Towards human-compatible autonomous car: A study of non-verbal Turing test in automated driving with affective transition modelling

Zhaoning Li
Sun Yat-sen University, Guangzhou, China

Qiaoli Jiang
Sun Yat-sen University, Guangzhou, China

Zhengming Wu
Guangzhou Intelligent Connected Vehicle Pilot Zone Operations Centre, Guangzhou, China

Anqi Liu
Johns Hopkins University, Baltimore, Maryland, United States

Haiyan Wu
University of Macau, Taipa, Macau, China

Miner Huang
Sun Yat-sen University, Guangzhou, China

Kai Huang
School of Data and Computer Science„ Guangzhou, China

Yixuan Ku
Sun Yat-sen University, Guangzhou, China

Abstract

We tested whether the AI driver could create a human-like ride experience for passengers based on 69 participants' feedback in a real-road scenario. We designed a ride experience-based version of the non-verbal Turing test for automated driving. Participants rode in autonomous cars as a passenger and judged whether the driver was human or AI. The AI driver failed to pass our test because passengers detected the AI driver above chance. We further investigated how human passengers ascribe humanness in our test. Based on Lewin's field theory, we advanced a computational model combining signal detection theory with pre-trained language models to predict passengers' humanness rating behaviour. We employed affective transition between pre-study baseline and corresponding post-stage emotions as the signal strength of our model. Results showed that the passengers' ascription of humanness would increase with the greater affective transition, indicating an important role of affective transition in passengers' ascription of humanness.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*