·智能网联汽车安全验证与评估专题·

自动驾驶运行模式转换的安全论证方法研究*

陈俊涛[1] 陈贞[2] 杨强[3] 武湘成[4]

（1.北京航迹科技有限公司，北京 100080；2.北京镝石数据科技有限公司，北京 100176；3.北京赛目科技股份有限公司，北京 100080；4.工业和信息化部装备工业发展中心，北京 100846）

【欢迎引用】陈俊涛，陈贞，杨强，等．自动驾驶运行模式转换的安全论证方法研究[J]．汽车文摘，2026(8)：16-26.

【Cite this paper】CHEN J T, CHEN Z, YANG Q, et al. Research on Safety Argumentation Method for Automated Driving Operational Mode Transitions[J]. Automotive Digest (Chinese), 2026(8): 16-26.

【摘要】针对多运行模式并存条件下，自动驾驶系统运行能力边界变化、驾驶责任转移及风险缓解策略触发等安全问题，提出了一套面向运行模式转换的安全论证方法。该方法通过构建“运行模式-运行状态-行为响应”三层有限状态机模型，对多运行模式及其转换逻辑进行形式化建模，并进一步提出“状态机路径-场景类型-安全证据链”的三维验证模型，将法规安全要求与工程设计、测试验证及安全证据进行系统关联。通过典型场景示例，系统展示从安全要求分解、模型构建、验证实施到安全论证形成的完整流程，为多运行模式自动驾驶系统的安全设计、监管审查与合规决策提供工程化参考。

关键词：自动驾驶；符合性评估；运行模式转换；安全验证；安全论证

中图分类号：U461.91　　**文献标志码：**A　　**DOI:** 10.19822/j.cnki.1671-6329.20260007

Research on Safety Argumentation Method for Automated Driving Operational Mode Transitions

Chen Juntao[1], Chen Zhen[2], Yang Qiang[3], Wu Xiangcheng[4]

(1. Beijing Voyager Technology Co., Ltd., Beijing 100080; 2. Beijing Dishi Data Technology Co., Ltd., Beijing 100176; 3. Beijing Saimo Technology Co., Ltd., Beijing 100080; 4. Ministry of Industry and Information Technology Equipment Industry Development Center, Beijing 100846)

【Abstract】To address safety challenges arising from dynamic variations in operational capability boundaries, shifts in driving responsibility, and the activation of risk mitigation strategies in automated driving systems under multiple operational modes, this paper proposes a safety argumentation method for operational mode transitions. The method constructs a three-layer finite state machine model of “operational mode-operational state-behavioral response” to formally describe multi-mode operation and transition logic. Furthermore, a three-dimensional verification model of “state machine path, scenario type, safety evidence chain” is proposed to systematically link regulatory safety requirements with engineering design, testing and validation activities, and safety evidence. Typical scenarios are used to demonstrate the complete process from safety requirement decomposition and model construction to validation implementation and safety argumentation, providing an engineering reference for the safety design, regulatory review, and compliance decision-making of multi-mode automated driving systems.

Key words: Automated driving, Compliance assessment, Operational mode transition, Safety validation, Safety argumentation

0 引言

近年来，自动驾驶技术逐步由研发验证阶段迈向受控场景下的有限规模应用，车辆在多运行模式间的切换问题成为安全设计与监管关注的前瞻性议题[1-2]。尽管高级别自动驾驶功能尚未实现规模化运行，但在

*基金项目：新一代人工智能国家科技重大专项(2022ZD0116311)。

通信作者：陈贞(1984—)博士，高级工程师，研究方向为智能网联汽车安全测试与评估方法，chenzhen@dystech.cn。

当前组合驾驶辅助系统的实际运行中，功能退出、人工干预及直接控制等环节模式转换阶段的风险越来越凸显。这类现实运行经验为未来自动驾驶系统在多模式下运行模式及其转换机制方面的安全设计提供了重要实践依据。

在监管层面，自动驾驶相关法规与管理体系正经历范式转变：由早期以单一功能性能测试为主的合规模式，逐步演进为以系统安全可论证性为导向的监管范式[3-4]。以联合国自动驾驶全球技术法规（GTR ADS）及我国智能网联汽车准入与上路通行试点为代表，监管机构通过明确自动驾驶功能定义、设计运行范围（Operational Design Domain，ODD）、风险缓解状态（Mitigated Risk Condition，MRC）及控制权交接机制等关键要素，要求制造商系统性证明其产品在不同运行状态及其潜在转换过程中的安全可控性，并形成可被审查、复核和采信的安全论证证据。相关监管框架的共同特征在于秉持技术中立原则，即不限定具体技术实现路径，而是将合规目标转化为系统工程层面的可验证与可论证约束[1,3]。

在此背景下，自动驾驶系统的多运行模式转换安全性，已成为贯通实际运行数据与前瞻监管要求的核心工程问题。GTR ADS将自动驾驶功能划分为依赖后援用户（ADS Feature of type 1，ADSF-1）与不依赖后援用户（ADS Feature of type 2，ADSF-2），并允许2类功能在统一系统架构下共存与切换[5]。由于不同运行模式在系统能力边界、责任分配及风险处置策略上存在本质差异，其转换过程中的一致性、可解释性与可验证性，将直接影响系统安全论证是否成立。

然而，现有研究与工程实践多聚焦于单一运行模式下的功能验证或场景适配[6-9]，针对多运行模式并存条件下运行模式转换的系统建模方法、验证证据组织方式以及面向监管审查的安全论证结构缺乏系统性研究，难以充分满足前瞻性监管对安全论证可审查性与可复核性的要求。

针对上述问题，本文提出一种基于运行模式状态机的运行模式转换安全论证方法，将法规安全要求内嵌于运行模式定义、状态转换条件及路径中，并通过结构化验证与证据组织机制，实现对尚未规模化运行场景下运行模式转换安全性的前置论证，从而为自动驾驶系统的安全设计与合规评估提供工程化支撑。自动驾驶系统完整的运行模式转换安全验证与论证体系如图1所示。

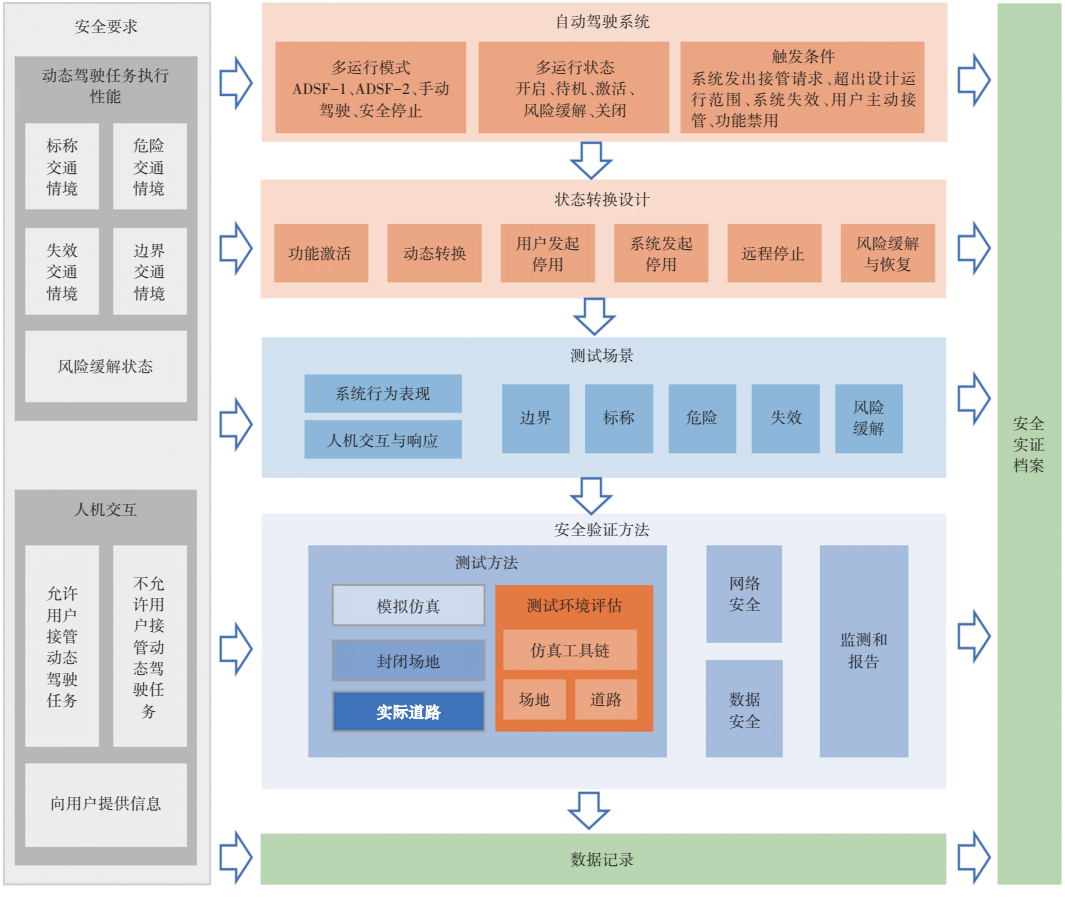


图1　运行模式转换安全验证与论证体系概览

1 多模式协同安全要求与设计挑战

在GTR ADS等法规约束下，运行模式及其状态转换成为系统工程中需明确设计与验证的对象。本章从工程视角出发，一方面梳理法规对多运行模式协同及状态转换提出的核心安全要求，另一方面分析在ADSF-1与ADSF-2共存条件下运行模式转换所面临的关键设计挑战，为后续验证方法与安全论证结构奠定基础。

1.1 核心安全要求

在GTR ADS及国内实施指南中，运行模式及状态转换是连接系统功能、安全责任及风险控制的关键要素，法规针对制造商的车辆ADSF模式转换的核心要求包括[1]：

（1）责任明确性：法规要求在不同运行模式下，系统与用户（后援用户）在动态驾驶任务（Dynamic Driving Task，DDT）中的责任边界应清晰、可判定。

（2）安全状态可达性：当系统无法维持当前运行模式时，应能够自动执行风险缓解策略，使车辆进入风险缓解状态，或通过清晰、及时的提示将控制权安全移交后援用户。

（3）模式转换可控性：任何运行模式的激活、停用与切换，均应在系统控制下有序进行，避免非预期的系统行为或用户不可感知的模式变化，并要求用户（或后援用户）能够清晰地感知系统运行状态及其变化过程。

（4）可验证与可论证性：系统在不同运行模式及其转换过程中的安全性能，需通过测试验证，并形成可与具体运行模式转换路径一一对应的安全实证档案，以供监管审查与采信。

1.2 多运行模式协同设计挑战

将ADSF-1与ADSF-2集成于同一自动驾驶系统，并支持两类模式间的动态转换，涉及系统架构、安全策略与人机交互的深层重构，带来以下工程挑战：

（1）控制逻辑融合：ADSF-1与ADSF-2在功能特性与安全概念上存在差异，两类功能详细对比内容如表1所示。前者依赖后援用户作为回退策略，后者则要求系统具备完全自主的风险处置能力。在同一控制框架内融合两类模式，并确保转换逻辑一致、行为可预测，是系统设计的首要难题。

（2）责任与人机交互（Human-Machine Interaction，HMI）边界动态管理：运行模式转换瞬间，系统需完成控制权分配、执行责任归属、功能接口、控制接口与HMI接口的同步一致性管理，并对转换过程中的中间状态进行精细化定义，以满足法规对控制权唯一性、责任可判定性、用户状态可感知性的约束要求。

（3）人机交互与接管安全：在涉及后援用户的模式转换场景中，系统需通过多模态人机交互手段，确保提示信息的可感知性、可理解性以及用户接管行为的及时性和平顺性。

（4）系统化验证与证据生成：运行模式转换涉及多种触发条件、状态路径与异常组合，需系统性地覆盖所有关键转换路径与边界工况，并将验证结果组织为可审查和合规采信的安全证据。

表1 ADSF-1与ADSF-2功能特性对比[1,10]

维度	ADSF-1	ADSF-2
后援用户依赖	需要后援用户接管，用户须具备接管能力	无需后援用户接管，用户可选择是否接管
技术关注	用户交互机制、接管能力监测、接管策略、接管请求（Take-Over Request，TOR）响应	风险缓解策略、ODD边界判断、MRC执行与恢复机制
回退策略	TOR触发，后援用户接管；如后援用户在限定时间内未接管则进入MRC	系统自动进入MRC
典型风险	用户未及时响应TOR、误触发接管或系统误判	系统感知或决策失效、ODD判断错误

2 基于FSM的多模式安全转换建模方法

2.1 设计思路

基于有限状态机（Finite State Machine，FSM）构建多运行模式安全转换框架，其核心是以运行模式及其状态转换作为工程对象。通过在状态机建模与转换路径定义中，将法规安全要求转化为可执行的系统行为约束，实现“法规要求-工程设计-验证活动-安全证据”的闭环映射[11-12]。

2.2 基于有限状态机的多层形式化模型

为了分析自动驾驶系统在多运行模式并存条件下的状态演化与安全转换行为，本文构建了一种3层递进式有限状态机模型，分别为运行模式层、运行状态层与行为响应层。该模型实现了从宏观控制责任与安全边界，到系统内部执行状态，再到具体控制与人机交互行为的逐级形式化建模。系统运行模式、运行状态及其转换路径之间的整体关系如图2所示。

2.2.1 运行模式层

运行模式层描述系统顶层功能配置及车辆控制权归属，是法规安全要求与系统设计之间的直接映射层。典型运行模式集合$\boldsymbol{M}$定义为：

$$\boldsymbol{M}=\{\text{ADSF-1, ADSF-2, Manual, SafeStop}\} \quad (1)$$

式中：$\boldsymbol{M}$ 为向量，代表系统全部顶层运行模式集合；ADSF-1 为依赖后援用户自动驾驶功能；ADSF-2 为无后援用户自动驾驶功能；Manual 表示人工驾驶模式；SafeStop 表示系统进入风险缓解状态或安全停车后的终态[1]。系统实现中可引入必要的内部子模式或工程状态。

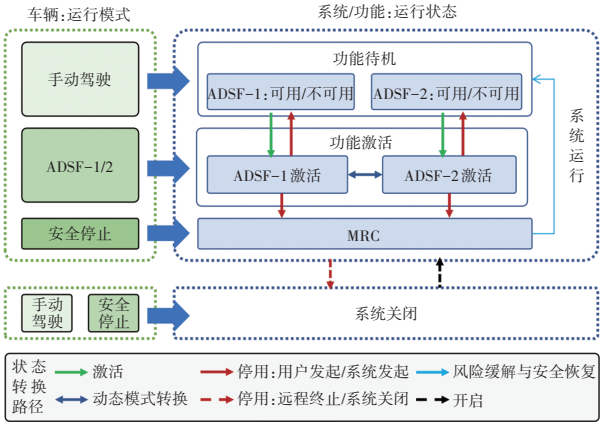


图 2　运行模式–运行状态–转换路径关系示意

2.2.2　运行状态层

运行状态层用于刻画各运行模式下系统内部的精细执行阶段，系统状态覆盖功能待机、激活及风险缓解等阶段，其中待机状态区分为可用与不可用，激活状态区分为 ADSF-1 与 ADSF-2。运行状态集合 $\boldsymbol{S}$ 定义为：

$$\boldsymbol{S}=\{(\text{Standby, available}),(\text{Standby, unavailable})\}\cup\{(\text{Active, ADSF}-1),(\text{Active, ADSF}-2)\}\cup\{\text{MRC}\} \quad (2)$$

式中：$\boldsymbol{S}$ 为系统内部的全部运行状态集合，Standby 为功能待机状态，available 为功能可用，unavailable 为功能不可用，Active 为功能激活状态，MRC 为风险缓解状态。

运行状态是运行模式在系统内部的工程展开状态。该约束用于保证运行状态的定义不脱离其所属运行模式，从而避免在验证与安全论证过程中出现模式和状态语义不一致的问题。

2.2.3　行为响应层

行为响应层定义系统在特定运行状态下的具体执行动作与 HMI 策略，包括功能激活提示、控制权交接反馈、风险缓解阶段的控制策略表现等。该层用于确保系统控制行为、HMI 提示与运行状态之间的一致性，是状态机模型向测试验证、合规审查及运行监测落地的关键接口。

2.3　系统架构与核心转换路径

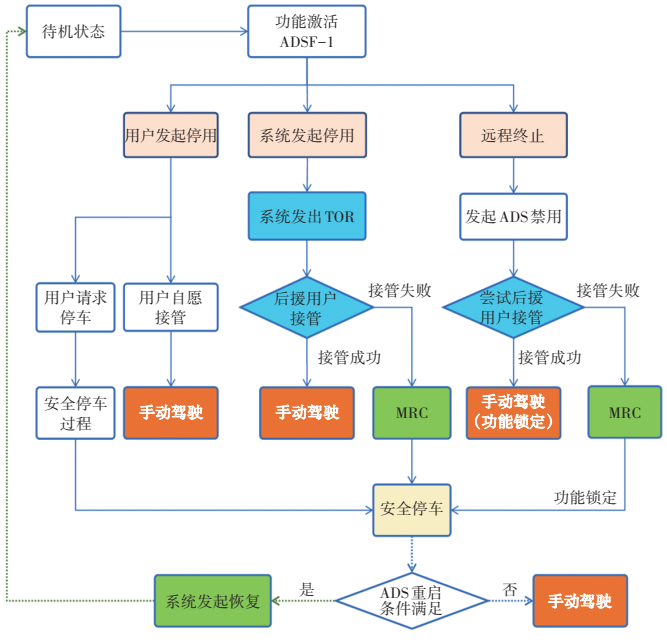
2.3.1　ADSF 系统架构类型

为明确不同自动驾驶系统架构对运行模式转换设计与验证复杂度的影响，本文根据支持的自动驾驶功能类型，将系统划分为 3 类典型架构：

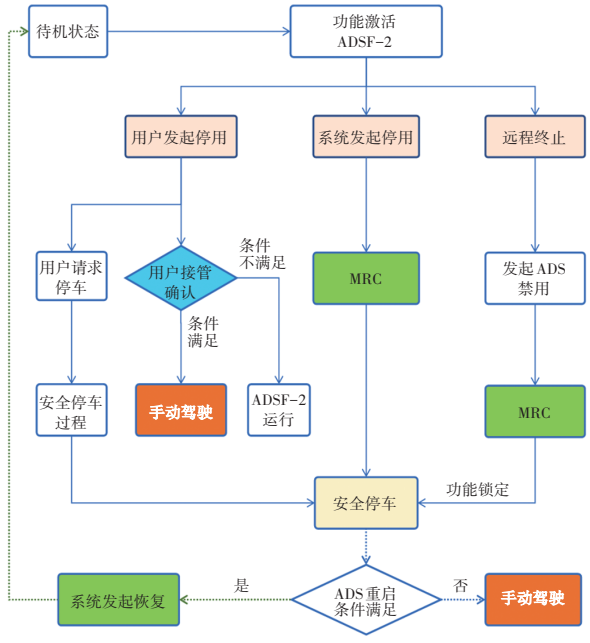
（1）ADSF-1 单模式系统：仅支持 ADSF-1，设计重点在于后援用户监测与接管安全性（图 3a）。

（2）ADSF-2 单模式系统：系统在 ODD 内独立运行，异常或边界条件下直接进入风险缓解状态（图 3b）。

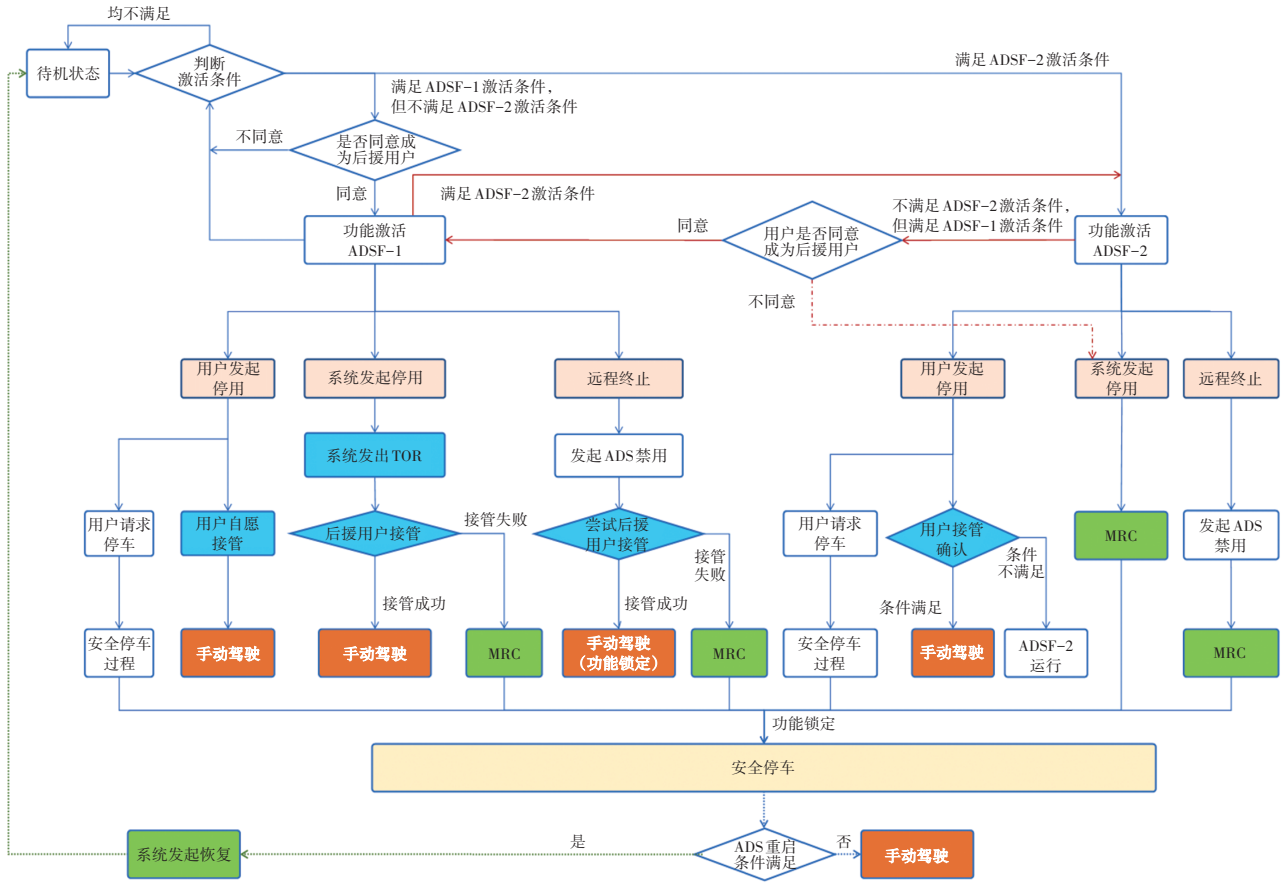
（3）ADSF-1/2 混合模式系统：支持两类模式共存与动态切换，是 GTR ADS 重点关注，且对安全论证提出更高要求的工程形态（图 3c）。



（a）仅包含 ADSF-1 的自动驾驶系统



（b）仅包含 ADSF-2 的自动驾驶系统



（c）ADSF-1及ADSF-2共存的自动驾驶系统

图3　状态转换统一逻辑框架

2.3.2　核心转换路径

结合法规要求与典型工程实践[1]，本文将自动驾驶系统运行模式相关的状态转换归纳为6类核心路径：功能激活、模式转换、用户发起停用、系统发起停用、远程终止以及风险缓解与安全恢复。上述路径覆盖系统全生命周期内的主要安全相关状态变化，构成后续验证与安全论证的基础转换路径集合。图4为6类路径构成GTR ADS审查中运行模式转换相关安全风险的主要覆盖对象。

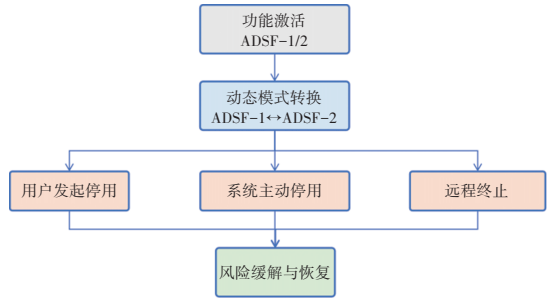


图4　6类转换路径

2.4　关键设计参数

2.4.1　输入变量空间

本文所定义的输入变量空间用于运行模式与状态转换条件的形式化表达与验证分析，不直接对应具体感知或控制算法的实现接口。为保证状态转换条件的确定性与可验证性，本文将触发条件形式化为输入向量$\boldsymbol{X}(t)$：

$$\boldsymbol{X}(t)=[E(t), D(t), H(t), R_c(t), U(t), R(t)] \quad (3)$$

式中：$\boldsymbol{X}(t)$为t时刻系统全部输入判据变量集合；$E(t)$为t时刻环境适配状态；$D(t)$为t时刻后援用户接管能力状态；$H(t)$为t时刻系统硬件与功能健康状态；$R_c(t)$为t时刻远程管控指令信号；$U(t)$为t时刻用户操作输入信号；$R(t)$为t时刻系统实时风险评估值。

各变量的定义、数据来源及阈值约束如表2所示。通过对输入变量的统一评估，可为运行模式与状态转换条件提供结构化、可验证的判据基础。

2.4.2　模式与状态转换函数

运行模式层的宏观转换逻辑函数定义为：

$$mode(t+1)=T(mode(t), \boldsymbol{X}(t)) \quad (4)$$

式中：$mode(t)$为t时刻系统当前运行模式，$mode(t+1)$为$t+1$时刻迁移后的目标运行模式，$T(\ )$为运行模式顶层状态迁移函数。

运行状态层的微观迁移函数定义为：

$$state(t+1)=f(state(t), \boldsymbol{X}(t)) \quad (5)$$

式中：$state(t)$为t时刻系统内部运行状态，$state(t+1)$为$t+1$时刻迁移后的内部运行状态，$f(\ )$为系统内部运行状态迁移函数。

两者通过映射关系$g(state(t))=mode(t)$保证模式与状态之间一致性。该形式化定义支持对关键状态空间的可达性、路径覆盖性及逻辑一致性的系统性分析。

表2　关键输入变量及其判据属性

输入变量	含义	判据性质	阈值说明
$E(t)$	环境适用性	ODD合规性	$E\geq E_{req}$表示满足功能运行条件
$D(t)$	后援用户状态	可接管能力	$D\geq D_{min}$表示用户具备接管能力
$H(t)$	系统健康状态	功能完整	$H=H_{ok}$表示系统处于可运行状态
$R_c(t)$	远程控制/管控信号	外部强制约束	$R_c(t)=1$时触发规定模式响应
$U(t)$	用户输入信号	人机交互有效性	$U(t)=1$表示有效用户请求
$R(t)$	风险评估值	安全触发判据	$R\geq R_{min}$触发安全转换

2.4.3　风险评估模型

为支持运行模式与状态转换条件的统一触发判据定义，并便于开展验证与安全论证阶段的路径触发一致性分析，研究中引入以验证与安全论证为目的的风险评估函数$R(t)$：

$$R(t)=\sum_i P_i(S(t+1)|S(t))\cdot w_i\cdot C_i \quad (6)$$

式中：$R(t)$为t时刻系统综合风险评估值，i为第i条状态转换路径，$P_i(\ |\ )$为第i条转换路径发生的场景条件概率，$S(t)$为t时刻系统状态，$S(t+1)$为$t+1$时刻目标状态，w_i为第i条转换路径风险权重系数，C_i为第i条转换失效对应的后果严重度。当$R(t)$超过阈值R_{min}时，系统触发相应的安全转换机制。

本文并不试图构建通用风险评估模型，而是将该风险评估判据作为运行模式转换验证与安全论证中的统一触发基准，用于支撑路径覆盖分析与证据组织，并不替代车辆在实际运行中的实时感知、决策或控制算法。在此框架基础上，可进一步引入不确定性建模方法（如部分可观测马尔可夫决策过程[13]），用于分析复杂环境下的路径覆盖与风险演化问题。

3　多维验证闭环方法

3.1　三维验证模型

在GTR ADS框架下，运行模式转换验证的核心目标：自动驾驶系统在其已定义的运行模式及其转换过程中，不会引入不合理的安全风险，且相关安全结论具备可验证性与可追溯性[1]。为系统性覆盖运行模式转换中的安全风险，本文提出“状态机路径-场景类型-安全证据链”三维验证闭环模型。该模型以有限状态机定义的运行模式转换路径为基本验证单元，在多类风险场景下开展验证活动，并将验证结果组织为与安全声明直接对应的证据链。

（1）状态机路径维度：基于有限状态机对运行模式及其转换路径进行形式化定义，明确不同触发条件下的状态迁移关系。验证重点包括：状态迁移逻辑的一致性与无歧义性；控制权转移的唯一性与可解释性；在任意路径下安全状态的可达性。

（2）场景维度：针对每条状态机路径，构建覆盖标称、边界、危险、失效及风险缓解的场景集合，引入环境扰动、系统能力退化及人机交互不确定性，确保验证结论是否在不同风险条件下均成立。

（3）安全证据链维度：将仿真、场地试验、道路试验及运行监测等多源验证结果[14,15]，与具体状态机路径进行绑定，形成可追溯的路径级安全证据链，用于支撑工程结论与后续安全论证。

三维验证模型在典型运行模式转换路径下的应用示例如表3所示。该三维验证模型中，“状态机路径”确立验证的基本逻辑单元与流程框架；“场景类型”（标称、边界、危险、失效、风险缓解）为每条路径构建完整的风险谱系测试集；“安全证据链”则将仿真、测试、运行监测等多源证据与具体的“路径-场景”组合进行结构化关联与追溯。三者共同构成一个立体验证空间：以路径为纲，以场景为目，以证据为支撑，确保运行模式转换安全性论证实现全面覆盖与深度追溯。

3.2　验证活动组织与闭环机制

验证活动以状态机路径为主线，按照“设计验证-测试验证-运行验证”的递进逻辑组织，并通过统一的状态机模型、状态标识与数据接口实现不同阶段验证结果的一致性与可衔接性。

（1）设计阶段：通过形式化分析与模型审查，验证状态机定义的完备性与逻辑正确性，重点识别死锁状态、不可达安全状态及控制权或责任主体不明确的中间状态。

（2）测试阶段：围绕关键状态转换路径构建覆盖多类场景的验证集。其中，仿真用于实现高覆盖率与极端工况验证，场地试验用于验证仿真筛选出的低置

表3　三维验证模型示例(状态机路径-场景类型-安全证据链)

状态机路径	标称场景	边界场景	危险场景	失效场景	风险缓解场景
功能激活	正常激活	ODD不足拒激活	激活误触发防护	模块自诊断失败	激活中断→安全停车
模式转换	平顺转换	ODD变化触发TOR	外部风险干预优先级	通信异常→降级	用户不接管时MRC锁定
用户发起停用	正确操作	非典型输入防误触	停用中风险优先级	指令未执行→系统保护	用户停用→安全停靠
系统发起停用	正常停用闭环	异常触发停用提示	高风险→紧急停用	核心功能失效→MRC	停用后锁定→复位授权
远程终止	合法认证链执行	通信延迟超时	指令冲突/篡改响应	通道中断→本地策略	远程MRC与功能锁定
风险缓解与系统恢复	MRC→安全停靠	部分恢复需人工/远程授权	事故应急与乘员保护	持续故障→维修	授权复位→重新激活

信度危险场景,道路试验用于检验验证结论在真实交通情境中的适用性[16-17]。不同验证手段通过状态迁移日志、时间序列与事件标记进行结果对齐。

(3)运行阶段:通过在用监测与日志追溯,持续确认运行中状态转换路径与设计假设的一致性,并将运行数据作为安全论证的补充证据输入闭环验证体系[16-17]。

3.3　核心验证目标与指标体系

基于安全要求与具体技术要求[1,3,18],运行模式转换验证聚焦三类核心目标:系统行为安全、人机交互安全及验证结果可追溯性,据此构建以"关键指标-度量方式-适用场景-验证方法"为主线的验证指标体系(表4)。

表4　车辆运行模式转换验证指标示例

核心验证目标	关键指标	指标定义与度量方式	适用场景	验证方法与证据来源
系统行为安全	逻辑一致性	状态机路径正确、时序一致、无竞态、死锁或异常循环;基于FSM模型进行路径覆盖与一致性检验	标称、边界、危险	形式化建模、模型检验、仿真对比;证据:状态转换日志、模型验证报告
	响应性能	TOR提前量、后援用户响应时间、MRC触发与执行时序;测量系统动态响应链路并统计成功率	边界、危险	仿真、场地试验;证据:时序记录、控制轨迹数据、自动驾驶数据记录系统(DSSAD)日志
	控制权交接安全	接管过程车辆动态平顺、无控制冲突、轨迹误差在限值内;采集车辆动态、操控输入与轨迹偏差	标称、边界	驾驶模拟器、道路试验;证据:CAN数据、冲突检测记录、HMI录像
	风险缓解闭环	风险识别率、MRC触发率、执行成功率、MRC时序、最小安全停靠偏差;故障注入验证冗余闭环	失效、风险缓解	故障注入试验、冗余链路验证;证据:事件包、视频回放、应急控制记录
人机交互安全	接管提示有效性	TOR提示多模态覆盖率、提示提前量;采集提示触发时间、模态一致性与响应关联	边界、危险	驾驶模拟器、人因试验;证据:HMI日志、受试者响应记录
	后援用户可用性	接管窗口内恢复并执行控制的能力(可达性、状态、理解);指标:接管成功率、反应时间、姿态恢复时间	边界、失效	人因实验、姿态/注意力监测;证据:行为记录、姿态/生理参数、操作日志
	认知负荷	心率变化率与视觉注意分布;采集心率、眼动与任务负荷参数评估认知压力	边界	人因分析、认知模型验证;证据:生理信号、视线追踪数据
安全证据与追溯	日志完整性	DSSAD日志完整、关键字段齐全、时间同步误差≤限值;验证记录项、时间戳一致性与签名正确性	全场景	日志审计、记录比对;证据:DSSAD记录、签名校验报告
	因果链可还原性	状态路径"激活→转换→停用→恢复"完整可回放;回放轨迹与实际轨迹一致性≥阈值	全场景	回放验证、合规审计;证据:运行日志、轨迹比对报告
	法规符合性	满足ISO 26262、ISO 21448、技术指南、GTR ADS要求;验证材料与试验数据满足审查要求	全场景	审查与符合性测试;证据:安全实证档案、符合性测试报告

(1)系统行为安全:验证运行模式及状态转换过程中的逻辑一致性、响应及时性、控制权交接平顺性与风险缓解闭环的有效性。

(2)人机交互安全:验证系统提示的可感知性、交互逻辑的清晰性以及后援用户在接管场景下的操作安全性。

(3)安全证据可追溯性:对关键状态转换过程进行全过程记录与回放复核,支撑可审计的安全实证档案。

3.4　运行期监测与验证结果的持续更新

运行模式转换相关风险具有长期性与动态演化特征,仅依赖开发阶段验证难以覆盖全部运行工况。

通过对TOR触发、模式降级、MRC执行及用户接管等关键事件的统计与分析，可持续识别长尾风险场景，并用于状态机参数修正、场景库补充及证据链更新，最终形成“开发验证–运行监测–验证更新”的持续闭环验证机制[19]。

借助人工智能技术可进一步提升场景覆盖度与测评效率，例如利用机器学习生成长尾场景、依托大语言模型自动生成测试用例，实现自动驾驶安全测评体系自适应迭代[20-21]。

4 运行模式转换的安全论证构建

针对搭载多运行模式的自动驾驶系统，运行模式转换贯穿功能激活、运行维持、异常处置及风险缓解全过程，其安全性需要通过结构化的安全声明与系统的证据链予以论证[1]。本章基于前文所提出的运行模式转换设计方法与多维验证结果，旨在构建一套面向监管审查、具备高度可采信度的安全论证结构。本章的核心目标并非引入新的工程设计或验证方法，而是致力于将已有的设计与验证要素，系统性地组织成一条清晰、严谨、可追溯的论证路径。

4.1 安全声明结构

安全论证的前提是建立一个层次分明、与工程实践紧密关联的安全声明体系。为确保论证内容与系统实现之间的一致性，本文基于第3章构建的运行模式状态机模型来实例化安全声明体系[1]，将法规中的抽象安全目标，映射至具体的运行模式定义、状态转换条件及转换路径中，实现安全声明的具体化与工程化锚定。明确顶层声明，并结合运行模式转换过程中的主要风险来源，分解为3个核心子声明（SC1、SC2、SC3），分别对应监管审查的不同关注维度。安全声明层级结构如图5所示。

上述子声明共同构成覆盖设计、验证与运行全链条的完整论证框架。表5针对具有代表性的运行模式转换路径，示例性展示安全声明与GTR ADS法规条款之间的对应关系。其余等效或派生路径可依据相同的状态定义、触发逻辑进行扩展。

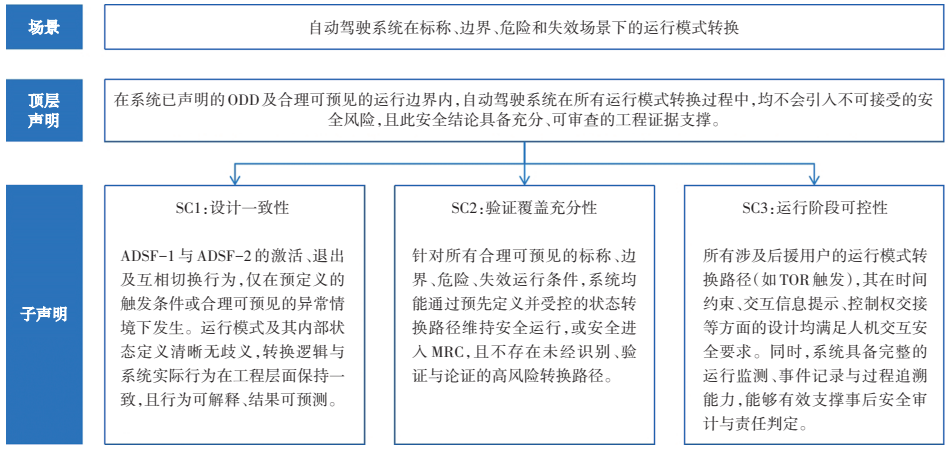


图5 安全声明结构

表5 安全论证要素与法规条款映射（示例）

转换类别	典型转换路径	关键系统安全行为与论证关注要素	对应法规条款
功能激活	未激活 → ADSF-1	系统自检、ODD判定、后援用户接管能力确认	§4.1.5.2 / §4.2.2.2
	未激活 → ADSF-2	系统健康状态诊断、通信链路完好性、无人运行可行性验证	§4.1.5.2 / §4.2.2.2
模式转换	ADSF-2 → ADSF-1	ODD边界识别、控制权与责任主体转移、用户确认机制	§4.2.2.2.4 / §4.1.6
	ADSF-1 → ADSF-2	ODD满足性验证、系统能力自评估、平滑切换控制	§4.2.2.2 / §4.3.7
用户发起停用	ADSF-1/2 → 手动 / MRC	用户接管意图有效性判断、异常输入处理机制	§4.2.2.3 / §4.2.3
系统发起停用	ADSF-1 → TOR → MRC	系统故障或性能退化识别、TOR触发与后援用户响应监测	§4.1.2 / §4.1.4 / §4.1.6
	ADSF-2 → MRC	无用户依赖条件下的自主风险识别、缓解策略执行	§4.1.5 / §4.1.6
远程终止	ADSF-1/-2 → MRC → 禁用	远程指令合法性认证、通信安全保障、功能持续禁用	§4.1.4.4 / §4.2.1.2
风险缓解与系统恢复	MRC → 安全状态 → 待机	系统恢复条件判定、复位授权机制、安全重新激活	§4.1.6.1 / §4.1.5.4 / §4.1.3.3.1

4.2 论证路径与证据组织

安全实证档案是结构化的文档体系，其核心是通过声明-论据-证据（Claims-Arguments-Evidence）的逻辑链，综合来自安全设计、仿真、场地和道路测试等多源信息，形成证明系统安全性的核心证据体[1,22]。在完成安全声明的层次化分解后，安全论证的关键任务在于构建一条逻辑严密、可被审查机构复核与采信的论证路径。该路径的核心是将第4章产生的多维验证结果与前文的设计定义进行系统性关联，组织成直接支撑各项安全声明的证据链。

本文采用“安全声明-状态转换路径-论证关注要点-证据引用”的结构化论证方法。该方法以运行模式状态机为工程锚点，将顶层安全目标逐层分解至具体的状态转换路径上，使得每一类转换场景均对应明确的论证侧重点与证据支撑要求。

在此框架下，不同类型的运行模式转换路径因风险特征差异，其论证焦点亦有所不同。例如，ADSF-1向ADSF-2的模式转换，论证焦点在于系统能力、无人后援的运行条件满足的充分性证明；而ADSF-2向ADSF-1或MRC的降级转换，则更侧重于异常识别的及时性、风险评估的准确性以及风险缓解策略的有效性。

4.3 典型状态转换路径的安全论证示例

为具体阐明第4.2节所提出的论证结构的可操作性，在当前以受控场景测试与示范运行为主的条件下，选取高速公路施工区作为具有代表性的复杂ODD边界场景，触发“ADSF-2向ADSF-1转换”这一转换路径，构建完整的安全论证示例。图6展示了从安全声明、论证逻辑到工程证据的完整映射关系。

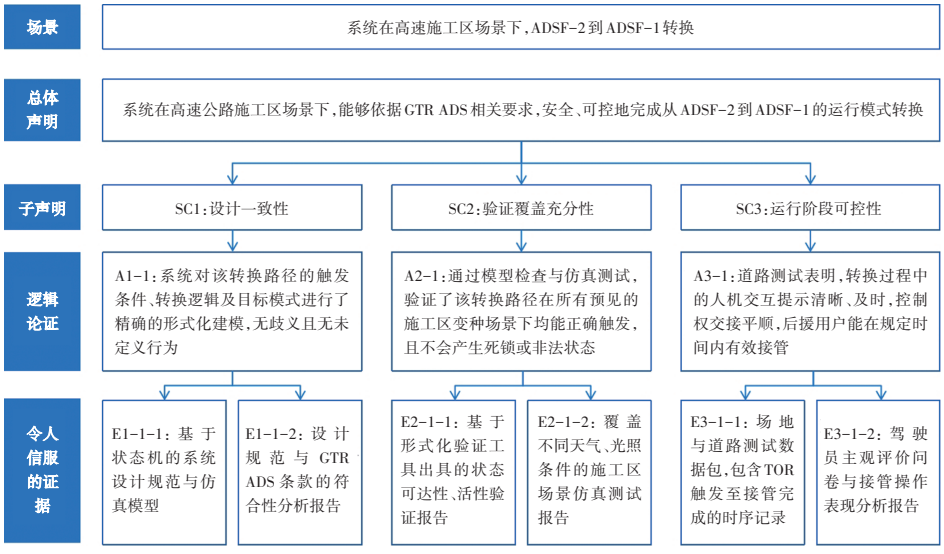


图6 典型状态转换路径的安全论证示例

通过上述结构化的论证示例可以看出，运行模式转换的安全性由多源、多层次的工程证据共同支撑，避免了依赖单一验证结论的风险，从而显著提升了安全论证的稳健性与监管可采信性。

4.4 与既有安全工程方法的关联

本文提出的运行模式转换安全论证方法，与功能安全（ISO 26262）、预期功能安全（ISO 21448）及系统安全分析技术（如FMEA、FTA、STPA等）构成互补与集成关系，旨在将既有安全实践的产出，系统性地整合为面向监管可采信的合规性安全实证档案。

（1）输入承接：基于ISO 26262/21448安全产出。本文的论证体系，其前提与约束直接来源于既有安全流程的核心输出，例如ISO 26262的危害分析与风险评估、安全目标与安全要求，以及ISO 21448的场景分析结果。这确保了本方法的论证基础与行业安全基准一致。

（2）分析聚焦：用状态机界定转换安全分析边界。“运行模式状态机”模型为FMEA、STPA等安全分析提供了以状态转换为核心的清晰分析边界。基于此，相关失效模式与不安全控制行为的识别，能够聚焦于模式激活、切换、降级等特定动态过程，提升了传统分析在转换安全问题上的针对性与效率。

（3）论证整合：结构化证据形成可审查安全实证档案。本文构建的“安全声明—状态转换路径—证据链”框架，核心功能是将设计、分析、测试、运行监测等多源异构证据，依照“运行模式转换”属性进行结构化组织、关联和追溯，最终形成可直接用于监管审查、闭环完整的顶层安全实证档案，实现从离散安全验证到系统级、可论证安全合规性的跃升。

因此，本文方法是对成熟安全工程体系的一次聚

焦性应用与论证化提升，专门为应对“运行模式转换”这一特定安全合规挑战，提供了具有工程可操作性的论证路径。

5 结论与展望

围绕多运行模式自动驾驶系统在监管审查与合规评估框架下面临的运行模式转换安全论证需求，针对ADSF-1与ADSF-2协同共存条件下的状态转换问题，提出了一套统一的运行模式转换建模与安全验证方法体系。主要结论如下：

（1）提出了面向多ADSF协同的运行模式转换形式化建模方法。基于FSM模型，将ADSF-1、ADSF-2及手动驾驶等运行模式纳入统一建模框架，通过明确状态定义、转换条件与行为响应，实现运行模式转换逻辑的形式化表达与一致性描述。该方法为多运行模式自动驾驶系统的工程设计、状态管理与验证提供了统一建模基础。

（2）构建了可支撑监管审查与合规采信的运行模式转换安全验证闭环。针对运行模式转换过程中标称、边界及异常工况并存的特点，本文提出“状态机路径-场景类型-安全证据链”的三维验证模型，将设计阶段逻辑验证、测试阶段行为一致性验证与运行阶段安全监测有机衔接，实现验证结果与法规条款、安全声明之间的结构化映射，为多运行模式自动驾驶系统提供了可追溯、可解释、可采信的安全验证方法。

随着多运行模式自动驾驶相关产品和应用场景的逐步推出，自动驾驶系统的验证与合规评估将进一步向系统级、一致性与全生命周期方向发展。在此背景下，未来研究可重点关注以下方面：一是结合动态ODD识别与约束能力，完善多ADSF协同条件下运行模式转换过程的可信度评估方法，明确模式切换的安全边界与触发条件；二是探索数字-物理融合的验证方法与结果采信机制，提升仿真验证在监管与合规中的支撑作用；三是探索人工智能技术在验证辅助与场景覆盖提升中的应用，例如利用机器学习进行长尾场景自动生成、基于AI的系统异常行为预测等，以提升验证的效率和覆盖度，为构建自适应安全测评体系提供支持；四是研究基于运行数据的持续验证与安全论证更新路径，构建与产品迭代和运行状态相适应的动态合规评估机制，以支撑自动驾驶系统的长期合规性与运行安全性。

参 考 文 献

[1] Informal Working Group on Automated Driving Systems. Proposal for a New United Nations Regulation on Uniform Provisions Concerning the Approval of Motor Vehicles with Regard to Their Automated Driving Systems[S]. Geneva: UNECE, 2026.

[2] Informal Working Group on Automated Driving Systems. Proposal for a New United Nations Global Technical Regulation on Automated Driving Systems (ADS) [S]. Geneva: UNECE, 2026.

[3] 工业和信息化部, 公安部, 住房和城乡建设部, 等. 关于开展智能网联汽车准入和上路通行试点工作的通知（工信部联通装〔2023〕217号）[EB/OL]. (2023-11-17)[2026-01-15]. https://www.miit.gov.cn/zwgk/zcwj/wjfb/tz/art/2023/art_be04a39345354b1ba4624262f86e7cfb.html.

[4] United Nations Economic Commission for Europe. Framework Document on Automated/Autonomous Vehicles: WP.29/2019/34/Rev.1[R]. Geneva: UNECE, 2019.

[5] Informal Working Group on Automated Driving Systems. ADS Feature Type Definitions(ADS-05-08) [C]// Proceedings of ADS IWG 5th Session, 2024.

[6] KBA. KBA Grants First Approval for Automated Driving[EB/OL]. (2021-12-09)[2025-10-27]. https://www.kba.de/DE/Presse/Pressemitteilungen/Allgemein/2021/pm49_2021_erste_Genehmigung_automatisiertes_Fahren.html.

[7] Mercedes-Benz Group AG. Drive Pilot Support Speed of up to 95 km/h on German Motorways[EB/OL]. (2024-12-17)[2025-10-27]. https://group.mercedes-benz.com/innovations/product-innovation/autonomous-driving/drive-pilot-95-kmh.html.

[8] Honda. Honda Receives Type Designation for Level 3 Automated Driving in Japan[EB/OL]. (2020-11-11)[2025-10-27]. https://global.honda/en/newsroom/news/2020/4201111eng.html.

[9] 工业和信息化部. 我部许可两款L3级自动驾驶车型产品[EB/OL]. (2025-12-15) [2025-12-16]. https://www.miit.gov.cn/jgsj/zbys/gzdt/art/2025/art_2f4447d835e14daeb6717276ec8bbf88.html.

[10] 国家市场监督管理总局, 国家标准化管理委员会. 汽车驾驶自动化分级: GB/T 40429 —2021[S]. 北京: 中国标准出版社, 2021.

[11] ISO. Road Vehicles-Safety of the Intended Functionality: ISO 21448: 2022[S]. Switzerland: ISO, 2022.

[12] ISO. Road Vehicles-Functional Safety: ISO 26262: 2018[S]. Switzerland: ISO, 2018.

[13] MOHAMMED C, ABDERRAOUF B, JULIE B, et al. A Risk-Based Decision-Making Process for Autonomous Trains Using POMDP: Case of the Anti-Collision

Function[J]. IEEE Access, 2024, 12: 5630-5647.
[14] The EDR/DSSAD Informal Working Groups. Draft Regulations for DSSAD (ADS-12-48)[C]// Proceedings of ADS IWG 12th Session, 2025.
[15] 邱彬, 李广友, 薛晓卿, 等. 智能网联汽车数据交互与综合应用公共服务平台研究与构建[J]. 汽车工程学报, 2024, 14(5): 829-838.
[16] United Nations Economic Commission for Europe. New Assessment/Test Method for Automated Driving (NATM) - Master Document[R]. Geneva: UNECE, 2021.
[17] 刘法旺, 徐晓庆, 陈贞, 等. 搭载自动驾驶功能的智能网联汽车安全测试与评估方法研究[J]. 汽车工程学报, 2022, 12(3): 221-227.
[18] 国家市场监督管理总局, 国家标准化管理委员会. 智能网联汽车 自动驾驶系统通用技术要求: GB/T 44721—2024[S]. 北京: 中国标准出版社, 2024.
[19] United Nations Economic Commission for Europe. Guidelines and Recommendations for Automated Driving System Safety Requirements, Assessments and Test Methods[R]. Geneva: UNECE, 2024.
[20] 陈贞, 李京泰, 郭煌, 等. 应用人工智能技术的自动驾驶系统安全测评方法研究[J]. 汽车工程学报, 2025, 15(4): 457-467.
[21] 周时莹, 陈霞, 张东波, 等. 基于大语言模型的智能驾驶测试用例自动生成系统[J]. 汽车文摘, 2025(10): 6-18.
[22] 陈泽众, 邓玉欣. 关于安全案例论证构建的综述[J]. 软件学报, 2024, 35(9): 4013-4037.

（责任编辑　**明慧**）

《汽车技术》征稿启事

《汽车技术》杂志是中国第一汽车集团有限公司主办的国内外公开发行的汽车前瞻与应用技术类月刊，为我国高质量科技期刊分级目录入选期刊、中国科学引文数据库（CSCD）来源期刊、中文核心期刊、中国科技核心期刊、RCCSE中国核心学术期刊（A）、Scopus数据库收录期刊、俄罗斯《文摘杂志》（AJ）收录期刊、日本科学技术振兴机构数据库入选期刊、EBSCO学术数据库收录期刊、欧洲学术出版中心（EuroPub）数据库收录期刊。

《汽车技术》杂志以报道汽车整车及其零部件设计、研究、试验等方面的前瞻与应用技术为主，并兼有理论研究内容，是中国汽车行业核心学术和知识传播与共享的平台。

《汽车技术》将在国家提出的“创新、协调、绿色、开放、共享”发展理念的指引下，把握《节能与新能源汽车技术路线图》和“低碳化、信息化、智能化”的汽车技术主流发展趋势，努力在传统内燃机汽车高效动力系统、轻量化、低阻力领域，新能源汽车和互联智能汽车技术领域，大力吸收优质稿源，为广大科研和工程技术人员服务，为我国汽车工程技术创新能力提升贡献力量。

《汽车技术》欢迎高等院校师生、研发工程技术人员、技术管理人员及相关人员不吝赐稿，反映国家重点扶持项目、自然科学基金项目和其他重点项目等研究成果的稿件将被优先选择刊登。

投稿要求：

1. 文章字数最好控制在6 000~8 000字范围之内；

2. 请按科技论文要求撰写文章摘要，摘要中文字数控制在180字左右；

3. 文章必须附有公开发表的、体现本领域最新研究成果的参考文献，且在文中应标注文献引用处；

4. 文章主要作者应提供其简介，包括出生年、性别、职称、学历、研究方向及技术成果等；

5. 来稿的保密审查工作由作者单位负责，确保署名无争议，文责自负；

6. 请勿一稿多投；

7. 本刊使用网站投稿，请先登陆网站注册成功后投稿，详细投稿要求见本刊网站中“下载中心”栏的“作者指南”，网址：http://qcjs.cbpt.cnki.net。

《汽车技术》编辑部