

团体标准

T/SHV2X 9—2025

汽车驾驶自动化系统通用语料库 第1部分：总体要求

General corpus for automotive driving automation system—
Part 1: General requirements

2025 - 07 - 25 发布

2025 - 07 - 25 实施

上海市车联网协会 发布

全国团体标准信息平台

目 次

前 言..... II

引 言..... III

1 范围..... 1

2 规范性引用文件..... 1

3 术语和定义..... 1

4 基本要求..... 1

 4.1 建设原则..... 1

 4.2 建设目标..... 1

5 质量要求..... 2

6 管理与安全..... 2

7 运行与维护..... 2

参 考 文 献..... 3

前 言

《汽车驾驶自动化系统通用语料库》系列标准拟分为六个部分：

- 第1部分：总体要求；
- 第2部分：术语和定义；
- 第3部分：语料数据采集；
- 第4部分：语料数据清洗；
- 第5部分：语料数据标注；
- 第6部分：语料数据测试；

本文件为第1部分。

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由上海市车联网协会提出并归口。

本文件起草单位：上海库帕思科技有限公司、智能汽车创新发展平台（上海）有限公司、智己汽车科技有限公司、地平线征程（上海）科技有限公司、上海临港绝影智能科技有限公司、上海机动车检测认证技术研究中心有限公司、上海金桥智能网联汽车发展有限公司、北京一辅智行科技有限公司、上海优咔网络科技有限公司、亿咖通（上海）技术有限公司、滴水智行科技有限公司、上海航盛实业有限公司、上海阶跃星辰智能科技有限公司、东华大学、上海交通大学、上研智联智能出行科技（上海）有限公司、北京赛目科技股份有限公司、苏州柏川数据科技有限公司、华为技术有限公司。

本文件主要起草人：山栋明、黄海清、施佳樑、郭辉、赵九花、贺锦鹏、蒋达夫、周剑鸣、黄剑其、邵亚萌、徐春雷、解瀚光、朱雷、谭龙欢、于峰、滕添益、张帆、林瑜、周轶、袁月明、李学根、张裕珍、曹宇、邓思文、贺仁驹、李晨歌、马昊、陈紫娟、杨闻博、丁楚晨、李想、蔡雨辰、黄鹏飞、刘建业、李勋宏、马骏、李轶刚、刘壹青、田浩、郭晓宾、董连飞、范昌琪、李璟、孙雯、陈巧慧、王娜、沈滨、孔令和、鲁江东、李成虎、汪日甜、何丰、谭哲、薛晓卿、刘兴、马东升、刘鹏宇、邓子涵。

本文件首批承诺执行单位：上海库帕思科技有限公司、智能汽车创新发展平台（上海）有限公司、智己汽车科技有限公司、地平线征程（上海）科技有限公司、上海临港绝影智能科技有限公司、上海机动车检测认证技术研究中心有限公司、上海金桥智能网联汽车发展有限公司、北京一辅智行科技有限公司、上海优咔网络科技有限公司、亿咖通（上海）技术有限公司、滴水智行科技有限公司、上海航盛实业有限公司、上海阶跃星辰智能科技有限公司、东华大学、上海交通大学、上研智联智能出行科技（上海）有限公司、北京赛目科技股份有限公司、苏州柏川数据科技有限公司、华为技术有限公司。

引 言

为了实现可靠的汽车驾驶自动化,大量准确的语料数据至关重要。语料数据是自动驾驶系统的基础,它能够帮助系统理解复杂的交通环境、做出明智的决策,并不断优化性能。随着端到端模型整合度越来越高,对训练语料的规模和质量要求也成倍增加。大规模、高质量的数据标注(特别是端到端感知数据和推理数据)是构建安全、可靠、高性能自动驾驶系统的基石。围绕“采、洗、标、测”一体化流程,制定《汽车驾驶自动化系统通用语料库》系列标准,包括:

- 第1部分:总体要求,明确系列标准建设的总体要求,为后续各部分标准的制定提供指导和基础;
- 第2部分:术语和定义,统一系列标准建设过程中涉及的关键术语和定义;
- 第3部分:语料数据采集,对语料数据资源格式提出规范要求;
- 第4部分:语料数据清洗,针对采集好的数据,对语料数据清洗的流程与质量要求提出规范要求;
- 第5部分:语料数据标注,针对清洗好的数据,对语料数据标注的适用场景、标注内容、标注方式、数据存储提出规范要求;
- 第6部分:语料数据测试,针对标注好的数据,对语料数据测试的流程和质量要求提出规范要求。

本文件为第1部分总体要求。通过本文件的制定,明确《汽车驾驶自动化系统通用语料库》系列标准建设原则与建设目标。通过构建可复用的标准化自动驾驶训练数据集,促进汽车驾驶自动化系统语料资源高效流通利用,降低业内企业在重复数据采集、清洗、标注等方面的巨额成本,提升汽车企业竞争力且推进汽车产业健康发展。

汽车驾驶自动化系统通用语料库 第1部分：总体要求

1 范围

本文件规定了汽车驾驶自动化系统通用语料库的基本要求、质量要求、管理与安全、运行与维护等内容。

本文件适用于企业、研究机构对汽车驾驶自动化系统通用语料库的研究、开发、维护、应用、评估等工作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 39732—2020 汽车事件数据记录系统

GB 44497—2024 智能网联汽车 自动驾驶数据记录系统

GB/T 14911—2008 测绘基本术语

GB/T 17159—2009 大地测量术语

GB/T 25000.12—2017 系统与软件工程 系统与软件质量要求和评价(SQuaRE)第12部分：数据质量模型

GB/T 36073—2018 数据管理能力成熟度评估模型

3 术语和定义

T/SHV2X 10—2025文件界定的术语和定义适用于本文件。

4 基本要求

4.1 建设原则

《汽车驾驶自动化系统通用语料库》系列标准建设原则应包括：

- 科学性：语料库数据标准建设应遵循科学、系统、规范的原则，确保数据的代表性、全面性和一致性；
- 开放性与共享性：宜鼓励数据资源开放共享，促进行业协同创新和资源高效利用；
- 安全性与合规性：应严格遵守国家相关法律法规，保障数据安全、隐私保护和合规流通；
- 可扩展性与可持续性：语料库结构和管理机制应具备良好的扩展性，支持持续更新和维护。

4.2 建设目标

4.2.1 语料数据采集

在汽车驾驶自动化系统通用语料库的语料数据采集过程中，采集的数据内容和方法应符合 GB 44497、GB 39732、GB/T 17159 以及 GB/T 14911 的规定。

语料数据资源的分类、采集要求、采集安全等方面的要求应满足 T/SHV2X 1—2025 的规定。

4.2.2 语料数据清洗

在汽车驾驶自动化系统通用语料库的语料数据清洗过程中，应对采集原始数据进行系统性处理，保留高价值数据资源用于数据标注。

语料数据预处理、数据清洗以及质量等方面的要求应满足 T/SHV2X 11—2025 的规定。

4.2.3 语料数据标注

在汽车驾驶自动化系统通用语料库的语料数据标注过程中，应通过人工或标注工具，对自动驾驶数据样本指定目标变量和赋值，支持多类型标注需求，实现为机器学习提供监督信号。

语料数据标注的适用场景、标注方式、标注内容及数据存储等方面的要求应满足 T/SHV2X 12—2025 的规定。

5 质量要求

汽车驾驶自动化系统通用语料库的质量应满足：

- a) 完整性：有效数据应覆盖多样化场景，包括但不限于不同道路类型、天气条件、交通参与者等；
- b) 准确性：采集、清洗和标注过程应有严格的质量控制，确保数据真实、准确；
- c) 一致性：数据格式、标注规范和管理流程应一致，便于数据的集成与复用。

6 管理与安全

语料产品数据的提供方应在组织、人员、制度、平台等方面实施安全策略、执行覆盖资源数据和语料产品全生命周期的安全控制措施，应执行并满足 T/SAIAS 015—2024《语料库建设导则》、T/CAAMTB 77—2022《汽车传输视频及图像脱敏技术要求与方法》、自然资发〔2024〕139号《自然资源部关于加强智能网联汽车有关测绘地理信息安全管理的通知》的相关数据安全规定。

汽车驾驶自动化系统通用语料库在管理与安全方面应满足：

- a) 在数据安全方面，应建立完善的数据安全管理体系，防止数据泄露、篡改和非法访问；
- b) 在隐私保护方面，需对涉及个人信息的数据进行脱敏和匿名化处理，保障用户隐私；
- c) 在权限管理方面，应设定合理的数据访问和操作权限，确保数据使用的安全性与合规性；
- d) 在数据备份与恢复方面，应建立数据备份机制，保障数据的可靠存储和灾难恢复能力。

7 运行与维护

汽车驾驶自动化系统通用语料库在运行与维护方面应满足：

- a) 持续维护：应建立语料库的更新、维护和质量评估机制，确保数据长期可用，反映最新的自动驾驶应用场景和行业需求。
- b) 技术支持：应配备专业的技术团队，负责语料库的建设、运行、维护和技术创新。

参 考 文 献

- [1] T/CAAMTB 77—2022 汽车传输视频及图像脱敏技术要求与方法
 - [2] T/SAIAS 015—2024 语料库建设导则
 - [3] T/SHV2X 10—2025 汽车驾驶自动化系统通用语料库 第2部分：术语和定义
 - [4] T/SHV2X 1—2025 汽车驾驶自动化系统通用语料库 第3部分：语料数据采集
 - [5] T/SHV2X 11—2025 汽车驾驶自动化系统通用语料库 第4部分：语料数据清洗
 - [6] T/SHV2X 12—2025 汽车驾驶自动化系统通用语料库 第5部分：语料数据标注
 - [7] 智能网联汽车时空数据安全处理基本要求
 - [8] 自然资发〔2024〕139号 自然资源部关于加强智能网联汽车有关测绘地理信息安全管理的通知
-