

智能驾驶试驾实验数据分析

2025.9.20 张乐晗

信任腐蚀剂的主题分析：关键负面事件与系统性缺陷

3.1 “黑箱”式失误的失衡影响：单一负面事件的力量

一次严重的、无法理解的系统失误，对信任的破坏力远超多次的成功表现。这种失衡效应的核心在于错误的“黑箱”本质——其发生是不可预测的，其原因也是不透明的。

- 高风险的关键负面事件：** 朱明宇在四元桥匝道的经历是此类事件的典型代表。在两次尝试向左变道失败后，系统突然执行了一次他描述为“一脚油门啪一下就往右边甩出去”的剧烈操作，险些冲向护栏。这一事件被他定义为“唯一一个预期之外的事情”和“完全超乎我们预期的”。正是这次事件，让他在试驾后依然对系统“保持警惕”，成为了他信任提升的天花板。
- 明确的规则违反：** 同样，蒋咸康观察到系统试图在实线处变道，这一行为直接违背了交通规则。这个单一事件让他立刻“意识到智能驾驶还远没有达到成熟的地步”，并认为“驾驶员仍需时刻保持警惕”。这种对明确规则的违背，被用户视为一种根本性的“违约”，直接冲击了用户对系统“正直性”和意图的信任。
- 不可预测性是核心症结：** 无论是朱明宇的危险操作，还是蒋咸康的违规企图，其共同点在于用户无法从系统界面或行为逻辑中理解其决策原因。这种不透明性使用户无法建立对系统行为的稳定预期，从而产生一种深层的不安——因为他们不知道下一次类似的“黑箱”失误会在何时、何地、以何种方式发生。

3.2 机器逻辑与人类直觉的冲突：“人车较劲”现象

当系统的操作逻辑与人类驾驶员根深蒂固的本能、习惯和社会规范发生冲突时，信任便会受到侵蚀。这种冲突常常表现为一种物理或心理上的“人车较劲”。

- 控制权交接协议的冲突：** 多个经验丰富的驾驶者，如刘劲和罗昌霖，都指出了系统接管协议与驾驶本能的矛盾。系统要求“先踩刹车再动方向盘”来退出辅助驾驶，但这与人在紧急情况下优先转动方向盘避险的本能反应相悖。这导致了他们在尝试直接转动方向盘接管时，会感受到方向盘传来巨大的反抗力，刘劲将其描述为高速行驶时很危险的“人车较劲”。罗昌霖更详细地描述了这种对抗可能带来的风险：方向盘在对抗中突然卸力，可能导致驾驶员因用力过猛而造成过大的转向输入，引发二次危险。
- 驾驶风格与社会规范的冲突：** 系统的驾驶风格虽然在技术上可能是“正确”的，但却可能违背人类驾驶的社会规范和防御性驾驶原则。例如，当一台大巴车贴着车道线行驶时，系统依然“固执地”保持在车道正中心，这让刘劲和罗昌霖都感到非常紧张和危险。他们作为人类驾驶员，会下意识地主动远离，哪怕暂时偏离车道中心。系统的这种“绝对守规矩”反而被视为一种缺乏社会智能的“死板”行为。此外，罗昌霖感觉系统的刹车点比他自己的习惯要晚，尤其是在靠近弱势交通参与者（如自行车）时，这种不

协调感也引发了他的不安。

3.3 驾驶的“恐怖谷”：系统被视为缺乏经验的新手

对于期望系统能达到甚至超越自己水平的专家驾驶员而言，当系统暴露出决策能力上的局限时，它会被迅速降格为一个“新手司机”，从而引发信任危机。

- **缺乏前瞻性与战略规划：** 驾驶经验丰富的陈俊泽对系统的核心批评集中在其“缺乏前瞻性”和“思考能力”。他观察到，在路口拥堵时，系统不会像经验丰富的人类驾驶员那样，提前观察全局路况并选择最优车道，而是会遵循“非常短期的”、“很固定的算法”做出次优决策（如先直行再强行并线）。
- **从期望到失望的认知偏差：** 陈俊泽的失望源于系统表现与他内心高标准参照系之间的巨大差距。他期望的是一个具备“全局观”和“预判能力”的专家，但实际体验到的却是一个行为“机械”、“不科学”的“新手”。对专家而言，系统的可预测性（例如，它总会保持在车道中心）并不足以建立信任；他们要求的是更高层次的战略能力和决策智慧。当系统无法展现这种能力时，它不仅没能“省心”，反而带来了需要时刻提防和纠正的“奇奇怪怪的麻烦”，最终导致信任的瓦解。

综合与深度洞察，构建用户信任体验模型

4.1 驾驶者经验悖论

- **画像A：“新手受益者”**
 - **参照系：** 自身有限的驾驶技能与经验。
 - **信任构建逻辑：** 他们对系统的核心期望是安全保障和能力补偿。因此，当系统展现出精准的车距控制、可靠的安全预警、清晰的辅助信息显示（如侧后方影像），以及执行他们自己不擅长的高级操作（如自信地加塞）时，会产生强烈的“赋能感”¹。
 - **信任侵蚀因素：** 主要是灾难性的、明显的安全失误，或过于复杂、难以理解的操作界面。
 - **心智模型：** 将系统视为一位“专家副驾”或“驾驶教练”，其主要价值在于弥补自身不足，提升驾驶安全与舒适度。
- **画像B：“专家怀疑论者”**
 - **参照系：** 一个理想化的、经验丰富的人类驾驶员（通常是他们自己）。
 - **信任构建逻辑：** 他们对系统的期望是达到甚至超越专家的驾驶水平。只有当系统展现出类似人类的战略前瞻性、对复杂交通流的深刻理解和灵活的防御性驾驶策略时，信任才可能建立。
 - **信任侵蚀因素：** 任何“新手式”的错误，如缺乏预判、决策死板、对社会规范的无视，以及无法解释的“黑箱”行为，都会被视为系统能力不足的证据，从而严重侵蚀信任。

- **心智模型：** 将系统视为一个需要通过严格考核的“实习司机”。他们会持续、批判性地评估系统的每一个决策，任何不符合其专家模型的行为都会被标记为负分项。

这一悖论揭示了信任评估的主观性：同样一个“固执地保持车道居中”的行为，在“新手受益者”看来是可靠和守规矩的体现，而在“专家怀疑论者”眼中则是缺乏变通和危险的信号。因此，不存在一个普适的、能满足所有用户的“完美”系统，信任校准必须考虑用户画像的差异。

画像对比分析

特征维度	“新手受益者”	“专家怀疑论者”
驾驶经验	较少，对自身能力可能缺乏自信	丰富，对自身驾驶能力有较强自信
对系统的主要期望	安全保障、能力补偿、减轻驾驶负担	战略能力、决策效率、超越人类的驾驶智慧
关键信任构建因素	精准控制、主动安全、清晰的HMI、赋能感（完成高难度操作）	展现前瞻性、拟人化的驾驶策略、对复杂路况的优雅处理
关键信任侵蚀因素	严重的安全失误、复杂的交互	机械死板的决策、缺乏预判、人机控制权冲突、“黑箱”式错误
偏好的HMI特性	提供丰富、易于理解的辅助信息（如侧后方影像）	提供决策逻辑的解释，增强系统的透明度
代表性引言	“（系统的精确控制）是一个非常加分的点。”	“（系统）不具备人类的‘前瞻性’和‘思考能力’。”

4.2 完全自动化信任的心理壁垒

访谈数据揭示了两个深层次的心理壁垒，它们为用户完全信任并依赖自动化系统设定了难以

逾越的上限。这些壁垒并非技术缺陷，而是植根于人类的道德、责任感和认知模式。

- **责任悖论：** 蒋咸康的观点精辟地阐述了这一悖论。他表示，尽管他认为系统在技术上“表现得非常优秀”，没有任何失误，但在市中心等交通复杂的区域，他“更愿意相信我自己的判断”。其根本原因并非对技术的不信任，而是一种深刻的个人责任感——他担心的核心是“我会伤害到别人”，而非自己受伤。这表明，只要人类驾驶员在法律和道德上仍是事故的最终责任人，那么无论技术发展到何种程度，完全的控制权让渡都将面临巨大的心理障碍。信任的上限，最终由责任的归属所定义。
- **可解释性优于完美性：** 这一洞察来自于对沈雨凡和朱明宇体验的对比分析。沈雨凡提出了一个深刻的观点：他父亲开车可能会闯红灯，但他并不会因此完全不信任父亲，因为人的行为是可预期的，并且有明确的“责任归属”。相比之下，机器的“黑箱”算法使其在极端情况下的行为变得不可预测，且责任归属不明，这才是影响深层信任的根本。这与朱明宇的经历形成了呼应：一个表现近乎完美的系统，因为一次无法解释的危险行为，就足以让用户心存芥蒂。这揭示了一个关键的用户心理：用户可能更愿意接受一个行为逻辑清晰、可解释、错误模式可预期的系统（哪怕它偶尔会犯“类人”的错误），而不是一个99%时间完美无瑕，但有1%概率发生灾难性、神秘错误的系统。一个值得信赖的系统，首先必须是一个可被理解的系统。

旨在实现信任校准的行动

5.1 关于人机自动化信任本质的核心结论

1. **信任的非对称性：** 信任的建立是缓慢的、渐进的，通常需要多次正面体验的累积；然而，信任的侵蚀可以是瞬时的，一次严重的、特别是无法解释的负面事件，其破坏力远大于多次的成功表现。
2. **信任的多维校准：** 用户的信任并非单一维度的增减，而是在“正面信任”（对系统能力的信心）和“负面不信任”（对系统风险的警惕）两个轴向上独立进行校准。一次成功的试驾可能会同时提升这两个维度的得分，形成一种“能力认可但保持警惕”的复杂信任状态。
3. **用户经验的决定性作用：** 驾驶者的先前经验是其评估系统性能的核心滤镜，直接导致了“驾驶者经验悖论”。对新手而言，系统是赋能的专家；对专家而言，系统是待考的学徒。这一差异决定了信任校准的基准和方向。
4. **心理壁垒的刚性约束：** 超越技术性能，“责任悖论”和对“可解释性”的根本需求，为完全的自动化信任设定了坚实的上限。在人类驾驶员仍需承担最终责任的框架下，追求无条件的、绝对的信任既不现实，也不安全。

5.2 数据驱动的安全教育视频内容设计建议

根据实验方案的目标，以下建议旨在通过教育视频内容，促进健康、恰当的信任校准。

- 针对“新手受益者”——主题：赋能与引导

- **建立基础信心：** 重点展示系统在处理复杂路况（如拥堵跟车、高速变道）时的稳定表现和主动安全能力（如主动避让）。通过“眼见为实”的方式，快速建立用户对系统核心能力的信任。
- **功能教学与赋能：** 详细介绍HUD、侧后方影像等辅助功能如何帮助驾驶员弥补经验不足，提升态势感知能力，将系统定位为一个可靠的“教练”和“伙伴”。
- **管理过度信任风险：** 必须明确展示系统的运行设计域（ODD）边界。例如，通过模拟雨雪天气、车道线模糊等场景，清晰告知用户系统在这些情况下性能会受限，并会主动请求接管。这是防止“习得性粗心”的关键。
- **针对“专家怀疑论者”——主题：解释与协作**
 - **揭示“机器逻辑”：** 视频内容的核心不应是“展示能力”，而是“解释原因”。例如，用动画清晰解释系统为何坚持车道居中（为了最大化传感器视野和冗余空间），为何刹车点偏晚（可能基于能耗最优化的算法），以及为何接管协议是“先刹车”（这是最明确、无歧义的控制权交接信号）。目标是“解密黑箱”，将用户的抱怨从“它很蠢”转变为“我理解它为何这么做”。
 - **建立伙伴心智模型：** 强调驾驶员的角色从“操作者”转变为“监督者”。用生动的比喻，如“您现在是机长，系统是自动驾驶仪，但您需要随时准备接管”，来重塑人机关系，鼓励用户与系统协作，而非对抗。
 - **演示失败模式：** 在安全可控的环境下，模拟系统因传感器被遮挡或遭遇极端场景而失效的情况，并清晰演示正确的接管流程。这种“失败演示”能有效校准专家的过高期望，将他们的注意力从“挑错”转移到“准备接管”上。

5.3 系统设计与人机交互（HMI）的战略改进建议

- **开发自适应驾驶风格：** 为解决不同用户对驾驶风格的偏好冲突，建议智驾车企的工程团队探索可定制的驾驶“个性”模式。例如，提供“舒适”、“标准”、“运动”或“防御性”等选项。“防御性”模式可以自动增加跟车距离，并与相邻车辆保持更宽的横向距离，以满足“专家怀疑论者”的需求。
- **优化控制权交接的灵活性：** 重新评估“人车较劲”问题。在保留刹车作为主要接管方式的同时，可以考虑将一个超过特定阈值的高扭矩方向盘输入也作为一种有效的、平滑的接管信号。这能更好地适应人类在紧急情况下的本能反应，减少用户的恐慌和对抗感。
- **增强HMI的可解释性：** HMI设计应超越“显示状态”，进化到“解释意图”。当系统减速时，HUD或中控屏上可以短暂显示一个图标或文字，如“前方慢车”或“前方弯道”。当系统规划变道时，可以提前在屏幕上高亮显示其目标车道和空间。这种“意图可视化”能将系统的“思考过程”透明化，是打破“黑箱”、建立深度信任的根本途径。