

# Human–Autonomous Taxis Interactions

Subjects: Computer Science, Artificial Intelligence | Robotics | Computer Science, Interdisciplinary Applications

Contributor: Mahmoud Mastouri, Zied Bouyahia, Hedi Haddad, Leila Horchani, Nafaa Jabeur

With the increasing deployment of autonomous taxis in different cities around the world, recent studies have stressed the importance of developing new methods, models and tools for intuitive human–autonomous taxi interactions (HATIs). Street hailing is one example, where passengers would hail an autonomous taxi by simply waving a hand, exactly like they do for manned taxis.

Keywords: human–autonomous taxi interaction ; explicit and implicit street-hailing recognition ; deep learning

---

## 1. Introduction

Despite the uncertainty regarding their ability to deal with all real-world challenges, autonomous vehicle (AV) technologies have been receiving a lot of interest from industry, governmental authorities and academia <sup>[1]</sup>. Autonomous taxis (also called Robotaxis) are a recent example of such interest. Different companies are competing to launch their autonomous taxis, such as Waymo (Google) <sup>[2]</sup>, the Chinese Baidu <sup>[3]</sup>, Hyundai with its Level 4 Ioniq 5s <sup>[4]</sup>, NuTonomy (MIT) and the Japanese Tier IV, to mention a few. Robotaxis are not in the prototyping stage anymore, and different cities have regulated their usage on the roads, such as Seoul <sup>[5]</sup>, Las Vegas <sup>[6]</sup>, San Diego, and recently San Francisco, where California regulators gave Cruise's robotic taxi service the permission to start driverless rides <sup>[7]</sup>. Uber and Lyft, two ride e-hailing giants, have partnered with AV technology companies to launch their driverless taxi services <sup>[8]</sup>. The Chinese DeepRoute.ai is planning to start the mass production of Level 4 autonomous taxis in 2024, to be available for consumer purchase afterwards <sup>[9]</sup>. In academia, recent research works have concluded that, besides the safety issue, the acceptance of Robotaxis by users is hugely affected by the user experience <sup>[10][11][12]</sup>. Researchers have identified different autonomous taxi service stages, mainly calling, pick-up, traveling and drop-off stages, and have conducted experiments to study the user experience during these different stages. Flexibility during the pick-up and drop-off stages is one of the issues raised by users. Currently, the users use mobile applications to "call" an autonomous taxi, and after boarding they interact either through a display installed in the vehicles or using a messenger app on their smartphones. One of the problems raised by participants is the difficulty of identifying the specific taxi they called, especially if there are many autonomous taxis around <sup>[10]</sup>. Mutual identification using QR code was inconvenient <sup>[10]</sup>, and some participants suggested that other intuitive ways would be more interesting, such as hailing taxis through eye movement using Google glasses <sup>[12]</sup> and hailing gesture signals using smartwatches <sup>[12]</sup>. Participants also requested more flexibility with respect to the communication of the pick-up and drop-off locations to autonomous taxis. The current practice is to select pick-up and drop-off locations from a fixed list of stands, but users have the perception that taxis can be everywhere <sup>[10]</sup>, and they expect autonomous taxis to be like traditional manned taxis in this respect <sup>[11]</sup>.

These research findings show that scientists from different disciplines need to join their efforts to propose human–autonomous taxi interaction (HATI) models, frameworks and technologies that allow for developing efficient and intuitive solutions for the different service stages. Researchers address the case of one service stage that has been explored to a very limited extent, which is street hailing. The current state of the art reveals that only two works have studied taxi street hailing from an interaction perspective. The first work is the study of Anderson <sup>[13]</sup>, a sociologist who explored traditional taxi street hailing as a social interaction between drivers and hailers. Based on a survey that he distributed to a sample of taxi drivers, he found that the hailing gestures used by passengers largely vary in relation to the visual proximity of the hailer to the taxi and to the speed at which the taxi is passing <sup>[13]</sup>. The second work, also from a social background, is the conceptual model proposed by <sup>[14]</sup> to design their vision of what would be humanized social interactions between future autonomous taxis and passengers during street-hailing tasks. Both works model taxi street hailing as a sequence of visually driven interactions using gestures that differ according to the distance between autonomous taxis and hailing passengers. Consequently, researchers believe that computer-vision techniques are fundamental in the development of automated HATI, particularly for the recognition of street-hailing situations. First, they are not intrusive, as they do not require passengers to use any device or application. Second, they mimic how passengers communicate their requests to taxi drivers in the real world, through visual communication. Third, like for manned taxis, they allow passengers to hail

autonomous taxis everywhere without being limited to special stands <sup>[14]</sup>. Finally, they allow for better accessibility of the service, given that passengers who cannot/do not want to use mobile applications—for whatever reason—can still make their requests.

## **2. Human–Autonomous Vehicle Interaction (HAVI) and Body Gesture Recognition**

Since its introduction by W. Myron in 1991 <sup>[15]</sup>, gesture detection and recognition have been widely used for the implementation of a variety of human–machine interaction applications, especially in robotics. With the recent technological progress in autonomous vehicles, gestures have become an intuitive choice for the interaction between autonomous vehicles and humans <sup>[16][17][18]</sup>. From an application perspective, gesture recognition techniques have been used to support both indoor and outdoor human–autonomous vehicle interactions (HAVIs). Indoor interactions are those between the vehicles and the persons inside them (drivers or passengers). Most of the indoor HAVI applications focus on the detection of unsafe driver behavior, such as fatigue <sup>[19]</sup> and on vehicle control <sup>[20]</sup>. Outdoor interactions are those between autonomous vehicles and persons outside them, such as pedestrians. Most of the outdoor HAVI applications focus on the car-to-pedestrian interaction <sup>[21][22]</sup> and car-to-cyclist communication <sup>[23]</sup> for road safety purposes, but other applications have been explored, such as traffic control gestures recognition <sup>[24]</sup>, where traffic control officers can request an autonomous vehicle to stop or turn with specific hand gestures.

From a technological perspective, a lot of research work has been conducted on the recognition of body gestures from video data using computer-vision techniques. Skeleton-based recognition is one of the most widely used techniques <sup>[25]</sup>, both for static and dynamic gesture recognition. A variety of algorithms have been used. Traditional techniques include Gaussian mixture models <sup>[26]</sup>, recurrent neural network (RNN) with bidirectional long short-term memory (LSTM) cells <sup>[27]</sup>, deep learning <sup>[28]</sup> and CNNs <sup>[29]</sup>. The current state of the art for indoor and outdoor gesture recognition builds on deep neural networks. A recent review of hand gesture recognition techniques in general can be found in <sup>[30][31]</sup>.

## **3. Predicting Intentions of Pedestrians from 2D Scenes**

The ability of autonomous vehicles to detect pedestrians' road-crossing intentions is crucial for their safety. Approaches of pedestrian intention detection can be categorized into two major categories. The first category formalizes intention detection as a trajectory prediction problem. The second category considers pedestrian intention as a binary decision problem.

Several models and architectures have been developed and deployed, aiming at achieving a high-accuracy prediction of pedestrian intention using a binary classification approach. Unlike other methods, binary classification utilizes different tools and techniques depending on the data source and the data characteristics and features. For instance, the models based on RGB input use either 2D or 3D convolutions. In 2D convolution, a sliding filter is used along the height and width, and in 3D settings, the filter slides along the height, width and temporal depth. Using 2D convolutional networks, the information is propagated across time either via LSTMs or feature aggregation over time <sup>[32]</sup>. For instance, the authors in <sup>[33]</sup> proposed a two-stream architecture that takes as input a single excerpt from typical traffic scenes bounding an entity of interest (EoI) corresponding to the pedestrian. The EoI is processed by two independent CNNs producing two feature vectors that are then concatenated for classification. Authors in <sup>[34]</sup> presented an extension of these models by integrating LSTMs and 3D CNNs, and those in <sup>[35]</sup> did so by feeding many frames into the future and carrying out the classification using these frames.

Other methods that use the skeletal data extracted from the frames have been proposed. These methods directly operate on the skeleton of the pedestrians. The main advantage of these methods is that the data dimensions are significantly reduced. The yielded models are therefore less prone to overfitting <sup>[36]</sup>. Recently, a new method was proposed based on the individual keypoints in order to achieve the prediction of pedestrian intentions but from a single frame. Another method proposed by <sup>[37]</sup> exploits contextual features, such as the distance separating the pedestrian from the vehicle, his lateral motion, and his surroundings as well as the vehicle's velocity as input to a CRF (conditional random field). The purpose of this model is to predict in an early and accurate fashion pedestrian's crossing/not-crossing behavior in front of a vehicle.

## **4. Identification of Taxi Street-Hailing Behavior**

The topic of recognizing taxi street hailing has been studied by sociologists in order to explore how taxi drivers perceive and culturally interact with their environment, including passengers <sup>[38]</sup>. An interesting work is the work of Anderson, who studied gestures, in particular, as a communication channel between taxi drivers and passengers during hailing

interactions <sup>[13]</sup>. Based on a survey that he distributed to a group of taxi drivers in San Francisco, CA, USA, the researcher wanted to explore how taxi drivers evaluate street hails in terms of clarity and propriety. Clarity refers to “the ability of clearly recognizing a hailing behavior and distinguishing it from other one”, such as waving to a friend. Propriety refers to “the ability to identify if the passenger can be trusted morally and socially” so the taxi driver can decide either to accept the hailing request or not <sup>[13]</sup>. In the context of the work, it is the clarity aspect that is more relevant, and with this respect, Anderson's results are interesting. He found that the method of hailing adopted varies largely in relation to the visual proximity of the hailer to the taxi, and to the speed at which the vehicle is passing <sup>[13]</sup>. When the driver and the hailer are within range of eye contact, the hailer “can use any waving or beckoning gestures to communicate his intention, such as raising one's hand, standing on the curb while sticking arms out sideways into the view of the oncoming driver” <sup>[13]</sup>, etc. However, if the hailer and the driver are too far from each other, “hailers need to make clear that they are hailing a cab as opposed to waving to a friend, checking their watch, or making any number of other gestures which similarly involve one's arm” <sup>[13]</sup>. Taxi drivers who participated in the survey specified that the best and most clear gesture, in this case, is the “Statue of Liberty”, where the “hailer stands on the curb, facing oncoming traffic, and sticks the street-side arm stiffly out at an angle of about 100–135 degree” <sup>[13]</sup>. However, taxi drivers pointed out that there are many other hailing gestures, and they may even depend on the hailers' cultural backgrounds. Similarly, the conceptual model of human–taxi street-hailing interaction proposed in <sup>[14]</sup> assumed that, depending on the distance between taxis and passengers, different hailing gestures can be used, such as waving or nodding.

---

## References

1. Faisal, A.; Kamruzzaman, M.; Yigitcanlar, T.; Currie, G. Understanding autonomous vehicles. *J. Transp. Land Use* 2019, 12, 45–72.
2. McFarland, M. Waymo to Expand Robotaxi Service to Los Angeles. Available online: <https://edition.cnn.com/2022/10/19/business/waymo-los-angeles-rides/index.html> (accessed on 4 March 2023).
3. CBS NEWS. Robotaxis Are Taking over China's Roads. Here's How They Stack Up to the Old-Fashioned Version. Available online: <https://www.cbsnews.com/news/china-robotaxis-self-driving-cabs-taking-over-cbs-test-ride> (accessed on 11 December 2022).
4. Hope, G. Hyundai Launches Robotaxi Trial with Its Own AV Tech. Available online: <https://www.iotworldtoday.com/2022/06/13/hyundai-launches-robotaxi-trial-with-its-own-av-tech/> (accessed on 11 December 2022).
5. Yonhap News. S. Korea to Complete Preparations for Level 4 Autonomous Car by 2024: Minister. Available online: <https://en.yna.co.kr/view/AEN20230108002100320> (accessed on 4 March 2023).
6. Bellan, R. Uber and Motional Launch Robotaxi Service in Las Vegas. Available online: <https://techcrunch.com/2022/12/07/uber-and-motional-launch-robotaxi-service-in-las-vegas/> (accessed on 4 March 2023).
7. npr. Driverless Taxis Are Coming to the Streets of San Francisco. Available online: <https://www.npr.org/2022/06/03/1102922330/driverless-self-driving-taxis-san-francisco-gm-cruise> (accessed on 11 December 2022).
8. Bloomberg. Uber Launches Robotaxis But Driverless Fleet Is ‘Long Time’ Away. Available online: <https://europe.autonews.com/automakers/uber-launches-robotaxis-us> (accessed on 13 December 2022).
9. Cozzens, T. DeepRoute.ai Unveils Autonomous ‘Robotaxi’ Fleet. Available online: <https://www.gpsworld.com/deeproute-ai-unveils-autonomous-robotaxi-fleet/> (accessed on 11 December 2022).
10. Kim, S.; Chang, J.J.E.; Park, H.H.; Song, S.U.; Cha, C.B.; Kim, J.W.; Kang, N. Autonomous taxi service design and user experience. *Int. J. Hum.-Interact.* 2020, 36, 429–448.
11. Lee, S.; Yoo, S.; Kim, S.; Kim, E.; Kang, N. Effect of robo-taxi user experience on user acceptance: Field test data analysis. *Transp. Res. Rec.* 2022, 2676, 350–366.
12. Hallowell, M.; Large, D.; Harvey, C.; Briars, L.; Evans, J.; Coffey, M.; Burnett, G. Deriving UX Dimensions for Future Autonomous Taxi Interface Design. *J. Usability Stud.* 2022, 17, 140–163.
13. Anderson, D.N. The taxicab-hailing encounter: The politics of gesture in the interaction order. *Semiotica* 2014, 2014, 609–629.
14. Smith, T.; Vardhan, H.; Cherniavsky, L. *Humanising Autonomy: Where are We Going*; USTWO: London, UK, 2017.
15. Krueger, M.W. *Artificial Reality II*; Addison-Wesley: Boston, MA, USA, 1991.
16. Ohn-Bar, E.; Trivedi, M.M. Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations. *IEEE Trans. Intell. Transp. Syst.* 2014, 15, 2368–2377.

17. Rasouli, A.; Tsotsos, J.K. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Trans. Intell. Transp. Syst.* 2019, 21, 900–918.
18. Holzbock, A.; Tsaregorodtsev, A.; Dawoud, Y.; Dietmayer, K.; Belagiannis, V. A Spatio-Temporal Multilayer Perceptron for Gesture Recognition. *arXiv* 2022, arXiv:2204.11511.
19. Martin, M.; Roitberg, A.; Haurilet, M.; Horne, M.; Reiß, S.; Voit, M.; Stiefelhausen, R. Drive&act: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, Republic of Korea, 27–28 October 2019; pp. 2801–2810.
20. Meyer, R.; Graf von Spee, R.; Altendorf, E.; Flemisch, F.O. Gesture-based vehicle control in partially and highly automated driving for impaired and non-impaired vehicle operators: A pilot study. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction*, Las Vegas, NV, USA, 15–20 July 2018; pp. 216–227.
21. Rasouli, A.; Kotseruba, I.; Tsotsos, J.K. Towards social autonomous vehicles: Understanding pedestrian-driver interactions. In *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, USA, 4–7 November 2018; pp. 729–734.
22. Shaotran, E.; Cruz, J.J.; Reddi, V.J. Gesture Learning For Self-Driving Cars. In *Proceedings of the 2021 IEEE International Conference on Autonomous Systems (ICAS)*, Montréal, QC, Canada, 11–13 August 2021; pp. 1–5.
23. Hou, M.; Mahadevan, K.; Somanath, S.; Sharlin, E.; Oehlberg, L. Autonomous vehicle-cyclist interaction: Peril and promise. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, 25–30 April 2020; pp. 1–12.
24. Mishra, A.; Kim, J.; Cha, J.; Kim, D.; Kim, S. Authorized traffic controller hand gesture recognition for situation-aware autonomous driving. *Sensors* 2021, 21, 7914.
25. Li, J.; Li, B.; Gao, M. Skeleton-based Approaches based on Machine Vision: A Survey. *arXiv* 2020, arXiv:2012.12447.
26. De Smedt, Q.; Wannous, H.; Vandeborre, J.P. Skeleton-based dynamic hand gesture recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Las Vegas, NV, USA, 27–30 June 2016; p. 1–9.
27. Brás, A.; Simão, M.; Neto, P. Gesture Recognition from Skeleton Data for Intuitive Human-Machine Interaction. *arXiv* 2020, arXiv:2008.11497.
28. Chen, L.; Li, Y.; Liu, Y. Human body gesture recognition method based on deep learning. In *Proceedings of the 2020 Chinese Control And Decision Conference (CCDC)*, Hefei, China, 22–24 August 2020; pp. 587–591.
29. Nguyen, D.H.; Ly, T.N.; Truong, T.H.; Nguyen, D.D. Multi-column CNNs for skeleton based human gesture recognition. In *Proceedings of the 2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, Hue, Vietnam, 19–21 October 2017; pp. 179–184.
30. Yuanyuan, S.; Yunan, L.; Xiaolong, F.; Kaibin, M.; Qiguang, M. Review of dynamic gesture recognition. *Virtual Real. Intell. Hardw.* 2021, 3, 183–206.
31. Oudah, M.; Al-Naji, A.; Chahl, J. Hand gesture recognition based on computer vision: A review of techniques. *J. Imaging* 2020, 6, 73.
32. Karpathy, A.; Toderici, G.; Shetty, S.; Leung, T.; Sukthankar, R.; Fei-Fei, L. Large-scale Video Classification with Convolutional Neural Networks. In *Proceedings of the CVPR*, Columbus, OH, USA, 23–38 June 2014.
33. Rasouli, A.; Kotseruba, I.; Tsotsos, J.K. Understanding pedestrian behavior in complex traffic scenes. *IEEE Trans. Intell. Veh.* 2017, 3, 61–70.
34. Saleh, K.; Hossny, M.; Nahavandi, S. Real-time intent prediction of pedestrians for autonomous ground vehicles via spatio-temporal densenet. In *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, 20–24 May 2019; pp. 9704–9710.
35. Gujjar, P.; Vaughan, R. Classifying Pedestrian Actions In Advance Using Predicted Video Of Urban Driving Scenes. In *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, 20–24 May 2019; pp. 2097–2103.
36. Shahroudy, A.; Liu, J.; Ng, T.T.; Wang, G. Ntu rgb+d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 1010–1019.
37. Neogi, S.; Hoy, M.; Chaoqun, W.; Dauwels, J. Context based pedestrian intention prediction using factored latent dynamic conditional random fields. In *Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–8.
38. Ross, J.I. Taxi driving and street culture. In *Routledge Handbook of Street Culture*; Routledge: Abingdon, UK, 2020.

