INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2019.1634317

Takeover Transition in Autonomous Vehicles: A YouTube Study

Feng Zhou [a], X. Jessie Yang [b], and Xin Zhang [b,c]

Q1 [a]Department of Industrial and Manufacturing Systems Engineering, University of Michigan, Dearborn, Michigan, USA; [b]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA; [c]Ross School of Business, University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT

Automated driving has many potential benefits, such as improving driving safety and reducing drivers' workload. However, from a human factors' perspective, one concern is that drivers become increasingly out of the control loop once they start to engage in non-driving-related tasks, which makes it difficult for the drivers to take over control in some situations. In the present study, we examined reviewers' comments of YouTube videos featuring takeover transitions on commercially available autonomous vehicles and categorized the comments into four topics: Non-driving related tasks, automation capability awareness, situation awareness, and warning effectiveness. Then we investigated people' opinions on the design of the takeover mechanism of commercially available autonomous vehicles using topic mining and sentiment analysis, and we found that 1) the topic of automation capability awareness received many more positive comments than both negative and neutral comments while the distributions of positive, negative, and neutral comments were fairly even in other topics and 2) people had extreme positive and negative opinions in non-driving related tasks than other topics. Finally, we discussed possible design recommendations in order to facilitate takeover transitions.

1. Introduction

Automated driving is becoming an engineering reality (Howard & Dai, 2014). According to the Society of Automotive Engineers (SAE), driving automation can be categorized into 6 levels with Level 0 being manual driving and Level 5 fully automated driving (SAE, 2016). Although there is still speculation on when Level 5 full automation will be realized eventually, Level 2 partial automation and near Level 3 conditional automation have been gradually implemented on vehicles (e.g., Volvo XC90, Tesla all models, Mercedes S class), and clear roadmaps toward Level 4 automation have been announced by major automotive manufacturers.

With automation capability advancing from Level 2 to Level 3 and above, human drivers will become increasingly out of the control loop in the dynamic driving task. With Level 2 automation, the driver is ultimately responsible for the driving task and has to actively monitor the road conditions. With Level 3 automation, by contrast, the autonomous vehicle is able to monitor the environment in some condition, which allows the driver to engage in other non-driving related tasks (Gold, Happee, & Bengler, 2017). If the autonomous vehicle reaches its system limit (e.g., automation failure, adverse weather, lane marks disappearance), however, the driver will be requested to resume control of the vehicle in a limited amount of time. According to Bainbridge (1983), takeover transition consists of two primary tasks, including monitoring and taking over control. Specifically, it involves the human driver receiving information, processing the information, and executing both lateral and longitudinal control of the vehicle.

Despite many potential benefits brought by automated driving, that drivers decoupled from the operational level of control makes it difficult for them to take over in situations with which the automation is not able to deal (Eriksson & Stanton, 2017; Gold, Körber, Lechner, & Bengler, 2016). With regard to this, in the present study, we aimed to 1) identify major human factors issues underlying takeover transitions by examining takeover events through YouTube videos and YouTube viewers' comments on these videos, to 2) investigate people's opinions over the design of takeover mechanism of commercially available autonomous vehicles, and 3) to suggest design improvements in order to facilitate takeover transitions. Compared with traditional experimental studies using driving simulators in this area, the present study identifies the human factors issues through videos filmed in a naturalistic driving environment. According to Barry and Eric (2017), Q2 such YouTube videos filmed by real users provided a remarkable source of data that focus on interaction and use of the autonomous vehicle technologies under different field circumstances. In addition, Siersdorfer, Chelaru, Nejdl, and Pedro (2010) conducted a large-scale (more than 6 million comments on 67,000 YouTube videos across 6 categories) in-depth study of YouTube comments and they demonstrated that YouTube comments were able to determine the community acceptance of particular events, topics,

CONTACT Feng Zhou fezhou@umich.edu Department of Industrial and Manufacturing Systems Engineering, University of Michigan, 4901 Evergreen Rd, Dearborn, MI 48128, USA

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

and content. Uryupina, Plank, Severyn, Rotondi, and Moschitti (2014) also presented a dataset of user comments on YouTube videos for sentiment analysis with regard to the video and the product discussed at the comment level. Asghar, Ahmad, Marwat, and Kundi (2015) summarized different techniques of sentiment analysis on YouTube comments. Therefore, we believe that by analyzing a large number of comments from the takeover transition videos (of Level 2/close to Level 3 autonomous vehicles) on YouTube in the market now, the issues identified can inspire possible design solutions that are helpful for the research and development of autonomous vehicles.

2. Related work

Research has shown that challenges have to be tackled for highly automated driving, especially for those associated with takeover transitions in order to secure the benefits brought by autonomous vehicles. For example, Casner, Hutchins, and Norman (2016) pointed out multiple challenges of partially automated driving, including the navigation system (e.g., brittleness and trust) and the driver warning system (e.g., complacency, nuisance alerts, and short time frames). They proposed that driving should be a shared task between humans and the vehicle with partial automation and a transparent interface that allows natural interaction between the driver and automation was needed. Borojeni et al. (2017) discussed how to design effective control transition interfaces in highly automated vehicles. By identifying the takeover procedures both from the driver to the automated vehicle and from the vehicle to the driver, they identified various challenges involved in the takeover transition period, such as tasks and actors involved, warning display modalities, urgency levels of takeover requests, situation awareness, and non-driving related tasks. McCall, McGee, Meschtscherjakov, Louveton, and Engel (2016) proposed a taxonomy of takeover situations in automated driving, including scheduled and non-scheduled situations. Compared to scheduled takeovers, non-scheduled takeovers tended to be more critical, especially in emergency situations. They also identified various challenges associated with takeover transitions, such as legal responsibility, situation awareness, drivers' driving skills, and in-vehicle contexts.

Among many factors, warning displays play an important role in taking drivers back into the control loop in the transition process. Usually, three types of displays are used, including visual, auditory, and tactile. For example, Naujoks, Mai, and Neukum (2014) examined the influence of urgency levels of takeover requests by using visual or visual and auditory warning displays. They found that when the takeover requests were presented in both visual and auditory forms, drivers had shorter hands-on time and better lateral vehicle control compared to those presented only in a visual form. Politis et al. (2018) examined four types of dialogue-based displays for takeover requests, i.e., a countdown-based system, a repetition-based system, a response-based system, and a multimodality-based system. Their experimental results showed that drivers liked displays with simplicity and, among the four tested displays, the countdown-based system resulted in the shortest takeover time with higher perceived usability and acceptance. Assuming that an increase in automation would increase safety, Hock, Kraus, Walch, Lang, and Baumann (2016) investigated different strategies to persuade drivers to engage in automation in order to improve safety. They compared three conditions, including a control condition (i.e., no feedback), and two treatment conditions (i.e., audio spoken feedback, and audio + a virtual co-driver). Their results showed that compared to the drivers in the control condition, those in the two treatment conditions had significantly longer automation engaged.

While many studies applied visual and/or auditory displays, tactile displays received increasingly more attention. Petermeijer et al. (2017) examined six different types of takeover requests in terms of auditory beeps, vibration in the driver seat, and their combinations. Furthermore, both non-directional and directional types of information were also provided (i.e., sounds produced from the right speaker or the left speaker and vibration of motors in the right or left column). They found that multimodal takeover requests produced shorter reaction time and higher self-reported ratings in terms of usefulness and satisfaction, while directional information did not result in a directional response, possibly due to the fact that overtaking on the right was not allowed on German highways. Borojeni et al. (2017) designed a shape-changing steering wheel that was able to convey contextual information during the takeover transition period in order to improve drivers' situation awareness. Although their results did not support the notion that haptic cues were able to assist drivers in decision making as such cues were not perceivable, they did reassure drivers of their decision making. The authors suggested that contextual haptic cues should be designed to contact the human body instead.

While it is allowed to perform non-driving related tasks in highly automated driving (e.g., SAE Level 3), its influence on takeover performance and quality, especially with a limited time budget, has to be investigated in order to inform design. Mok, Johns, Miller, and Ju (2017) investigated the influence of non-driving related tasks on takeover performance and they found that when drivers engaged in an active secondary task, i.e., playing a game on a tablet, they needed more time to take over control from automation. Radlmayr, Gold, Lorenz, Farid, and Bengler (2014) explored the influence of different types of non-driving related tasks on takeover performance, including the cognitive n-back task (a verbal-cognitive task that asks participants to verbally repeat a sequence of numbers with an offset of n steps) (Reimer, Mehler, Wang, & Coughlin, 2010) and the visual surrogate reference task (i.e., a visual-motoric task that asks participants to find the bigger circle among the smaller ones on a screen) (ISO14198, ISO14198, 2012). Compared to those in the cognitive n-back task, those in the visual surrogate reference task only had a significantly higher collision rate, while other measures (takeover time, longitudinal acceleration, and time to collision) had no significant differences. Such results indicated that both visual distraction and cognitive distraction led to worse takeover performance than that in manual driving.

Takeover time is also a critical issue in the takeover transition period and the optimal takeover time has been explored in different studies. Mok et al. (2017) found that participants with a 2-second lead time were not able to take over control safely while those with a 5- or 8- second lead time were able to take over control during a hazard situation successfully. Likewise, Gold, Damböck, Lorenz, and Bengler (2013) also found that those with a shorter takeover lead time (i.e., 5 seconds) did react faster than those with a longer takeover lead time (i.e., 7 seconds), but performed significantly worse. Kuehn, Vogelpohl, and Vollrath (2017) examined takeover times while drivers were performing non-driving related tasks. They found that when drivers were highly distracted, it took drivers 3–4 seconds to look at the road for the first time, and 6–7 seconds to have their feet on the pedals and hands on the steering wheel, 7–8 seconds to deactivate automation, and 12–15 seconds to glance at the mirror and speedometer to understand the driving situation. Compared to the participants assigned to drive manually, these highly distracted drivers were delayed up to 5 seconds to have the situation awareness needed to understand the driving situation.

3. Method

First, we conducted a YouTube video search to find videos containing transition moments of commercially available autonomous vehicles across different automotive manufacturers. YouTube.com contains a large number of third-party videos and those featuring autonomous driving experience are first hand, real-world footage by early adopters of this new technology. As such, we sifted through hundreds of videos and identified 20 YouTube videos and each contained at least one takeover transition across six automotive manufacturers. The collected videos had a combined number of over 140 minutes (Mean = 7.03 minutes and standard deviation = 10.41 minutes). They provided a satisfactory number of samples that were easily recognized by viewers to generate meaningful comments. Of these collected videos, we crawled 4464 comments and kept 3454 after a cleaning process by removing meaningless comments and those written in languages other than English.

Second, we conducted a systematic comparison between traditional experimental research in takeover transitions in automated driving and the comments provided by YouTube viewers after watching the takeover transitions in order to identify the human factors issues involved in this area. Traditional experimental studies explicitly listed the human factors issues in their studies, which guided us to identify the issues that were being discussed in the YouTube comments. However, we did recognize the differences between these two types of sources and the major ones were discussed in Section 4.

Third, in order to further automate this process, we conducted a topic mining analysis using fastText (Joulin, Grave, Bojanowski, & Mikolov, 2017) based on the human factors issues identified in step 2. fastText is a library created by Facebook and it is used to learn word representation and sentence classification. It was reported that its performance was on par with deep learning methods, but with extreme efficiency (Joulin et al., 2017). Figure 1 shows the model architecture of fastText. It first looks up the *N* word vectors (L2 short for Layer 2) in a preprocessed comment (L1), which are then averaged (O2 short for Operation 2) into a hidden comment representation (L3). The comment representation, shared among features and classes, is then fed into a linear classifier with rank constraint and a fast loss approximation (O3). Finally, the output is a softmax layer (L4) producing a probability distribution over labeled classes (O4), which are topic (L5) in this research. In the lookup step, fastText applies a hashing trick (Weinberger, Dasgupta, Langford, Smola, & Attenberg, 2009) (O1) that is a fast and space efficient way of vectorizing features by using the hash values of the features as direct indices of the vector. For ngram features, it makes use of sub-word information (i.e., character n-grams) so that wrongly spelled words (e.g., 'goood') can have a similar word vector to the correct one ('good'), which are often seen in social media. It also gives an option to use hierarchical softmax at the output layer when the number of classes is large, reducing the computational complexity from linear time to log time.

Fourth, based on each topic identified, we further performed a sentiment analysis using the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment tool (Hutto & Gilbert, 2014) to understand reviewers' opinions on each comment. VADER made use of an affective lexicon list and was built on syntactic rules, especially suitable for analyzing social media text data. The main advantage of

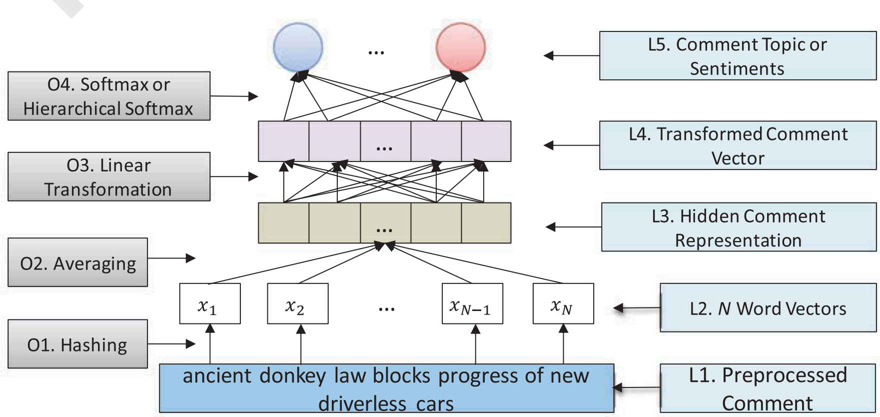

Figure 1. Model architecture of fastText, where L (in L1) represents layer and O (in O1) operation.

VADER is that it not only recognizes the polarity of the social media comment, but also quantifies its intensity on a − 1 to 1 scale with extremely good accuracy. In their study, Hutto and Gilbert (2014) showed that VADER (r = 0.881, F_1 = 0.96) performed as well as individual human raters (r = 0.888) in terms of correlation coefficient and outperformed individual human raters (F_1 = 0.84) in terms of F_1 accuracy when classifying tweets into positive, neutral, and negative.

Specifically, in order to calculate the intensity score of each comment, VADER summed the valence scores of each word in the lexicon and adjusted using the rules proposed and normalized between −1 (most extreme negative) and 1 (most extreme positive). We thus used scores between 0 and 1 as the intensity of positive comments and those between −1 and 0 as the intensity of negative comments. However, we used the absolute value of negative intensity scores so that both can be in the same scale from 0 and 1. The lexicons include those well-established word banks (e.g., LIWC, ANEW, and GI) as well as a full list of emoticons (e.g., ☺), sentiment-related acronyms, initialisms (e.g., LOL, BTW), and frequently used slangs (e.g., nah, meh). These lexicons were rated from −4 (extremely negative) to 4 (extremely positive) using Amazon Mechanical Turk with very good quality control (e.g., screening, training, selecting and data quality checking, evaluations and validation). For example, okay was rated as 0.9, good was rated as 1.9, and fantastic was rated as 2.6, while bad was rated as −2.5 and worst was rated −3.1. Five rules were used to modify the intensity of the sentiment, including punctuation (e.g., the service is great*!*), capitalization (e.g., the service is *GREAT*), degree modifiers (e.g., the service is *extremely* great), the contrastive conjunction (e.g., the service is great, *but* the food is not good), and the tri-gram before the sentiment-laden lexical features (e.g., the service isn't *really all that* great).

Fifth, based on the opinionated reviews, we finally discussed possible design improvements in order to facilitate the takeover transition process in highly automated driving. The goal for the design improvements is to ensure a smoother and better takeover transition, to reduce human error, to increase situation awareness, to reduce reaction time, and thus to achieve a safer and comfortable driving experience for drivers currently using Level 2 and near Level 3 automation systems.

4. Results

4.1. Difference between the YouTube study and literature

YouTube has a large number of users and thus the 20 selected transition videos attracted thousands of comments. The majority of the comments carried a sentiment tone to express their attitudes towards objects and things discussed in the video. Of all the comments collected, 63.6% of them were not neutral. After cleaning and removing the noisy data, 81.6% of them were not neutral. However, the text data provided by YouTube viewers were much less systematic and structured compared to academic publications. In addition, the data had lots of noise and the cleaning process removed 66.65% of the data. Nevertheless, the cleaned data tended to have interesting and *implicit* feedback knowledge about automated driving and takeover transitions. Unlike vehicle information provided by manufactures, which often highlighted vehicle performance using technical specifications, the comments generated by users assessed the autonomous vehicle in concrete use cases (e.g., highways, urban, and rural areas) with personal preferences and various user perspectives (Chen & Xie, 2008). In this sense, these user-generated comments have an important role for the designers. The human factors topics in the literature were often explicitly pointed out, which offered us overall guidance to sift through the YouTube comments to understand the topics (see Section 4.2 for detailed differences). Furthermore, with sentiment analysis and topic mining, we were able to extract the implicit issues with current autonomous vehicles and propose design recommendations.

4.2. Human factors topics

We watched the 20 YouTube videos and manually labeled 500 comments selected randomly from the 3454 comments made by the YouTube viewers. Using a grounded approach (Strauss & Juliet, 1994), we identified four topics pertinent to the Q7
takeover transitions in highly automated driving, including

- Non-driving related tasks,
- Automation capability awareness,
- Situation awareness,
- Warning effectiveness.

The topic of non-driving related tasks discusses what tasks, other than driving, drivers could do during automated driving and many viewers expressed their projected non-driving related tasks in the future automated driving. Examples include "*hollysh** thanks to Tesla I can text while driving, talk on phone, drink, smoke, have sex, sleep. The future is awesome!*", "*they will be doing full make up and texting …* ", "*Sit back, relax, text away, take a nap, let the car get you to your destination.*" Compared with the studies involving non-driving related tasks in the literature, the biggest difference was that these comments mentioned many more daily activities that could be done in the vehicle as evidenced by the example comments here. Previous studies often made use of standardized tasks, such as the cognitive n-back task, the visual surrogate reference task (e.g., Radlmayr et al., 2014), and/or a small number of non-driving related tasks (e.g., Mok et al., 2017). These tasks in the literature offered good experimental control in the laboratory environment and investigated their influence on takeover performance. However, a more comprehensive list of non-driving related tasks should be examined in order to better understand their influence on takeover performance.

Automation capability awareness refers to drivers recognizing whether the environment and the vehicle's operating conditions are suitable for turning or keeping the automated driving system on. Examples include "*I wonder how well it works in foggy or snowy conditions. How well will it brake on icy roads? Is that all programmed in the software? Amazing.*

Things like leaving yourself more room to brake and stop … " The naturalistic driving environment is very complicated and the drivers involved in the videos were not very sure if the vehicle was able to handle the driving task under different driving environments. This also aroused the curiosity of the viewers in their comments on automation capability in different driving conditions. The primary difference is that many published studies were conducted in driving simulators (e.g., Hock et al., 2016; Politis et al., 2018). The participants were usually told that they did not have to worry about anything when the automated mode was engaged and only were required to take over control from the automated driving when a takeover request was issued. This potentially involves one of the major human factors issues, i.e., trust in automation (Yang, Unhelkar, Li, & Shah, 2017). If the driver over-relies on the automation, under various driving conditions, the drivers would let automation be in control without taking over or ignoring the takeover request. If the driver distrusts automation, then the benefits claimed by autonomous vehicles would not be achieved. The simulation studies in the literature assumed that the participants would trust the vehicle to the largest possible extent whether the automated mode was engaged or the takeover request was issued, which tended to be not consistent with naturalistic automated driving in real life. Therefore, the risks of autonomous vehicles and the curiosities of the participants in automated driving might be missing in these simulation studies.

Situation awareness dicusses how to manage drivers' attention during automated driving so that the driver can maintain a high level of situation awareness of the driving environment and how to regain situation awareness effectively when the takeover request is initiated in order to take over control successfully. Examples include "*Man … I'm cringing at you folding your arms. Especially in traffic and on roads like that. There's a reason Elon said to have your hands on the wheel*" and "*Would you be actually able to watching videos and get back to road safely*". The YouTube videos available were mostly related to SAE Level 2 to near SAE Level 3 automated driving, and thus the drivers were still monitoring the driving environment from time to time although their hands might be off the steering wheel. Hence, many takeover transition scenarios tended to be less urgent. However, many studies (e.g., Borojeni, Wallbaum, Heuten, & Boll, 2017; Petermeijer et al., 2017) in the literature focused more on how to resume driver's situation awareness while they were immersed in non-driving related tasks both in emergency and non-emergency situations. In this aspect, these studies made use of different combinations of warning displays to explain the potential hazards, steering directions, and emergency levels, etc. Therefore, these studies explored many more possibilities than the existing autonomous vehicles in the collected YouTube videos.

Warning effectiveness describes the system's capability of alerting drivers for any impending dangers or conditions that require takeover. Examples include "*It does the beeping and sh** so if you were to fall asleep it would wake you up, and you would take control …* ", "*That's pretty annoying. The steer sign coming on a lot*", "*will it just alert you like it did when the lane markings disappeared!!*" The current warning displays mainly capitalized on the auditory and visual modalities while studies in the literature also investigated vibrotactile warnings and their combinations with other types of warnings during the takeover transition period (Petermeijer et al., 2017; Petermeijer, Hornberger, Ganotis, de Winter, & Bengler, 2017) and the results showed that vibrotactile warnings effectively helped drivers resume situation awareness and reduced takeover reaction time. However, vibrotactile warnings are still not widely deployed in the current autonomous vehicles as evidenced in the examples.

YouTube users also commented on a variety of other topics related to automated driving, including cybersecurity, joy of driving, government policies and laws. Others expressed their attitudes towards the posted videos or particular automotive manufactures. Example comments include, "*So, how long till someone hacks in to your car and makes it crash or commit crimes from a remote location?* (cybersecurity)", "*Driving is fun, it'll suck if it ever becomes completely autonomous* (joy of driving)", "*What if two auto-pilot cars do an accident, whose fault is it? The car company, the citizen, insurance, … ?* (policy and law)", "*Yes do more in depth videos. I enjoy watching your videos & that is a really nice car you have* (video)", and "*Tesla really awesome* (automotive manufacturer)".

In summary, we listed the major differences between the YouTube comments and previous studies identified in the literature in Table 1.

4.3. YouTube comment topic mining & sentiment analysis

We applied fastText with C++ to extract the four human factors topics in the YouTube comments. In order to do so, we manually labeled the selected 500 comments into five different categories, including the four topics mentioned above and one that did not fall into any of the four issues, named as "Others". These 500 comments were distributed as follows: 44 belonged to the category of non-driving related tasks, 92 belonged to the category of automation capability awareness, 80 belonged to the category of situation awareness, and 33 belonged to the category of warning effectiveness, and the remaining 251 belonged to the category of others. Using a 5-fold cross-validation strategy, we obtained 80% precision and 80% recall. We then trained the topic classification model using these 500 comments to classify the rest of the comments. During the predicting process, an incremental learning process was adopted, i.e., those predicted output with probabilities larger than certain threshold were used again to update the training model. The reason we adopted such a method was that the performance was enhanced with a relatively small number of labeled data and with extreme efficiency of fastText, the model was retrained and updated within seconds.

As a way for validation, Figure 2 depicts the incremental training process when we only randomly selected 400 labeled comments for training at the beginning and the rest 100 labeled comments for testing in each iteration. The left y-axis shows the number of comments left and the right y-axis shows the F_1 measure calculated as $2 \times Precision \times$

Table 1. Major differences between YouTube comments and typical studies identified in the literature.

Human Factor Topics	The focus of YouTube Comments	The focus of Typical Studies in Literature
Non-driving related tasks	A variety of daily activities was mentioned	Standardized non-driving related tasks or a small number of daily activities
Automation capability awareness	Unknown if the vehicle was able to handle the driving task in naturalistic driving environments	Known that the vehicle can handle the driving task or not in simulated driving environments
Situation awareness	How to resume situation awareness in SAE Level 2 to near SAE Level 3 automated driving, in which drivers monitor the driving situations continuously; Non-emergency takeover scenarios	How to resume situation awareness when drivers were involved in non-driving related tasks without paying attention to the driving conditions continuously; Both emergency and non-emergency takeover scenarios
Warning effectiveness	Warning displays with auditory and visual modalities and their combinations; The annoyance caused by the warning	Warning displays with auditory, visual, and vibrotactile modalities and their combinations; Warning modality and time budget

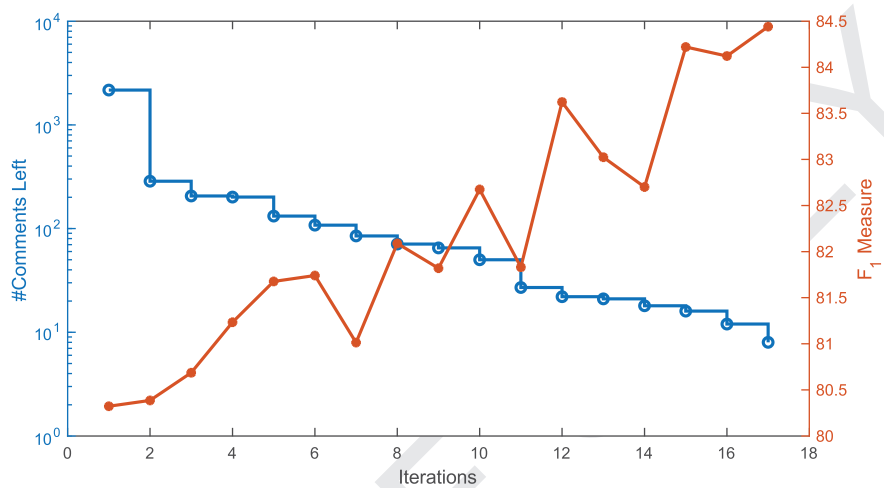

Figure 2. The incremental learning process involved in topic prediction. The left y-axis indicates the number of the comments left and the right y-axis indicates the F_1 measure as the number of iterations increases.

$Recall/(Precision + Recall)$ as the harmonic mean of precision and recall. First, we set the probability threshold as 0.8 for the first 10 iterations. Hence, those output results with prediction probabilities over 0.8 were added to the training data to update the model. Then for the 11th–15th iterations, we set the probability threshold as 0.6 and for the rest iterations, we set the probability threshold as 0.5. As the number of iterations increases, the number of comments left decreases and F_1 measure tends to increase. After 17 iterations, 32 comments were left, and we manually labeled them.

The predicted results were shown in Table 2 in terms of the numbers and the percentages of comments in each human factor topic. Among them, about two-thirds of them were categorized into others and 1152 comments were predicted for the four major topics identified above. This showed that the majority of the comments seemed to be noise in the YouTube comments which were discarded. Among the four topics, we found that the topic of automation capability awareness accounted for nearly 70% of the valid comments, whereas warning effectiveness only accounted for 4.34%. Both non-driving related tasks and situation awareness accounted for 17.53% and 9.03%, respectively. Such an uneven distribution tended to show YouTube viewers' emphasis was on whether the driving environment was suitable for the autonomous mode or not. These comments indicated the general public's caution and curiosity to test the boundary of the current technologies in automated driving in the market. Not many viewers paid attention to the importance of warning effectiveness. This seemed contradictory from what we found in the literature. The possible reason is that the takeover events involved in the YouTube videos were successful due to the fact that the driving conditions seemed to be less critical than emergency situations and the drivers tended to monitor the driving environment. However, the takeover scenarios in many studies (e.g., Merat, Jamson, Lai, & Carsten, 2012; Mok et al., 2017; Radlmayr et al., 2014) in the literature were critical and the drivers were involved in non-driving related tasks, which needed effective warning in terms of both a sufficient time budget and an effective presentation form to bring the drivers back to the control loop quickly.

Nevertheless, the topics of the comments did not directly show YouTube viewers' opinions and acceptance toward automated driving technologies. Hence, we performed sentiment analysis using the VADER sentiment analysis tools in Python for the 1152 relevant comments spanning across six different automotive manufacturers. The tool categorized the sentiment of a comment into positive, neutral, and negative. It also quantified sentiment intensity on a 0 to 1 scale. Among all the comments, 46.96% of them were positively evaluated, 18.40% were neutral, and 34.64% were negatively evaluated, indicating that almost half of the YouTube users tended to favor the current commercially available autonomous vehicles.

Table 2. Numbers and percentages of valid YouTube comments in each human factor topic.

Human Factor Topics	# Comments	%Comments
Non-driving related tasks	202	5.85%
Automation capability awareness	796	23.05%
Situation awareness	104	3.01%
Warning effectiveness	50	1.45%
Others	2302	66.65%
Grand Total	3454	100.00%

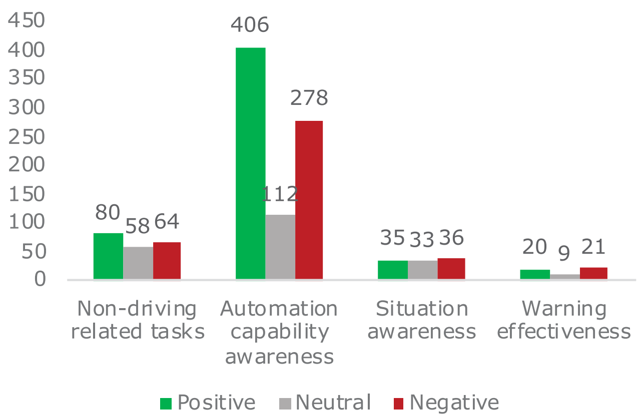

Figure 3. Numbers of positive, neutral and negative comments in four different topics.

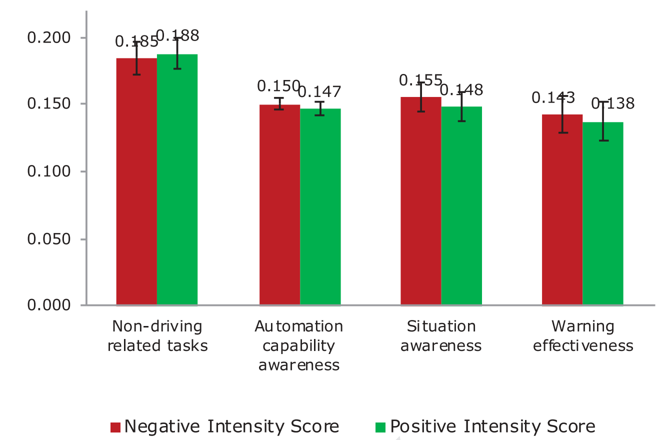

Figure 4. Comparisons of sentiment intensity scores among different topics.

Figure 3 summarizes the distributions of positive, negative, and neutral comments associated with each of the four human factors topics. Chi-squared tests were applied to test the distributions of positive, negative, and neutral comments across the four topics. We found that automation capability awareness was significantly different from situation awareness ($\chi^2(1) = 23.46, p < 0.001$) and non-driving related tasks ($\chi^2(1) = 25.07, p < 0.001$). The automation capability awareness topic had many more positive comments than both negative and neutral comments, while the distributions of positive, negative, and neutral comments were fairly even in the topics of situation awareness and non-driving related tasks.

Figure 4 shows the average intensity scores for the positive comments and the negative comments with their standard error. To compare the intensity scores among the four human factors topics, we applied one-way analysis of variance (ANOVA) for negative and positive intensity scores. For negative intensity scores, we found a significant difference ($F(3, 395) = 3.29, p < 0.05$). Post-hoc analysis was carried out with Bonferroni correction and the results revealed that the negative intensity score of the automation capability awareness topic was significantly smaller than that of non-driving related tasks ($p < 0.05$). This demonstrated that people had a stronger negative opinion in non-driving related tasks than that in automation capability awareness. For positive intensity scores, we also found a significant difference $F(3, 537) = 7.14, p < 0.001$. Post-hoc analysis was also conducted with Bonferroni correction and the results showed that the positive intensity score of non-driving related tasks was marginally, significantly, and significantly larger than that of situation awareness ($p < 0.10$), that of warning effectiveness ($p < 0.05$), and that of automation capability awareness ($p < 0.001$), respectively. This showed that people had a stronger positive opinion in non-driving related tasks than that in all other three topics. In other words, people had extreme positive and negative opinions in non-driving related tasks than other topics. Furthermore, we compared positive intensity scores with negative intensity scores within each topic. We only found that positive intensity score was marginally smaller than negative intensity score in automation capability awareness, $F(1, 682) = 2.65, p < 0.100$. It showed that people evaluated automation capability awareness marginally less positively than negatively.

Figure 5 shows the overall evaluation of six automotive manufacturers by aggregating their positive and negative intensity scores. Among them, Tesla received 859 comments out of 1152 comments, followed by Mercedes, while others including Volvo, BMW, Audi, and Honda had relatively small numbers of comments. Such a distribution tended to be consistent with the market share of autonomous vehicles. It seemed that the positive evaluation and negative evaluation of Tesla canceled out for all the four topics, while Volvo and Honda all received positive evaluation across four topics. Audi had all positive evaluation except situation awareness while BMW was the opposite. Mercedes was positively commented on non-driving related tasks and automation capability and negatively commented on situation awareness and warning effectiveness. However, due to the limited number of comments collected for Volvo, BMW, Audi, and Honda, it should be cautious to interpret such results.

5. Potential design improvement

5.1. Non-driving related tasks

The non-driving related tasks that the drivers were performing seem to have a major impact on the takeover performance and thus are widely examined in the literature (e.g., Mok et al., 2017; Radlmayr et al., 2014). However, the viewer's comments on this topic were primarily focused on non-driving related tasks in terms of what they could do during automated driving while maintaining the driving performance. For example, one commented, "*Women must be so excited, now they can do their makeup all the*

Topic	Tesla	Volvo	Mercedes	BMW	Audi	Honda
Non-driving related tasks	-0.010	0.089	0.302	-0.817	0.238	0.827
Automation capability awareness	0.076	0.132	0.258	-0.085	0.301	0.283
Situation awareness	-0.073	0.140	-0.083	0.672	-0.083	NA
Warning effectiveness	-0.038	0.448	-0.224	-0.296	0.584	0.505
#Sample	859	34	206	14	28	11

Figure 5. Evaluation of comments on different automotive manufacturers.

way to work" [0.4795], "*Now you can text until you kill yourself!*" [-0.7177], and "*Nice. I can finally watch my YouTube vids while driving woot woot:D*" [0.6808], where the numbers inside of [] indicated the predicted sentiment intensity scores. Other viewers expressed their interests in automated driving under the influence of drinking and other substances (e.g., drugs) as well as sleeping. Example comments include "*I just need a self driving car so I can roll blunts while driving*" [0] and "*Stuck in traffic on the interstate that would be great. I could just recline the seat and go to sleep*" [0.6249]. However, previous studies have shown that under the current automation capabilities (e.g., SAE Level 2 and near 3 automation), such behaviors are still dangerous. For example, Wiedemann et al. (2018) showed that drivers with 0.08% blood alcohol content had significantly worse takeover performance than those without any alcohol or with a level of 0.05%. Vogelpohl, Kühn, Hummel, and Vollrath (2018) showed that sleep-related fatigue made those in automated driving respond to takeover requests much slower those in manual driving. Nevertheless, studies did show that automated driving without any non-driving related tasks tended to induce disengagement related fatigue (Gold et al., 2017; Vogelpohl et al., 2018), such as boredom. In this sense, drivers are recommended to engage in certain types of non-driving related tasks in order to keep them at a certain level of alertness. In addition, regulations and policies still need to be made in order to make sure which non-driving related tasks should and should not be allowed in the current Level 2 – Level 3 autonomous vehicles.

In addition, one good way to both engage users with non-driving related tasks and to maintain driving performance is to integrate the warning system into the non-driving related tasks or design a lockout system. For example, warnings can be presented to the drivers' tablets or smartphones on which they are playing games, reading books, sending/editing emails, etc. Melcher, Rauh, Diederichs, Widlroither, and Bauer (2015) showed that such a method could overcome the disadvantages of visual displays and reduce their reaction time to takeover requests. A lockout is often initiated by the system at the moment of the takeover request that blocks drivers' non-driving related tasks, which are often provided by the vehicle via its infotainment system. Such a lockout is also able to bring drivers back to the control loop quickly and Wandtner, Schömig, and Schmidt (2018) found that a task lockout had significant advantages for reaction times, and was highly acceptable during the takeover transition period.

5.2. Automation capability awareness

In spite of many positively reviewed comments on this topic, viewers' major concerns were that human drivers did not know the boundary of the autonomous vehicles. Example comments include "*Will the car do an emergency stop and lane change? For example, someone pulls out in front of you, will the car dodge the other vehicle or is that not implemented yet?* [−0.6322]", "*I wonder how well it works in foggy or snowy conditions. How well will it brake on icy roads? Is that all programmed in the software? Amazing. Things like leaving yourself more room to brake and stop …* [0.9089]", "*What if one of the censors starts to malfunction though?* [−0.2960]", "*looks good in "ideal" conditions but theres no way id let a computer control my life in heavy rain or snow/ice* [−0.2144]", and "*It's so cool that you can actually see the car learning to an extent the way to drive on a road like that!* [0.6581]".

Although human drivers will become increasingly out of the control loop as the automated driving technology develops, it plays an important role in making the passengers and drivers aware of the automation capability in order to make the takeover transition decision correctly and promptly. The key for further improvement, therefore, is to help human drivers better understand the automation system and its interaction with the dynamic external environment. One possible improvement is to increase automation transparency so as to help the driver build a proper level of trust during the takeover transition process (Lee & See, 2004). With increasing levels of automation introduced and built into autonomous vehicles, the system must clearly indicate what level of automation is currently engaged so that the driver knows his/her responsibilities and roles. For example, in SAE Level 2 automation, the driver needs to actively monitor the driving condition while in SAE Level 3 automation, the system is actively monitoring the driving condition and the driver is allowed to perform non-driving related tasks. Another way to understand the system boundary or improve transparency is to offer explanations for each takeover request. For example, Körber, Prasch, and Bengler (2018) showed that post-hoc explanations of takeover requests improved their understanding of the system. In such a way, the takeover requests can be predictable in similar driving situations to those the driver has experienced before. The increased automation transparency (i.e., its intent, performance, future plans, and reasoning process) will then help users develop an accurate mental model of the automaton and its behavior, which leads to a higher level of trust and acceptance.

5.3. Situation awareness

Examining the comments, we found that YouTube viewers were very concerned about how to quickly bring back human drivers' attention to the driving task and manage the takeover request successfully during the transition period. Example comments consisted of "*I think I prefer to actually pay attention when I drive*" [- 0.1027], "*Man … I'm cringing at you folding your arms. Especially in traffic and on roads like that. There's a reason Elon said to have your hands on the wheel*" [- 0.3612], and "*Despite Elon Musk saying that people should ABSOLUTELY keep their hands on the wheel, I suspect that a lot of Tesla Owners will be like this numb nut and take their hands OFF the wheel. Note to other motorists: AVOID all Teslas on the road. There might be an idiot behind the wheel, like the one in this video*". [- 0.7338]. In order to successfully take over control from automation, the driver needs to have a good sense of situation awareness. Regarding this, many assistive technologies may be helpful. For example, augmented reality (AR) has been used in takeover transitions to improve drivers' situation awareness. Lorenz, Kerschbaum, and Schumann (2014) designed two AR assistive concepts, i.e., AR green and AR red. The AR green concept highlighted a passage for users to safely navigate during the takeover transition period while the AR red one showed a corridor to be avoided. Their results showed no significant differences between the two concepts in terms of takeover performance, but the AR green concept did help drivers to use the brake more and take a consistent path to guide them through the takeover transition period. Another possible way is to offer environment cues directly to the driver during the takeover transition period. For example, Wright et al. (2017) made use of auditory-based environment cues to help improve drivers situation awareness. Among the four types of assistive cues, i.e., an environment cue (e.g., work zone ahead), a threat cue (e.g., scan for workers), a combined cue (work zone ahead; scan for workers), and a general cue (take over control), the environment cue had better takeover performance than others. This was possibly due to the fact that the environment cue provided the specific location of the hazard while the combined cue cost extra time and resources for the driver to process.

Due to the capabilities of current autonomous vehicles (SAE Level 2 to near SAE Level 3), many concerned that they were not able to monitor the car all the time when automated driving was engaged. For example, one commented that "*i think we will be super tired from the constant monitoring of the car with the tension that it may screw it up. It will feel like we are letting a kid drive and have to constantly watch them*" [−0.2023]. In order to make sure that the driver has enough time to resume situation awareness, one possible way is to monitor drivers' state using their facial expressions, physiological data, and head orientations. Such a system is able to tell whether the driver is highly distracted or actively monitoring the driving situation. While the driver is highly distracted (by non-driving related tasks), the system should provide early warnings and give extra time for the driver to prepare, understand, and predict the driving environment.

5.4. *Warning effectiveness*

Three important factors are identified in the literature, including warning experience, time budget, and alert types. Warning experience refers to how annoyed the alert is when it goes off at the takeover request. Time budget refers to the lead-time of the takeover request, or how much time the request should be provided to the driver ahead of the danger. Alert type involves the presentations and modalities of the warning, i.e., whether it is visual, auditory, tactile, or a combination of them with various forms of presentations. Unlike the studies in the literature which emphasized on the latter two important factors, including time budget and alert type, the YouTube comments mainly focused on the annoyance caused by the warning. Example comments included "*It does the beeping and sh** so if you were to fall asleep it would wake you up, and you would take control …* [−0.03]", "*That's pretty annoying. The steer sign coming on a lot*" [−0.34], "*If you don't want to have to touch the steering wheel every 6 seconds, buy a Neodriven and install Comma ai OpenPilot on it for free. I just got one for my Civic w/Sensing and its 100x better at self-driving than Honda's system …* [0.7165]", and "*The autopilot alarm keeps telling sounding alerting the driver to the fact that it doesn't trust itself (difficulty reading road markings?) so why is this driver trusting it?*" [−0.4141]. These responses tended to be consistent with the capabilities of current autonomous vehicles on the road which required the driver to monitor the driving constantly. Whenever the driver did not put their hands on the steering wheel for a certain time, the system may warn the driver although it may not necessarily ask the driver to take over control from automated driving. However, the warning in the form of visual (e.g., red takeover sign flashing) and auditory (e.g., beeping) displays tended to be annoying.

First, visual displays seemed not enough and many drivers tended to miss visual information while they were engaged in non-driving related tasks (de Waard, van der Hulst, Hoedemaeker, & Brookhuis, 1999; Politis, Brewster, & Pollick, 2015). Therefore, auditory and tactile displays should also be explored due to their advantages over visual displays in that they are gaze-free. Second, for auditory displays, countdown-based warnings (Politis et al., 2018) with clear spatial or environmental cues (Wright et al., 2017) are preferable, while tactile displays should be placed in contact with drivers back in a repeated pattern as human backs tend to be more sensitive to vibrations than human hips and thighs and repeated vibrations are effective than directional ones (Wan & Wu, 2017). Under critical situations, multiple modalities of signals should be combined not only to increase the urgency level to improve drivers' reaction time but also to provide redundancy. Such integration of visual, auditory, and tactile displays are preferred by a large number of participants (Bazilinskyy & de Winter, 2015).

Although the comments did not mention much about the time budget needed to warn the driver at the takeover transition period, it is extremely important that enough time is provided, especially when the driver is engaged in non-driving related tasks or even falls asleep. Takeover transitions happen in a relatively short timeframe and thus the lead time is directly associated with the outcome of the transition. We recommend that the takeover lead time should be at least between 6.5 and 8 seconds (Clark & Feng, 2015; Eriksson & Stanton, 2017; Gold et al., 2013; Mok et al., 2015; Mok et al., Q10
2017; Mok et al., 2015), which lead to better takeover performance, quality, and comfort in various scenarios. Although Q11
a shorter time budget (e.g., 5 seconds) can result in shorter

reaction time and shorter takeover time, a longer time budget tends to facilitate drivers with a higher level of trust and ease (e.g., Mok et al., 2015). From the design perspective, it should accommodate the 5th to the 95th percentiles of the population as is in anthropometrics and the average lead time would exclude a large number of drivers (Eriksson & Stanton, 2017). As a summary, providing ample lead time allows more time to react and using multiple modalities of stimuli (e.g. visual, auditory, hepatic) increases the driver's efficiency and effectiveness of processing the alerts.

Finally, we summarized the major design improvement in Table 3 for the four topics identified in YouTube comments.

6. Limitations and future work

6.1. YouTube comments

YouTube is one of the most popular video sharing websites and videos with autonomous vehicles were also received a large number of audiences. Despite a large number of comments on autonomous vehicles and takeover events, a large portion of the comments were not related to the issues involved in this study (e.g., comments on the video quality rather than content, funny but irrelevant statements), and many of them vaguely discussed part of the issues. As seen from Table 2, over 66% of the data were categorized as others and thus were not included in the analysis. Therefore, a systematic cleaning and refining process is necessary before the data can be analyzed. The videos provided were mostly related to SAE Level 2 to near SAE Level 3 automated driving, and thus the comments and the issues derived from these comments tended to target these vehicles. However, by comparing and contrasting with the studies in the literature, we aimed to complement each other and proposed design recommendation accordingly.

6.2. Topic mining and sentiment analysis

Although the fastText method used for topic mining of the takeover transition-related comments had over 80% accuracy in terms of precision and recall. The training set only had 500 samples, which tended to be small. In the future, we plan to crawl a large number of related comments and manually label them to better train the models in order to improve the accuracy of this model. Moreover, the VADER model is an unsupervised model and generally performs well for social media text data. However, it was not created specifically for the YouTube comments on autonomous vehicles and the takeover transition-related topics. The model can be tailored to such type of comments to further improve its accuracy in the future. Furthermore, viewers often made different types of comments on these YouTube videos, such as ideas, concerns, and requirements related to the takeover events and/or the vehicles involved. Future work can consider developing machine learning models to automatically classify such types of comments to better understand how we can further improve the current design of autonomous vehicles.

6.3. Results and design recommendations

We analyzed the issues of currently available autonomous vehicles which were mainly SAE Level 2 to near SAE Level 3 automation from the YouTube comments. It should be cautious to generalize such results to other levels of automation. However, the design improvements are helpful for automotive manufacturers in general as the takeover transitions can happen in SAE Level 3 and Level 4 automated driving. Compared to the well-controlled lab studies with structured objectives, such a YouTube study can understand the general public's concerns and acceptance towards autonomous vehicles in a bigger picture. However, in order to have a systematic comparison between the lab studies and YouTube studies, a thorough literature review is needed in order to identify those most frequently studied human factor topics involved in the takeover transition process, which will further guide us to examine the issues in YouTube comments. On the contrary, those in the YouTube comments also can inspire new studies in the lab in the future.

7. Conclusions

In this paper, we identified four major human factors topics from YouTube videos and their comments of takeover events

Table 3. Major design improvements suggested for the four identified human factor topics.

Human Factor Topics	Major Suggested Design Improvements
Non-driving related tasks	(1) Clarification and regulations about what types of non-driving related tasks are allowed and not allowed in the vehicle (2) A lockout system that blocks drivers' non-driving related tasks
Automation capability awareness	(1) Increase automation transparency to help the driver build a proper level of trust (e.g., indicating the current level of automation) (2) Offer explanations for each takeover request
Situation awareness	(1) Offer assistive technologies (e.g., AR and directional warning) (2) Provide a driver monitoring system to tell when drivers are distracted
Warning effectiveness	(1) Combine visual warning with other types of warning (e.g., auditory and vibrotactile displays) (2) Combine multiple modalities of warning to indicate warning urgency (3) Provide takeover lead time at least between 6.5 and 8 seconds (4) Accommodate the 5th to the 95th percentiles of the population in design warning displays

in commercially available autonomous vehicles, including 1) non-driving related tasks, automation capability awareness, situation awareness, and warning effectiveness. We then investigated viewers' positive and negative opinions on each topic using topic mining and sentiment analysis. We found that YouTube viewers commented on automation capability awareness the most among the four topics, and had similar numbers of positive, neutral, and negative comments on other topics. Furthermore, viewers went to two extremes in terms of sentiment intensity scores of non-driving related tasks involved in automated driving, compared with other topics. By identifying the differences between the comments on the collected YouTube videos and experiment studies in the literature, we finally suggested possible design recommendations in order to improve takeover performance of automated driving.

ORCID

Feng Zhou http://orcid.org/0000-0001-6123-073X

References

Asghar, M. Z., Ahmad, S., Marwat, A., & Kundi, F. M. (2015). Sentiment analysis on YouTube: A brief survey. *MAGNT Research Report*, *3*(1), 1250–1257.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*, 775–779. doi:10.1016/0005-1098(83)90046-8

Bazilinskyy, P., & de Winter, J. (2015). Auditory interfaces in automated driving: An international survey. *PeerJ Computer Science*, *1*, e13. doi:10.7717/peerj-cs.13

Q13 Borojeni, S. S., Meschtscherjakov, A., Mirnig, A. G., Boll, S., Naujoks, F., Politis, I., & Alverez, I. (2017). *Control transition workshop: Handover and takeover procedures in highly automated driving*. Paper presented at the Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct, Oldenburg, Germany.

Borojeni, S. S., Wallbaum, T., Heuten, W., & Boll, S. (2017). *Comparing shape-changing and vibro-tactile steering wheels for take-over requests in highly automated driving*. Paper presented at the Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany.

Q14 Brown, B., & Laurier, E. (2017). *The trouble with autopilots: Assisted and autonomous driving on the social road*. Paper presented at the Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA.

Casner, S. M., Hutchins, E. L., & Norman, D. (2016). The challenges of partially automated driving. *Communications of the ACM*, *59*(5), 70–77. doi:10.1145/2830565

Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, *54*(3), 477–491. doi:10.1287/mnsc.1070.0810

Clark, H., & Feng, J. (2015). Semi-autonomous vehicles: Examining driver performance during the take-over. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 781–785. doi:10.1177/1541931215591241

de Waard, D., van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver behavior in an emergency situation in the automated highway system. *Transportation Human Factors*, *1*(1), 67–82. doi:10.1207/sthf0101_7

Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*, *59*(4), 689–705. doi:10.1177/0018720816685832

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1938–1942. doi:10.1177/1541931213571433

Gold, C., Happee, R., & Bengler, K. (2017). Modeling take-over performance in level 3 conditionally automated vehicles. *Accident Analysis & Prevention*. doi:10.1016/j.aap.2017.11.009

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density. *Human Factors*, *58*(4), 642–652. doi:10.1177/0018720816634226

Hock, P., Kraus, J., Walch, M., Lang, N., & Baumann, M. (2016). *Elaborating feedback strategies for maintaining automation in highly automated driving*. Paper presented at the Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA.

Howard, D., & Dai, D. (2014). *Public perceptions of self-driving cars: The case of Berkeley, California*. Paper presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, DC.

Hutto, C. J., & Gilbert, E. E. (2014). *VADER: A parsimonious rule-based model for sentiment analysis of social media text*. Paper presented at the Eighth International Conference on Weblogs and Social Media, Ann Arbor, MI.

ISO14198. (2012). Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems. *ISO/TS 14198:2012(en)*. Q15

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). *Bag of tricks for efficient text classification*. Paper presented at the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain.

Körber, M., Prasch, L., & Bengler, K. (2018). Why do i have to drive now? Post hoc explanations of takeover requests. *Human Factors*, *60*(3), 305–323. doi:10.1177/0018720817747730

Kuehn, M., Vogelpohl, T., & Vollrath, M. (2017). *Takeover times in highly automated driving (level 3)*. Paper presented at the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration. Q16

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. doi:10.1518/hfes.46.1.50_30392

Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving:How does augmented reality support the driver to get back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1681–1685. doi:10.1177/1541931214581351

McCall, R., McGee, F., Meschtscherjakov, A., Louveton, N., & Engel, T. (2016). *Towards a taxonomy of autonomous vehicle handover situations*. Paper presented at the Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA.

Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-over requests for automated driving. *Procedia Manufacturing*, *3*, 2867–2873. doi:10.1016/j.promfg.2015.07.788

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors*, *54*(5), 762–771. doi:10.1177/0018720812442087

Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015, September 15–18). *Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles*. Paper presented at the 2015 IEEE 18th International Conference on Intelligent Transportation Systems. Q17

Mok, B., Johns, M., Miller, D., & Ju, W. (2017). *Tunneled in: Drivers with active secondary tasks need more time to transition from automation*. Paper presented at the Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA.

Mok, B. K. J., Johns, M., Lee, K. J., Ive, H. P., Miller, D., & Ju, W. (2015, June 28–July 1). *Timing of unstructured transitions of control in* *automated driving*. Paper presented at the 2015 IEEE Intelligent Q18 Vehicles Symposium (IV).

Naujoks, F., Mai, C., & Neukum, A. (2014). *The effect of urgency of take-over requests during highly automated driving under distraction conditions*. Paper presented at the 5th international conference on applied human factors and ergonomics, Kraków, Poland.

Petermeijer, S., Bazilinskyy, P., Bengler, K., & de Winter, J. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics*, *62*, 204–215. Q19 doi:10.1016/j.apergo.2017.02.023

Petermeijer, S. M., Hornberger, P., Ganotis, I., de Winter, J. C. F., & Bengler, K. J. (2017). *The design of a vibrotactile seat for conveying take-over requests in automated driving*. Paper presented at the International Conference on Applied Human Factors and Q20 Ergonomics, Cham.

Politis, I., Brewster, S., & Pollick, F. (2015). *Language-based multimodal displays for the handover of control in autonomous cars*. Paper presented at the Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Nottingham, United Kingdom.

Politis, I., Langdon, P., Adebayo, D., Bradley, M., Clarkson, P. J., Skrypchuk, L., … Stanton, N. (2018). *An evaluation of inclusive dialogue-based interfaces for the takeover of control in autonomous cars*. Paper presented at the 23rd International Conference on Intelligent User Interfaces, Tokyo, Japan.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 2063–2067. doi:10.1177/1541931214581434

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2010). The impact of systematic variation of cognitive demand on drivers' visual attention across multiple age groups. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(24), 2052–2055. doi:10.1177/154193121005402407

SAE. (2016). SAE on-road automated vehicle standards committee taxonomy and definitions for terms related to on-road motor vehicle automated Q21 driving systems. *Technical Report J3016_201609, SAE International*.

Siersdorfer, S., Chelaru, S., Nejdl, W., & Pedro, J. S. (2010). *How useful are your comments? Analyzing and predicting youtube comments and comment ratings*. Paper presented at the Proceedings of the 19th international conference on World wide web, Raleigh, North Carolina, USA.

Uryupina, O., Plank, B., Severyn, A., Rotondi, A., & Moschitti, A. (2014). *SenTube: A corpus for sentiment analysis on YouTube social media*. Paper presented at the Ninth International Conference on Language Resources and Evaluation, Reykjavik, Iceland.

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2018). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention*. doi:10.1016/j.aap.2018.03.013

Wan, J., & Wu, C. (2017). The effects of vibration patterns of take-over request and non-driving tasks on taking-over control of automated vehicles. *International Journal of Human–Computer Interaction*, 1–12. doi:10.1080/10447318.2017.1404778

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of non-driving related task modalities on takeover performance in highly automated driving. *Human Factors*, *60*(6), 870–881. doi:10.1177/0018720818768199

Weinberger, K., Dasgupta, A., Langford, J., Smola, A., & Attenberg, J. (2009). *Feature hashing for large scale multitask learning*. Paper presented at the Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, Quebec, Canada.

Wiedemann, K., Naujoks, F., Wörle, J., Kenntner-Mabiala, R., Kaussner, Y., & Neukum, A. (2018). Effect of different alcohol levels on take-over performance in conditionally automated driving. *Accident Analysis & Prevention*, *115*, 89–97. doi:10.1016/j.aap.2018.03.001

Wright, T. J., Agrawal, R., Samuel, S., Wang, Y., Zilberstein, S., & Fisher, D. L. (2017). Effects of alert cue specificity on situation awareness in transfer of control in level 3 automation. *Transportation Research Record: Journal of the Transportation Research Board*, *2663*, 27–33. doi:10.3141/2663-04

Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017). *Evaluating effects of user experience and system transparency on trust in automation*. Paper presented at the Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria.

About the Authors

Feng Zhou is an Assistant Professor in the Department of Industrial and Manufacturing Systems Engineering at the University of Michigan, Dearborn. He obtained a Ph.D. degree in Engineering Design from Georgia Institute of Technology in 2014 and another Ph.D. degree in Human Factors Engineering from Nanyang Technological University, Singapore, in 2011.

X. Jessie Yang is an Assistant Professor in the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. She obtained a Ph.D. degree in Mechanical and Aerospace Engineering (Human Factors) from Nanyang Technological University, Singapore, in 2014.

Xin Zhang is a product marketing manager at Microsoft's Cloud division. His work spans from product management, business strategy to marketing for the cutting-edge cloud services at Microsoft. Xin earned a BS degree in Industrial Engineering from the Pennsylvania State University and an MBA and MS degree in Industrial & Operations Engineering from University of Michigan.