。通过基于目标市场的统计研究，可以得到危害行为事件的平均行驶里程，再考虑合理的设计余量，作为自动驾驶总体安全风险的接受准则。

3.3 自动驾驶里程累积测试终止原则

在自动驾驶里程累积测试过程中，通常会伴随危害行为事件的出现，特别是随着新功能、新设计的实施，发生危害行为事件的平均里程数会出现先下降后逐步上升的情况，如图 14 所示，从统计规律定性描述了引入新功能后，由于该新功能应对各种场景的能力较低，因此安全行驶里程相对较短，但通过预期功能安全的迭代开发、功能改进，危害事件发生率下降，从而无危害事件发生的安全平均行驶里程增加，也就是通过预期功能安全的迭代开发，车辆发生危害事件的次数降低，安全行驶里程增加，安全能力得到提升。

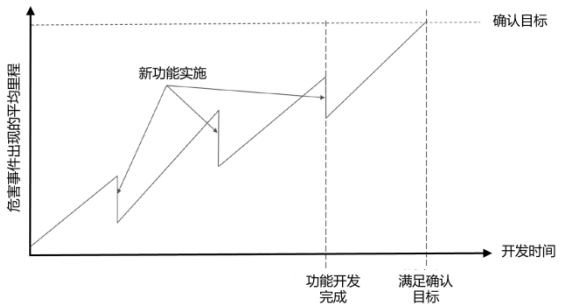

图 14 自动驾驶危害行为事件出现的平均里程

如果在达到累积测试里程目标前出现了危害行为事件，修复后为了继续确认自动驾驶系统是否可以满足初始设定的安全目标（相同危害行

为事件发生率和置信度水平），后续测试里程数会比无危害行为事件发生时更长。假设在里程达到τ_1时发生危害行为事件，当修复后，验证达到同等置信度水平α的危害行为事件发生率目标λ_0所需要的总里程 s 可由式（3）确定：

$$\int_0^{\lambda_0} P(\lambda|1,\tau_1+s)d\lambda = 1-\alpha \quad （3）$$

例如，定义危害行为事件发生率为λ_0=0.001 次/km，置信度水平α=99%，则发生 j 次危害行为事件后，需要测试的总里程 s 如表 3 所示。

表 3 自动驾驶危害行为事件发生次数与测试里程

危害行为事件发生次数j（次）	测试总里程s(km)
0	4 605.17
1	6 638.35
2	8 405.95
3	10 045.12
4	11 604.63

自动驾驶整车行为危害的安全接受准则（第一层安全接受准则）的得出，应基于具体可行的试验方案，而试验方案中场景的选取和搭建、试验车辆条件、试验条件如车速/车距等条件的建立，关键在于基于中国目标市场的交通场景数据统计分析，而自动驾驶总体安全风险接受准则（第二层安全接受准则）的得出关键在于交通事故数据统计分析。

六、 结束语

我国高度重视汽车强国战略，安全是命脉根基。国家层面陆续出台了《交通强国建设纲要》、《智能汽车创新发展战略》、《新能源汽车产业发展规划（2021-2035 年）》等一系列规划举措，重点强调健全车辆运行安全保障体系，开展整车、零部件安全技术研究、加快功能安全等新型安全标准研制，完善企业负责、政府监管、行业自律和社会监督相结合的安全生产机制，落实安全生产责任，加强安全生产监督管理。

功能安全(Functional Safety)和预期功能安全(SOTIF)技术和方法论为确保新能源汽车、自动驾驶汽车、传统汽车在故障、非故障情况下的安全运行提供了根本保障，已成为国际共识。ISO 21448 的多项中国提案为完善自动驾驶预期功能安全(SOTIF)方法论提供了有力补充，而推动 SOTIF 方法论在我国目标市场(Target market)的落地研究和应用是当务之急。

接下来，将继续落实第四章中提到的各项中国提案，包括量化思想的 SOTIF 双层安全接受准则、SOTIF 的量化开发、建立基于场景优先度子集的优先场景库及自动驾驶测试方法。同时，开展 SOTIF 几个特殊方面的研究：驾驶策略（Driving Policy）作为决策环节，如何设计驾驶策略以支持 SOTIF 设计、验证和确认（V&V）；应用机器学习（Machine Learning）功能时，如何应对 SOTIF 风险；高精地图作为安全相关的功能，SOTIF 对于高精地图局限性的考量；用于提高道路安全和运行效率的 V2X 功能，作为复杂的感知系统，SOTIF 对于 V2X 局限性的考量。

基于以上，建立以我国目标市场为主体的自动驾驶汽车预期功能安

全(SOTIF)设计开发、验证确认、测试评价、发布的闭环技术和标准体系，从整车、系统到感知、决策、执行各个环节，将自动驾驶系统设计不足、性能局限在已知/未知危害场景下导致的整车风险控制在合理可接受的范围内，避免竞相推高无止境的累积测试里程及大量无效的场景库、降低开发成本、提高开发效率，从源头保障车辆运行安全。

参考文献

[1] ISO 21448 Road vehicles — Safety of the Intended Functionality (DIS).

[2] LITTLEWOOD B, WRIGHT D. Some Conservative Stopping Rules for the Operational Testing of Safety-Critical Software[J]. IEEE Transactions on Software Engineering, 1997, 23(11), 673-683.

[3] 自动驾驶预期功能安全(SOTIF)接受准则的建立. 汽车技术, 2020(12).

网址：www.catarc.org.cn

地址：天津市东丽区先锋东路 68 号

邮编：300300

联系方式：

全国汽车标准化技术委员会（SAC/TC114）　　姓名：付越

道路车辆功能安全标准研究制定工作组　　电话：022-84379288

ISO/TC22/SC32/WG8 中国专家组　　邮箱：fuyue@catarc.ac.cn