# 评估智能辅助驾驶系统信任度的试驾环节执行方案

张乐晗 2025.8.24

用户试驾研究，是整个智能辅助驾驶系统的信任问题研究中的实验环节。我们设计了一套结合主观、客观与生理指标的多模态数据收集方案，通过数据三角验证法将这些数据转化为可操作的安全教育视频设计建议，最终实现人机交互中的“信任校准”。

## 第一部分：智能驾驶信任度的概念基础

在设计任何评估方案之前，必须对信任这一复杂概念建立深刻的理解。本部分将深入探讨人机信任的理论维度、其动态演变过程以及校准不当所带来的潜在安全风险。

### 1.1 信任的定义与维度

信任在自动化系统成功应用中扮演着关键角色，它影响着人们对日益自动化的系统的接受度和依赖程度。

通过Jian 等人的研究我们发现，将社会心理学中成熟的信任框架和理论应用于人机交互领域提供了坚实的经验基础。这意味着我们不必从零开始构建理论模型，而是可以借鉴并调整已有的研究成果。

然而，Jian 等人最初关于信任和不信任是单一维度的结论也受到后续研究的挑战。一些研究认为，信任和不信任可能作为两个独立的、可以同时存在的构念。在数据分析中，需要对此保持敏锐，因为“低信任得分”可能不等于“高不信任得分”，它们可能反映出不同的心理状态。

### 1.2 信任校准

信任并非一成不变的静态概念，而是一个动态的心理过程，它会随着用户的经验和具体的驾驶情境而不断变化。这种变化过程被称为“信任校准”，即用户的信任水平动态调整以匹配自动化系统的实际能力和局限性。一个理想的信任水平是，它既不过高也不过低，恰好与系统的实际性能相匹配。

要全面捕捉这种动态性，研究设计需要超越单次的事后评估。 Hoff 和 Bashir 提出的信任三层模型提供了一个全面的分析框架，本研究方案将以此为基础：

1. **倾向性信任（Dispositional Trust）：** 个人固有的、先天的信任倾向，可以作为驾驶员对自动化系统信任的基线衡量。
2. **初始习得信任（Initial Learned Trust）：** 用户在正式试驾前通过简短接触（例如听取系统介绍）所形成的初步信任。
3. **动态习得信任（Dynamic Learned Trust）：** 在持续交互过程中，基于系统实际表现而实时波动的信任水平。这是试驾环节需要重点捕捉的数据。

仅仅测量试驾结束时的平均信任得分是不够的。真正的价值在于理解信任**如何**随时间波动。例如，在平稳、常规的驾驶路段，驾驶员的信任度可能稳步上升，但一次突发的、意外的自动驾驶事件（例如系统突然的急刹车或转向）可能会导致信任度急剧下降。这种信任的动态变化，而不是最终的平均值，才是理解系统行为如何影响用户信任的关键。因此，试驾场景必须精心设计，以刻意引发这种信任的波动。

### 1.3 信任校准不当的危害：过度依赖与不使用

信任校准的目标并非是实现最高的信任度，而是要达到**恰当的信任校准**。信任校准不当会导致两种潜在的危险状态：

1. **过度依赖（Over-reliance）或滥用（Misuse）：** 这是一种自动化偏见（automation bias），指驾驶员过度信任系统，导致其警惕性降低，在系统出现故障或超出能力范围时，响应迟缓或不当 2。研究表明，如果自动化系统表现出高度可靠性，驾驶员可能会形成“习得性粗心”（learned carelessness），从而不再警惕，但如果系统首次出现故障，用户的信任度会急剧下降，这种现象被称为“首次故障效应”（first-failure effect）。
2. **不使用（Disuse）或信任不足（Under-trust）：** 这种情况发生在驾驶员对系统信任度过低时，即使系统高度可靠，他们也拒绝使用或依赖其功能 1。这会削弱智能驾驶技术在安全和效率方面带来的潜在益处。

由此可见，信任评估不仅仅是一个用户体验指标，更是一个至关重要的**安全指标**。校准不当的信任水平与驾驶员的危险行为（如延迟接管或注意力不集中）直接相关，这些行为可能导致交通事故。因此，本研究的发现可以直接解决核心安全问题，这对任何汽车制造商或技术公司都具有重大的商业和工程价值。

## 第二部分：信任评估方案设计

### 2.1 试驾环节的构建

#### 2.1.1 设定运营设计域（ODD）与情境

在设计试驾路线时，首先需要明确智能驾驶系统所处的运营设计域（Operational Design Domain, ODD），即系统被设计用于安全运行的环境限制。这包括：

* **地理区域：** 高速公路、城市街道、特定封闭区域。
* **道路条件：** 弯道、坡道、出入口匝道。
* **交通状况：** 拥堵、车流顺畅、混合交通。
* **环境条件：** 白天、夜间、雨雪雾等恶劣天气。

因此，我们常选择工作日下午进行试驾实验，此时车流量处于正常值，不会影响基础城区智驾功能的使用；车企接待压力较小，更有可能长时间体验高速等特殊路段。

#### 2.1.2 设计受控场景与关键事件

为了有效评估信任，试驾不能仅仅是常规的日常驾驶。研究方案必须包含精心设计的**关键事件（critical incidents）**，以探查系统的能力边界并引发驾驶员的信任响应，例如：

* **交通事件：**
  + **自动紧急制动（AEB）：**在前车突然制动或虚拟障碍物出现时，系统自动触发制动。
  + **意外变道：**为了避让突然出现的障碍物（如行人、自行车），系统执行一次急剧的自动变道。
  + **加减速变化：**系统在跟车时表现出不寻常的、非平滑的加减速风格。
* **环境事件：**
  + **复杂视觉环境：** 系统需要识别并避让施工路段的复杂视觉标志。
  + **对行人的反应：** 在有行人和自行车混行的路段，测试系统对弱势交通参与者的识别和反应能力。

**（事实上我们最后发现，只要正常开车，这些场景全都会遇到）**

#### 2.1.3 试驾前准备：基线评估与任务说明

在正式试驾开始前，需要对参与者进行充分的准备和基线评估：

* **任务说明：**
  + 清晰地向参与者解释系统的能力和局限性，确保他们对系统的功能有正确的认知。
  + 明确参与者在与车企员工交涉时保持“智驾体验者”的身份，减少可能的干扰性因素，充分体验智能驾驶。
* **基线评估：**
  + 使用简短问卷来测量参与者的**倾向性信任**。
  + 在使用前简短访谈，记录他们过往使用驾驶辅助系统的经验。

### 2.2 综合数据收集策略

#### 2.2.1 主观数据收集：衡量感知信任

主观数据主要通过量表和访谈来收集，以捕捉驾驶员对系统信任度的感知。

**1. 利用 Jian et al. 的自动化系统信任量表**

该量表是衡量人机信任最常用的标准化工具之一。研究将使用其12个项目，采用7点李克特量表（1 = 完全不同意，7 = 完全同意）进行评分。为确保量表的有效使用，必须区分正向和反向评分项目。

以下是根据 Jian et al. 的研究成果构建的12项量表：

|  |  |  |  |
| --- | --- | --- | --- |
| 项目编号 | 量表项目（中文翻译） | 原始英文项目 | 评分方向 |
| 1 | 该系统具有欺骗性。 | The system is deceptive. | 反向 |
| 2 | 该系统行为鬼鬼祟祟。 | The system behaves in an underhanded manner. | 反向 |
| 3 | 我怀疑该系统的意图、行动或输出。 | I am suspicious of the system’s intent, action, or output. | 反向 |
| 4 | 我对该系统保持警惕。 | I am wary of the system. | 反向 |
| 5 | 该系统的行为将导致有害或伤害性结果。 | The system’s action will have a harmful or injurious outcome. | 反向 |
| 6 | 我对该系统充满信心。 | I am confident in the system. | 正向 |
| 7 | 该系统提供安全保障。 | The system provides security. | 正向 |
| 8 | 该系统具有正直性。 | The system has integrity. | 正向 |
| 9 | 该系统是可靠的。 | The system is dependable. | 正向 |
| 10 | 该系统性能稳定。 | The system is reliable. | 正向 |
| 11 | 我可以信任该系统。 | I can trust the system. | 正向 |
| 12 | 我对该系统很熟悉。 | I am familiar with the system. | 正向 |

*注：反向评分项目（1-5）在计算总分时需要将得分反转，例如，1分计为7分，2分计为6分，以此类推，以确保高分始终代表高信任度。*

**2. 进行结构化试驾后访谈**

访谈是理解定量数据背后“为何”的必要补充。它能捕捉到量表无法触及的个体化感受、思维过程和决策动机。

* **访谈策略：** 访谈采用“关键事件技术”作为框架，要求参与者回忆并详细描述试驾过程中那些令其感到信任度最高或最低的特定时刻。
* **访谈问题举例：**
  + **“请描述一下您对这次试驾的总体体验。”**
  + **“当系统在遇到行人时突然制动时，您当时在想什么？系统这个行为让您感觉如何？”**
  + **“您提到对系统感到警惕，请问在刚才的试驾过程中，是系统的哪一个具体行为让您产生了这种感觉？”**
  + **“这次试驾过程中有哪些节点是让您印象深刻的？在这几个事件中，智驾系统如何反应？这些事件分别使您对这个系统的认识产生了什么变化？”**

#### 2.2.2 客观数据收集

为了超越主观报告的局限性，必须收集客观的行为和生理指标。

|  |  |
| --- | --- |
| 数据点 | 与信任的关联 |
| **接管时间（Takeover Time）** | 驾驶员在系统请求接管或发生故障时的反应速度。信任度高或过低都可能导致接管延迟。（事实证明也可能是因为操作不熟练） |
| **手部位置（Hands on Wheel）** | 衡量驾驶员的警惕性水平。手部长时间脱离方向盘可能表明过度依赖。 |
| **行程距离与时长** | 衡量自动化功能的使用频率和时长，是信任的直接行为体现。 |
| **系统提示次数与类型** | 记录系统向驾驶员发出的所有警告、接管请求或信息提示。 |

## 第三部分：数据整合与可操作洞察

如何将第二部分收集到的多维数据进行整合和分析，从而得出具有高可信度的结论并转化为具体的产品设计？

### 3.1 针对量表的数据分析

传统的总分比较虽然能显示信任的整体变化，但它提供的洞察力是有限的 。为了设计有效的安全教育视频，我们需要深入理解信任失衡的具体原因。我们将采用**正向与负向条目的分维度分析**，将TIAS-AD量表中的12个条目，分为两个子维度进行分析。

**具体分析步骤：**

* **计算子维度变化分数：** 分别计算每个参与者在驾驶前后，正向信任子维度和负向不信任子维度的平均分变化。
  + **信任变化分数** = 驾驶后信任子维度总分 - 驾驶前信任子维度总分
  + **不信任变化分数** = 驾驶后不信任子维度总分 - 驾驶后不信任子维度总分
* **解读数据结果：** 结合两种分数的具体变化，你可以得出更细致的结论：
  + **健康信任校准（校准成功）：** 驾驶后，信任子分数显著上升，不信任子分数显著下降。这表明系统性能良好，交互体验清晰，用户建立了健康的信任。
  + **过度信任风险（教育重点）：** 驾驶前信任子分数很高，但驾驶后不信任子分数没有显著变化，甚至有所下降。这表明用户可能存在“过度信任”的倾向。你的教育视频需要重点提醒用户系统能力的边界，避免在驾驶过程中过度分心 。
  + **不信任倾向（教育重点）：** 驾驶前信任子分数很低，不信任子分数很高。这表明用户存在“不信任”倾向，他们可能对新兴技术持怀疑态度。教育视频需要重点展示系统的可靠性、透明度，并通过真实场景演示来建立用户的信心 。
  + **信任失衡（设计与教育并重）：** 驾驶后，信任子分数没有显著提升，但负面不信任子分数却显著上升。这表明系统在驾驶过程中出现了让用户感到“可疑”或“不透明”的行为。这不仅仅是教育问题，更是系统设计缺陷的反映。你的视频需要解释这些“可疑”行为背后的原理，同时为工程师提供优化系统设计的依据。

### 3.2 数据三角验证法：整合混合方法数据

本方案的核心思想是“方法论三角验证”，即将定量数据（量表）和定性数据（访谈）结合起来。定量数据回答了“什么”和“多少”的问题（例如，有多少驾驶员在系统急刹时接管了控制？），而定性数据则解释了“为什么”（例如，为什么他们选择在那个特定时刻接管？他们当时经历了怎样的心理斗争？）。

我们将构建一个“关键事件分析矩阵”，把一次特定事件的数据（例如，车辆因前方突然出现障碍物而执行自动紧急制动）在不同数据维度上进行对齐，以提供一个全面的图景。

**关键事件分析矩阵：**

|  |  |  |  |
| --- | --- | --- | --- |
| 数据类型 | 数据来源 | 数据点 | 分析与洞察 |
| **主观感知** | 量表评分 | 试驾后信任量表得分 | 参与者对系统的“可靠性”和“信任”项目评分较低。 |
| **定性叙述** | 试驾后访谈 | “我当时觉得系统不会停下来，所以我自己踩了刹车。” | 用户的口述解释了量表评分低的原因：他们对系统的制动能力缺乏信心，或觉得系统响应不及时。 |
| **客观行为** | 试驾过程记录 | 驾驶员接管时间、踩刹车踏板次数、接管频率 | 驾驶员在系统自动制动前0.5秒接管控制。 |

通过这个矩阵，我们可以发现系统性能（定量）与用户感知（定性）之间的脱节。例如，系统可能在客观上表现完美，但如果驾驶员的主观感受表明他们并不信任，那么这本身就是一个需要解决的设计问题。

### 3.3 结论推导与安全教育视频设计

我们的最终目标是“校准信任”，即确保信任水平与系统的实际能力相匹配。当信任水平过低时，我们希望通过安全教育视频建立适当的信任；当信任水平过高时，我们的优化重点是“管理信任” 。

**核心原则：**

1. **告知系统边界：** 明确系统的能力边界和局限性。
2. **培养用户能力：** 鼓励和训练用户保持警惕，并能在关键时刻有效地接管。

* **当“系统很可靠”得分显著低于预期时：**
  + **洞察：**驾驶员对系统的核心性能（如车道保持、自动跟车）产生了怀疑。这可能是因为系统在某些特定场景（如弯道、恶劣天气）表现不佳。
  + **视频内容：**
    - **正确校准信任：**明确告知用户智能驾驶系统的适用范围和局限性。例如，在视频中展示当系统处于雨雪或光线昏暗环境时，其性能可能受限，并提醒驾驶员此时应保持警惕，随时准备接管。
    - **提升可靠性感知：** 视频可以展示系统在不同复杂场景下（如交通拥堵、频繁变道）如何平稳运行，通过“眼见为实”的方式重建用户的信心 。
* **当“这个系统有欺骗性”或“行为不正当”得分显著上升时：**
  + **洞察：** 用户对系统行为的意图难以理解。例如，系统可能在未提前告知的情况下突然加速或减速，导致用户感到被动或被欺骗。
  + **视频内容：**
    - **增强透明度：** 视频需要详细解释系统在何种情况下会做出何种决策。例如，当系统检测到前方有物体靠近时，会提前通过仪表盘或抬头显示（HUD）发出预警，并以视觉和听觉信号告知驾驶员即将采取制动措施 。
    - **展示意图：**通过动画或图解方式，清晰地展示系统“内部”的决策过程，例如系统如何识别前方的车辆、行人和骑行者，以及如何据此规划行车路径，从而让抽象的信任转化为具体的可理解行为。
* **当“我熟悉这个系统”得分普遍较低时：**
  + **洞察：**驾驶员对系统的功能、模式和人机交互（HMI）界面不熟悉，导致无法建立有效的心理模型。
  + **视频内容：**
    - **功能教学：**视频应作为用户手册的动态补充，系统地介绍各个功能按钮、显示图标的含义。
    - **情境化教学：**模拟真实驾驶场景，展示系统在不同模式下（如自适应巡航、车道保持辅助）如何运行，以及如何通过简单的操作在自动驾驶和人工驾驶之间进行平滑切换，以提升系统的可学习性。（但这个是否真实可行？）

同时，我们还需要警惕“过度信任”带来的潜在风险。当正面信任子分数显著上升时，需要进一步探究这种信任是否合理。

* **结合负面不信任子分数进行交叉分析：**
  + **健康的信任提升：** 如果正面信任分数上升的同时，负面不信任分数显著下降，这表明系统成功消除了用户的疑虑和警惕。这种信任提升是健康的。
  + **潜在的过度信任风险：** 如果正面信任分数上升，但负面不信任分数变化不大或甚至略有上升，这可能是一个危险信号。它意味着用户虽然对系统的能力（如可靠性、安全性）有了更强的信心，但内心深处的警惕（如“系统有欺骗性”、“行为不正当”）并未完全消除 。这可能是因为系统的透明度或可解释性还存在不足，而这种“隐性”的疑虑在未来可能被放大，导致信任突然崩溃。
* **结合行为数据进行分析：**
  + **衡量“自动化倦怠”风险：** 评估驾驶员在自动驾驶模式下的行为。例如，如果他们开始频繁地进行非驾驶任务（如玩手机、看视频），或者在需要接管时反应迟钝，这可能表明他们对系统产生了过度依赖（也称为“自动化偏见”或“习得性粗心”）。
  + **分析接管行为：** 观察并记录驾驶员在系统请求接管时的反应时间。一个健康的信任值应该使驾驶员在系统请求时能迅速且有效地接管 。如果他们因为过度信任而延迟接管，则可能存在安全隐患。
* **当“很可靠”和“值得依赖”得分过高时：**
  + **视频洞察：** 用户可能认为系统在任何情况下都无懈可击。
  + **视频内容：**
    - **情景模拟：** 展示系统在极端或非设计场景下的表现，例如在暴雨、大雾、冰雪天气，或者在没有清晰车道线的施工路段 。明确告知用户：在这种情况下，系统可能无法正常工作，并会向用户发出接管请求。
    - **失败演示：** 在一个安全可控的环境下，模拟系统因传感器被遮挡或故障而无法正常工作的场景。通过清晰的视觉和听觉提示，演示用户如何在此类情况下迅速且安全地接管。这种“失败演示”可以有效降低用户的过度信任，同时强化他们对自身能力的信心 。
* **当“有信心”和“提供安全保障”得分过高时：**
  + **视频洞察：** 用户可能将智能驾驶系统等同于“绝对安全”的保障。
  + **视频内容：**
    - **法律与责任：** 明确告知用户，尽管系统是自动化的，但在大多数情况下，驾驶员仍然是最终的责任人，需要持续监控驾驶环境 。
    - **驾驶员角色转变：** 解释驾驶员的角色从“主动驾驶”转变为“监控与准备接管”。用生动的比喻，如“你不再是司机，而是副驾驶，但你随时需要准备好成为司机”。