

本次研究的文献调研，主要围绕着以下几个方面来展开：

## 1. 信任的概念和理论基础

信任在人机交互中扮演着核心角色，尤其在自动驾驶系统中，它不仅是用户接受技术的前提，更是影响驾驶行为和安全的关键心理构念。在自动驾驶中，“信任”可被定义为用户对系统能力、可靠性和意图的积极预期，进而愿意在不确定情境下依赖该系统。在自动驾驶语境下，信任不仅涉及对技术本身的信心，还包括对设计者、监管者以及整个社会技术系统的信赖（Henschke, 2020）<sup>1</sup>。

早期研究基于技术接受模型（TAM）和信任理论，指出信任是用户采纳自动驾驶技术的重要预测变量（Choi & Ji, 2015）<sup>2</sup>。此外，信任也被视为一种动态心理状态，随着系统表现、交互方式和环境变化而不断调整（Azevedo-Sa et al., 2021）<sup>3</sup>。从社会反应理论出发，当系统具备拟人化特征时，用户更容易将其视为具有意向性的“社会智能体”，从而增强信任（Niu et al., 2018）<sup>4</sup>。

## 2. 信任的影响因素

信任的形成受到多种因素的影响，可归纳为系统特性、交互设计、用户认知与情境因素四大类。

### （1）系统特性

智能驾驶系统的透明度、技术能力和情境管理能力被证实对信任有显著正向影响（Choi & Ji, 2015）。此外，系统的可靠性和可预测性也是关键因素，尤其是在系统出现错误时，信任会受到显著冲击。值得注意的是在低风险情境中发生的失误，可能比在高风险情境中的失误对信任的摧毁力度更强，因为它是更加直接地影响了用户对系统可靠性和可预测性的感知（Tan & Hao, 2023）<sup>5</sup>。

### （2）交互设计

交互设计可以通俗地理解为智能驾驶汽车向人传递各种信息的方式设计和呈现方式<sup>1</sup>这对信任构建至关重要。研究表明，拟人化的交互设计（如虚拟形象与人性化语音）能显著提升用户的信任水平（Niu et al., 2018）<sup>6</sup>。此外，解释类型（如无解释、简单解释、归因

---

<sup>1</sup> Henschke, A. (2020). Trust and resilient autonomous driving systems. *Ethics and Information Technology*, 22(1), 81–92. <https://doi.org/10.1007/s10676-019-09517-y>

<sup>2</sup> Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction*, 31(10), 692–702. <https://doi.org/10.1080/10447318.2015.1070549>

<sup>3</sup> Azevedo-Sa Hebert, Jayaraman, S. K., Esterwood, C. T., Jessie, Y. X., Robert, Lionel P., Jr, & Tilbury, D. M. (2021). Real-time estimation of drivers' trust in automated driving systems. *International Journal of Social Robotics*, 13(8), 1911–1927. doi:<https://doi.org/10.1007/s12369-020-00694-1>

<sup>4</sup> Niu, D. F., Terken, J., & Eggen, B. (2018). Anthropomorphizing information to enhance trust in autonomous vehicles. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 28(6), 352–359. <https://doi.org/10.1002/hfm.20745>

<sup>5</sup> Tan, H., & Hao, Y. Y. (2023). How does people's trust in automated vehicles change after automation errors occur? An empirical study on dynamic trust in automated driving. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 33(6), 449–463. <https://doi.org/10.1002/hfm.21001>

<sup>6</sup> Niu, D. F., Terken, J., & Eggen, B. (2018). Anthropomorphizing information to enhance trust in autonomous vehicles. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 28(6), 352–359. <https://doi.org/10.1002/hfm.20745>

解释)在不同风险情境下对信任的影响也不同(Ha et al., 2020)<sup>7</sup>。然而,并非信息越多越好,过多的反馈反而可能降低用户的信任感知(Mackay et al., 2020)<sup>8</sup>。

### (3) 用户认知

用户对风险的认知和情感反应是信任的重要预测变量。Eitrheim 等人(2024)<sup>9</sup>发现,风险认知影响对技术本身的信任,而情感担忧则影响对管理主体的信任。此外,用户在接触系统前的初始信任受到宣传方式和安全边界介绍的影响(Cui et al., 2025)<sup>10</sup>。

### (4) 情境因素

驾驶环境的风险水平调节了解释类型对信任的影响(Ha et al., 2020)。系统错误发生的情境风险也会影响信任恢复的难度,低风险情境下的错误更容易导致信任崩塌(Tan & Hao, 2023)。

## 3. 信任的程度变化和偏失问题

信任不是静态的,而是随着使用经验和系统表现动态变化的。Azevedo-Sa 等人(2021)提出了一种基于卡尔曼滤波的动态信任建模框架,通过眼动、系统使用时长和非驾驶任务表现等多模态数据实时估计信任水平。该研究揭示了信任在连续交互中的波动特性,并为构建“信任感知型”系统提供了技术支持。

信任校准问题包括信任不足与过度信任,两者都会影响系统的合理使用(Azevedo-Sa et al., 2021)。研究表明,系统错误会显著降低信任,尤其是在低风险情境中(Tan & Hao, 2023)。此外,信任的恢复需要系统在透明性、可靠性和沟通策略上有所改进(Eitrheim et al., 2024)。

## 4. 未来挑战和研究空白

目前,智能驾驶和对其信任心理的研究都尚处于发展阶段,尽管现有研究在信任构建、测量与校准方面取得了显著进展,但仍存在若干重要挑战与研究空白:

### (1) 弱势道路使用者的信任视角缺失

目前大多数研究聚焦于驾驶员对自动驾驶系统的信任,而忽视了行人、骑行者等弱势道路使用者的视角(Saleh et al., 2017)<sup>11</sup>。这些群体与自动驾驶车辆的互动方式、信任形成机制及其对系统行为的理解,尚未得到充分探索。未来的研究应更多关注车外用户的信任构建,以促进自动驾驶系统与所有道路使用者的安全共存。

<sup>7</sup> Ha, T., Kim, S., Seo, D., & Lee, S. (2020). Effects of explanation types and perceived risk on trust in autonomous vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, 73, 271–280. <https://doi.org/10.1016/j.trf.2020.06.021>

<sup>8</sup> Mackay, A., Fortes, I., Santos, C., Machado, D., Barbosa, P., Boas, V. V., Ferreira, J. P., Costa, N., Silva, C., & Sousa, E. (2020). The impact of autonomous vehicles' active feedback on trust. In P. M. Arezes (Ed.), *Advances in safety management and human factors* (Vol. 969, pp. 342–352). Springer. [https://doi.org/10.1007/978-3-030-20497-6\\_32](https://doi.org/10.1007/978-3-030-20497-6_32)

<sup>9</sup> Eitrheim, M. H. R., Nordfjærn, T., Log, M. M., & Torset, T. (2024). Risk perception and trust in autonomous vehicles: A case study of ground handling employees at an international airport. *Journal of Risk Research*. Advance online publication. <https://doi.org/10.1080/13669877.2024.2423202>

<sup>10</sup> Cui, Z., Yang, F., Zhou, H., Li, X., & Itoh, M. (2025). Using video introduction to mitigate affect-based initial trust in automation. *Japanese Psychological Research*. Advance online publication. <https://doi.org/10.1111/jpr.12586>

<sup>11</sup> Saleh, K., Hossny, M., & Nahavandi, S. (2017). Towards trusted autonomous vehicles from vulnerable road users perspective. In 2017 11th Annual IEEE International Systems Conference (SysCon) (pp. 537–543). IEEE.

### （2）车与车之间信息共享的信任机制

在车联网环境中，车辆之间需要共享感知数据以扩展视野，但如何评估所接收信息的可信度仍是一个关键问题（Guo et al., 2020）<sup>12</sup>。尽管已有研究提出基于 Dirichlet-Categorical 模型的信任评估框架，并利用 ESRGAN 提升图像分辨率以增强信任判断能力，但在实际道路环境中，信息真实性、数据安全性与隐私保护仍是未被充分解决的挑战。

### （3）对不同用户群体的差异化研究

现有研究样本多集中于普通驾驶员或特定工作场所员工（如 Eitrheim et al., 2024），缺乏对老年人、残障人士、儿童等特殊群体的信任研究。这些群体对自动驾驶技术的认知、信任形成路径和交互需求可能存在显著差异，未来研究应拓展用户多样性，以构建更具包容性的信任模型。

---

<sup>12</sup> Saleh, K., Hossny, M., & Nahavandi, S. (2017). Towards trusted autonomous vehicles from vulnerable road users perspective. In 2017 11th Annual IEEE International Systems Conference (SysCon) (pp. 537-543). IEEE.