CrossMark

Anticipating acceptance of emerging technologies using twitter: the case of self-driving cars

Christopher Kohl[1] · **Marlene Knigge**[1] · **Galina Baader**[1] · **Markus Böhm**[1] · **Helmut Krcmar**[1]

Abstract In an early stage of developing emerging technologies, there is often great uncertainty regarding their future success. Companies can reduce this uncertainty by listening to the voice of customers as the customer eventually decides to accept an emerging technology or not. We show that risk and benefit perceptions are central determinants of acceptance of emerging technologies. We present an analysis of risk and benefit perception of self-driving cars from March 2015 until October 2016. In this period, we analyzed 1,963,905 tweets using supervised machine learning for text classification. Furthermore, we developed two new metrics, risk rate (RR) and benefit rate (BR), which allow analyzing risk and benefit perceptions on social media quantitatively. With our results, we provide impetus for further research on acceptance of self-driving cars and a methodological contribution to acceptance of emerging technologies research. Furthermore, we identify crucial issues in the public perception of self-driving cars and provide guidance for the management of emerging technologies to increase the likelihood of their acceptance.

✉ Christopher Kohl
christopher.kohl@in.tum.de

Marlene Knigge
marlene.knigge@in.tum.de

Galina Baader
galina.baader@in.tum.de

Markus Böhm
markus.boehm@in.tum.de

Helmut Krcmar
krcmar@in.tum.de

1 Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany

🙶 Springer

1 Introduction

The evolution of transportation has faced numerous trials as it has grown and
expanded over time. It seems safe to assume that this steady chain of development
of faster and safer vehicles with improved technological features continues (e.g.,
Burns 2013). Over the past decade, a vast amount of research has been conducted
regarding the topic of self-driving cars (Fagnant and Kockelman 2015; Kyriakidis
et al. 2015), which are being vigorously pursued by companies in the automotive
and related industries. Even native IT companies such as Google are pursuing the
development of self-driving cars (Spinrad 2014). However, for a technology to be
successful, we must remember a significant key factor for the success of emerging
technologies, technology acceptance (Davis et al. 1989).

In recent years, self-driving cars have become a controversially discussed topic.
Ethical, regulatory, and liability concerns (Zmud and Sener 2017; Gogoll and Mül-
ler 2017), centering on who is driving and who assumes responsibility for accidents,
are hotly debated issues. Nevertheless, the automotive and related industries seem
convinced that self-driving cars will be the future of mobility and may underesti-
mate the public's concerns and misconceptions related to this emerging technology
(Piao et al. 2016) that often differ from the perceptions of experts (Blake 1995).
With the first intelligent vehicle handling systems, a pre-stage of technologies that
enable self-driving cars, Conover (1994) already discussed that risk and benefit per-
ceptions could be an issue. Research regarding other technologies also shows that
risk and benefit perceptions are central determinants of their public acceptance
(Siegrist 2000; Butakov and Ioannou 2015). Public perceptions, therefore, eventu-
ally determine whether self-driving cars will be used and, thus, are a crucial factor
that needs to be considered especially for initial acceptance of emerging technolo-
gies such as self-driving cars (Butakov and Ioannou 2015; Pendleton et al. 2015;
Bansal et al. 2016).

However, studies addressing public perception and acceptance of this emerging
technology, especially across several countries and its change over time, remain
scarce. We address this paucity of research by first outlining the results of previ-
ous research on public perception and acceptance of self-driving cars. Second, we
describe a new approach to measure and react to public perceptions facilitating the
voice of the customer (VOC). This approach gives us the opportunity to utilize the
vast amounts of data publicly available in social media to anticipate acceptance of
emerging technologies and answer the following research questions:

RQ1: How can we measure public perceptions of self-driving cars to antici-
pate acceptance?

RQ2: How do events influence the public perception of self-driving cars?

To address these research questions, we created an approach for automatically determining and monitoring perceived risks and benefits of emerging technologies from short 140-character text messages published on the social media platform Twitter. We build on scientific literature and text mining methods, which allow the extraction of knowledge from text documents (Tan 1999) and, more specifically, sentiment analysis (Hopkins and King 2010). We use the social media platform Twitter to collect a stream of opinions about self-driving cars, one instance of currently emerging technology. Based on the perceived risks and benefits of self-driving cars, we identify events and issues crucial for the future acceptance of this emerging technology and guide the management of emerging technologies.

The remainder of this paper is structured as follows: First, we provide an overview of current literature on technology acceptance, self-driving cars, and previous research on the acceptance of self-driving cars. Second, we describe the data extraction from Twitter, the preprocessing of the data, and the model generation including its evaluation. Third, we describe and discuss the results of extracting the relevant data and applying our machine learning model to this data. We conclude with a summary of the results, limitations of our work, possibilities for further research, and the contributions to research and practice.

2 Theoretical background

In this section, we provide an overview of current literature disclosing the significance of acceptance towards self-driving cars from both an Information Systems (IS) and public acceptance perspective. An introduction to self-driving cars, the current scientific knowledge about them, and studies assessing the acceptance of self-driving cars are presented as well. We conclude this section by summarizing the theoretical background for our study.

2.1 Technology acceptance

Technology acceptance is one of the main research streams of IS research. The technology acceptance model (TAM) originates from this research stream and is a crucial source of numerous research endeavors (Venkatesh et al. 2007). TAM explains and predicts if and why IS will be used by individuals using three basic constructs: perceived usefulness, perceived ease of use, and behavioral intention to use the system under consideration (Davis et al. 1989). *Perceived usefulness* is defined as the probability that a specific IS increases the user's performance for a given task. *Perceived ease of use* describes the effort a user expects when using the IS to solve a given task. The main hypotheses of TAM are that perceived usefulness and perceived ease of use determine the strength of the behavioral intention of using a specific IS to solve a given task (Davis 1989). Behavioral intention then leads to actual use as described in the Theory of Reasoned Action (Fishbein

and Ajzen 1975), an influential theory from social psychology on which TAM is based (Davis et al. 1989).

Several researchers have extended TAM to consider the importance of risk perception for user acceptance (Venkatesh et al. 2016). For example, Martins et al. (2014) study internet banking adoption and conclude that risk perception is an important factor. They found that privacy risk and the risk of being subject to Internet banking fraud are important for Internet banking acceptance. Lancelot Miltgen et al. (2013) study end-user acceptance of biometrics and find that privacy risk is important for acceptance of biometrics. These studies show that assessing risk perception requires domain knowledge to identify risks that are relevant to a certain technology. However, despite some studies including risk perception as an additional factor, it has not been included in any of the central IS acceptance models (Venkatesh et al. 2016). Risk perception depends on the emerging technology itself and, therefore, is difficult to determine with standardized questionnaire items, as it is usually the case with TAM-based acceptance research.

Public acceptance research recognizes the central role of risk perception for acceptance. Previous research shows that many technologies have been rejected by people because of societal controversies, causing negative consequences for the commercialization of technologies (Gupta et al. 2012). Considering the vast investments in research and development of self-driving cars and the potential benefits of this technology for society, rejection of this technology could have severe consequences. The events and accidents that were recently reported with self-driving cars, such as the first human casualty (Yadron and Tynan 2016), could lead to fear and reluctance to accept, let alone try out this new technology (Hohenberger et al. 2016, 2017). Even traditional automotive companies, who are investing in research and development of self-driving cars, may suffer serious damage. If the technology failed to find wide acceptance, they would not get sufficient return on their investments.

Besides the research in the IS field, there is an influential model of technology acceptance in the public acceptance field proposed by Siegrist (2000). It specifically focuses on the relationship between perceptions of risks and benefits, trust, and public acceptance. Siegrist (2000) found that trust influences perceptions of risks and benefits, which in turn directly influence technology acceptance. To understand their model, it is important to differentiate between actual risks and benefits and their perceptions. The seminal work of Slovic (1987) describes the perception of risks associated with emerging technologies that are unfamiliar and incomprehensible to most people. People rely on intuitive judgments based on media reports rather than technologically sophisticated analyses to assess the risks more objectively. These judgments are often prone to biases caused by heuristics that may not lead to optimal or rational decisions as described in prospect theory (Kahneman and Tversky 1979). Furthermore, the perception of risk and benefits is confounded (Alhakami and Slovic 1994), which means that people do not differentiate between risks and benefits when evaluating new technology. Thus, we cannot expect people to make rational decisions based on facts but rather consider their perceptions when anticipating their behavior. These perceptions are

also influenced by trust, which helps people reduce cognitive complexity when evaluating new technologies (Earle and Cvetkovich 1995). Instead of their evaluation, people trust other entities to evaluate and apply emerging technologies correctly. In the case of self-driving cars, this could be the trust of people in regulatory authorities and the law to ensure that self-driving cars are safe to use (Choi and Ji 2015).

Risks and benefits of an emerging technology perceived by laypeople may vary significantly from the risks and benefits determined by experts. Emerging technologies and products, in particular, may cause anxiety and resistance to using them (Bongaerts et al. 2016; Zmud et al. 2016). Emerging technologies such as nanotechnology and genetically modified food struggle for acceptance, even if the benefits outweigh the risks from a scientific perspective, because of subjective (mis)perceptions (Gupta et al. 2012, 2015). Identifying perceived risks and benefits in an early product development stage, as it currently is the case with self-driving cars, allows companies to counteract precociously; that is, to adjust their products or product communication to address perceived risks and to exploit perceived benefits.

For anticipating and explaining technology acceptance, both IS and public acceptance research mainly rely on questionnaires. A questionnaire usually consists of several items for each construct (e.g., behavioral intention to use new technology), which were validated in prior research and adapted to the domain of application (e.g., telecommunication, banking) (Venkatesh et al. 2003). The questionnaire is then administered to a sample of the population and analyzed using regression or structural equation modeling after the data collection. This time-consuming and laborious approach comes with limitations. For example, an artifactual covariance between measures is caused by common methods, which can cause inflation of observed correlations (Sharma et al. 2009). Artifactual covariance is a major validity threat for IS acceptance as well as general social sciences research, which is often based on surveys (Sharma et al. 2009).

2.2 Self-driving cars

Driving automation, as in the case of self-driving cars, can be categorized into different levels of automation. There are three commonly used definitions of these levels. The German Federal Highway Research Institute (BASt) defines four levels of driving automation (Gasser and Westhoff 2012), the US National Highway Traffic Safety Administration (NHTSA) defines five levels of automation (NHTSA 2013), and the Society of Automotive Engineers (SAE) defines six levels of automation (SAE International 2014). Besides the different number of levels of automation, the definitions are similar. Kyriakidis et al. (2015) provide a comparison of the three definitions. We use the NHTSA definition to elaborate the different levels of automation exemplarily.

The idea of all definitions is to differentiate between driving automation systems providing no automation, systems providing only driving assistance, up to systems providing full automation. Table 1 summarizes the levels of driving automation. Current driving automation systems, such as Tesla's Autopilot, require that drivers

Table 1 Levels of driving automation (NHTSA 2013)

Level	Description
Level 0: non-autonomous	The driver is in complete control of the vehicle
Level 1: function specific automation	Automation involves only specific control functions (e.g., pre-charged brakes, electronic stability control)
Level 2: combined function automation	Automation of two primary control functions to work in unison (e.g., lateral and longitudinal control, to relieve driver of control of these functions)
Level 3: limited self-driving automation	The driver has the choice to give up control of all safety–critical functions under certain conditions, yet the driver is expected to be available for occasional control
Level 4: full self-driving automation	The vehicle has full control of all safety–critical driving functions under all conditions. The driver's availability is completely unnecessary

are in control of the car at any time and regardless of the current conditions. Hence, they need to be considered advanced driver-assistance systems (ADAS) that only provide level 2 driving automation according to the NHTSA definitions. However, the term "self-driving" is commonly associated with these cars. To emphasize when we specifically refer to level 4 automation, we use the term "full self-driving" in this paper.

The major drawback of current driving automation systems is that the driver must be able to take control of driving at any time (e.g., Yadron and Tynan 2016). Drivers are misusing those systems for example by leaving the driver's seat while driving on public roads using a level 2 driving automation system (Krok 2015). Considering how difficult it is for the driver to get back in the loop and properly react to certain traffic situations (Gold et al. 2013; Körber et al. 2016), such reports are troubling. They show that exaggerated benefit perceptions can have negative implications for driving safety and, thus, public acceptance.

A survey of public opinion about self-driving cars in the U.S., the U.K. and Australia with 1,533 respondents indicates that 56% of people have positive opinions towards self-driving cars while 13.8% express negative concerns and 29.4% are neutral towards the topic (Schoettle and Sivak 2014). The Consumer Technology Association (CTA) stated that 70% of drivers in the U.S. expressed interest in testing a self-driving car, and more than 60% of drivers showed a willingness in replacing their cars or trucks with a completely self-driving vehicle (Markwalter 2015). A study among 421 French drivers showed that 68.1% would be willing to use self-driving cars (Payre et al. 2014). Supporters argue that since 93% of car accidents are due to driver errors (Treat et al. 1977), the use of self-driving cars could reduce car accidents by that exact amount (Markwalter 2015; Fagnant and Kockelman 2015). However, opponents of this view state that self-driving cars might introduce new and currently unknown risks such as system failures or offsetting behaviors. Schoettle and Sivak (2014) concluded that self-driving cars may be no safer than an average driver and that they may increase the number of total vehicle accidents if self- and human-driven vehicles use the same roads.

In general, surveys show that people are accepting self-driving cars (e.g., Fraedrich et al. 2016) although they know only little about them. Previous research indicates that benefit perception positively influences technology acceptance (Hohenberger et al. 2017). However, focusing only on the benefits of self-driving cars might not be a sustainable strategy to increase their initial acceptance. If self-driving cars become widely available, people may begin to recognize potential safety issues and risks when they use them as in the case of active cruise control. When active cruise control was introduced in production vehicles, people began to recognize their loss of control, resulting in a lack of acceptance (Eckoldt et al. 2012). Therefore, car manufacturers need to communicate risks and limitations of self-driving cars. Misconceptions about both risks and benefits need to be avoided or even counteracted.

2.3 Summary

Currently, developers and manufacturers of full self-driving cars are racing to be the first ones on the market. They see the enormous potential of this emerging technology and the technological challenges, but they forget about the acceptance of customers (Rogers 2003). The prevalent method to measure acceptance is administering questionnaires, which might not be suitable for emerging technologies such as self-driving cars. Respondents to an acceptance questionnaire probably neither have detailed knowledge nor experience regarding self-driving cars, leading to biased results (Fraedrich and Lenz 2014). Rather than using standardized questionnaires, exploratory and qualitative research should be conducted in this field at an early stage (König and Neumayr 2017).

The voice of future customers provides interesting research opportunities for emerging technologies (Griffin and Hauser 1993). The risk and benefit perceptions of future customers are likely to play a central role in the acceptance of self-driving cars (Ward et al. 2017). Even before public availability, risk and benefit perceptions should be closely monitored to identify any issues with an emerging technology or its public perception. Previous research has conducted qualitative exploratory analyses of textual data about risk and benefit perceptions of self-driving cars and shown that this is a promising approach (Fraedrich and Lenz 2014; Bazilinskyy et al. 2015). The respective researchers put considerable effort in the manual coding of all data but struggled with this "cumbersome and time- consuming process" (Bazilinskyy et al. 2015, p. 2450), which still "could be biased" (Bazilinskyy et al. 2015, p. 2450). Furthermore, they suggest studying perceptions over time as they are likely to change as people become more familiar with this technology (Kauer et al. 2012; Haboucha et al. 2017; König and Neumayr 2017). In this context, it would be particularly interesting to study the effect of critical incidents with self-driving cars on public perceptions (Woisetschläger 2016). To address these findings and suggestions of previous research, we address the following propositions to answer our research questions:

P1: Machine learning algorithms can be used to analyze the publics' risk and benefit perceptions regarding self-driving cars on Twitter.

P2a: News concerning the benefits of self-driving cars (e.g., increased safety, reduced mobility costs) increases benefit perception of self-driving cars on Twitter.

P2b: News concerning the risks of self-driving cars (e.g., accidents, hacker attacks) increases risk perception of self-driving cars on Twitter.

3 Method

We use a novel approach to identify risks and benefits by analyzing the vast amount of existing data about self-driving cars on Twitter encoded in tweets. This approach is theoretically founded in the quantitative content analysis (Neuendorf 2016), which allows conducting quantitative data analyses based on qualitative data and extends previous qualitative approaches (Fraedrich and Lenz 2014; Bazilinskyy et al. 2015). While content analysis has been used to analyze mainly unstructured social media before (e.g., McCorkindale 2010), we automate most of the coding process using machine learning. This method is similar to sentiment analysis in marketing research (Okazaki et al. 2014) and has the advantage that only a small portion of the data needs manual coding when using supervised machine learning classification. By using this method, we avoid certain issues with questionnaires and studying technology acceptance, for example, common method variance (Sharma et al. 2009).

We follow the analysis process suggested by Okazaki et al. (2014) for sentiment analysis. It consists of data extraction, data preparation, model generation, model validation, and model application. Our approach to risk and benefit perception analysis is similar to sentiment analysis, which allows us to follow a common sentiment analysis process. Variations in the sentiment analysis process, for example, combining the steps of model generation and validation into one step (Feldman 2013) usually do not differ much.

We implemented the process of Okazaki et al. (2014) as follows: First, we obtain tweets using the Twitter Search API (data extraction). Second, we preprocess the tweets to improve data quality, reduce dimensionality, and avoid misclassification (data preparation). Third, we train the machine learning algorithm (model generation) and evaluate it using cross-validation (model validation). Fourth, we apply the machine learning algorithm to classify the tweets (model application). We then analyze the classified tweets qualitatively and quantitatively to address our research propositions.

3.1 Twitter mining

Twitter has often proven to be a valuable source of data for prediction and monitoring of diverse phenomena ranging from disease outbreaks (St Louis and Zorlu 2012) to political elections (Tumasjan et al. 2010). Users of Twitter face a limit of 140 characters per message, referred to as "tweet," to include all relevant information. Despite their brevity, tweets contain valuable information encoded in natural

language (Pak and Paroubek 2010). It is an ongoing challenge to extract this information from the vast amount of noise present on Twitter. We build on previous findings from sentiment analysis (Pak and Paroubek 2010) and machine learning classification to extract information from tweets. We need to extend previous approaches, as sentiment analysis is not directly applicable to the extraction of risk and benefit perceptions. It traditionally only assigns a polarity, i.e., positive or negative sentiment, to a given statement (Medhat et al. 2014).

New developments on Twitter include Twitter bots that are difficult to discern from real persons (Boshmaf et al. 2011; Haustein et al. 2016) and Internet of Things (IoT) devices that communicate over Twitter (Kranz et al. 2010). Therefore, results of Twitter analyses require careful consideration. The Twitter bots have especially become increasingly good at emulating human communication and writing style. Researchers are concerned about the large-scale infiltration of so-called "socialbots" that are hardly discernable from humans (Boshmaf et al. 2011). Socialbots make a quantitative analysis, for example analyzing tweet counts, not only from Twitter but also other social platforms such as Facebook, challenging (Haustein et al. 2016). However, we will compare the results of our analysis with previous studies to detect manipulations of the Twitter data.

A further issue with tweets is that they are not directly accessible from the authors. We only retrieve tweets in this study that are returned by the Twitter Search API, which are determined by proprietary algorithms and are not a representative sample of the overall tweets (Ruths and Pfeffer 2014). Furthermore, Twitter users are not a representative sample of the population (Ruths and Pfeffer 2014). However, Twitter has a broad audience from different social and interest groups and, thus, is a valuable source to assess people's perceptions (Pak and Paroubek 2010). We expect Twitter users to be more open to new technology, which could lead to slightly more positive results compared to previous surveys based on representative samples of the population.

Despite the limitations of Twitter mining, our approach allows accessing the VOC of 328 million monthly active Twitter users (Twitter 2017), which results in considerably more statements concerning the acceptance of self-driving cars than in previous research. To cope with the huge amounts of data from Twitter, we use machine learning to automate the classification of tweets. Thereby, we avoid the laborious manual coding of qualitative content, thus, making this research feasible.

3.2 Data extraction

Our dataset comprises tweets written in English concerning self-driving cars that were obtained using the Twitter Search API (Twitter 2016a). Furthermore, we developed a Java application as the Twitter Search API only allows retrieval of tweets not older than 1 week (Twitter 2016b). A meaningful longitudinal analysis, however, requires being able to collect tweets for longer intervals by collecting the tweets daily and storing them in a database. A MongoDB NoSQL database was used to store the complete tweets returned by the Twitter Search API including their date of creation, the username of the tweet creator, the message body, and a unique

identifier of the tweet. The tweets have then been transferred to an in-memory database to process them efficiently. We used a SAP HANA database SPS10 with 1 TB RAM and 32 CPU cores. SAP HANA offers R-integration, which allowed us to process the data using external R-libraries. The R server was installed on a virtual machine connected to the SAP HANA database. We started the data collection for this analysis on March 3, 2015, with the last tweets being collected on October 21, 2016. We used the following set of search queries (SQ) in our Twitter Search API requests:

- SQ1: self driving OR driverless OR autonomous OR automated
- SQ2: tesla OR google OR apple OR icar OR ford OR opel OR gm OR general motors
- SQ3: volkswagen OR vw OR daimler OR mercedes OR benz OR bmw OR audi OR porsche.

The search queries were fixed before the data collection. They consist of a combination of topic-related keywords (SQ1), names of US-based companies working on self-driving cars and U.S. car manufacturers (SQ2), as well as German car manufacturers (SQ2 and SQ3). Especially SQ2 and SQ3 resulted in many tweets that were not concerned with self-driving cars. However, at the beginning of our research in March 2015, we wanted to make sure that the search queries would find the relevant tweets without having to change the search queries over time. In total, we collected 1,963,905 tweets. For the data analysis, the tweets were filtered using the following regular expression:

```
(driver.?less|self.?driving|autonomous.?driving|auto
mated.?driving|autonomous.?car|automated.?car)
```

The filtering ensures that only tweets containing one of the following terms are included in the data analysis: driverless, self-driving, autonomous driving, automated driving, autonomous car, and automated car. The regular expression also ensures that slight variations of the terms are included, such as "driver less" or "driver-less." This filtering method reduced the number of relevant tweets to 642,033. For training the machine learning classifier, we used a dataset of 7,482 manually categorized tweets ranging from the beginning of January 2010 to June 2014, which we collected by scraping the "top tweets" about self-driving cars from the Twitter website. Table 2 shows the descriptive statistics of the training dataset.

Top tweets are "popular tweets that many other Twitter users have engaged with and thought were useful" (Twitter 2016c). Analyzing those tweets, we got

Table 2 Descriptive statistics of the training dataset

	Class		
	Risk	Benefit	Neutral
N	751	701	6030
%	10.0	9.37	80.6

an overview of the discussion of this topic on Twitter, which helped to design this study. However, we refrain from analyzing these tweets since they only represent a small fraction of the actual tweets published from January 2010 to June 2014 and are probably highly biased through the proprietary selection algorithms implemented and used by Twitter. Instead we do use them as "training data" for machine learning classification. In 80.6% of the tweets in the training dataset, no information about risk and benefit perceptions was present and were, thus, categorized as "Neutral."

3.3 Data preprocessing

We performed data preprocessing on the tweets to reduce dimensionality and avoid misclassification, a common step in text classification (Okazaki et al. 2014). That means we changed parts of the tweets so that the machine learning algorithm can extract relevant information from the tweets better, which results in better classification results. We use the text mining package "tm" for preprocessing, which provides a text mining framework for the statistics software R (Feinerer et al. 2008) to apply the preprocessing steps to both the training data and the tweets we want to classify automatically.

First, we transformed all characters in the body of the tweets to lower case. Like most of the preprocessing steps, this step decreases readability for humans. However, machine learning classifiers for text classification mainly rely on statistical features of the provided textual data and, thus, profit from such transformations. Second, we removed punctuation, numbers, and hyperlinks. Punctuation is not required to determine the classification of the tweets as we will not perform a grammatical analysis. Third, we removed English stop words as provided by the tm package. Stop words are terms that do not contain relevant information for the text classification (e.g., "a," "by," and "was"), so they are not needed in the further analysis. In addition to removing the English stop words, we removed the Twitter-specific stop words "via" and "rt." Fourth, we used stemming to reduce the dimensionality of the tweets further. Stemming reduces words with the same stem to the same word by stripping derivational and inflectional suffixes, for example: "driving" is stemmed to "drive."

The transformed tweets should now mainly contain words that are meaningful for machine learning classification. In the last step, we transformed the textual representation of the tweets into a document-term-matrix (DTM). In the scope of our analysis, terms are single words (i.e., unigrams). Only words that occur at least twice in the remaining set of tweets were included in the analysis. The entries in the DTM are the frequency of all terms (TF) in the tweet weighted by the inverse document frequency (IDF), i.e., how often a term occurs in the other tweets (Ramos 2003).

3.4 Model generation and evaluation

The basic idea of text classification with supervised machine learning is to assign classes to documents automatically using a much smaller set of training data

compared to the overall number of documents to classify. The training data usually contains manually classified documents. Based on these, the machine learning algorithm creates a model that determines how to classify new documents. Many different machine learning algorithms could be used for this task such as Naïve Bayes, Maximum Entropy Classification, or Support Vector Machines (SVM) (Pang et al. 2002).

We decided to use the SVM algorithm for text classification, which has been shown to be highly effective for this task (Joachims 1998; Pang et al. 2002). It does not require extensive parameter tuning and copes well with large feature vectors as it is usually the case with text classification problems (Joachims 1998). The basic idea of SVM is to find a hyperplane that separates the documents (tweets) according to their classification with a margin that is as large as possible, which is an optimization problem (Pang et al. 2002). We use the *Library for Support Vector Machines (LIBSVM)* implementation of SVM that allows classification, regression, and other learning tasks in R and other programming languages (Chang and Lin 2011).

We use a linear kernel function since this is recommended for datasets consisting of many instances and features (Hsu et al. 2016). Furthermore, text classification problems are often linearly separable (Joachims 1998). For linear kernels, only one parameter of the SVM algorithm can be varied to improve the classification performance: the regularization parameter C. We tested several different values ranging from 0.03125 to 100 using 10-fold cross-validation. A k-fold cross-validation randomly splits the training data into k mutually exclusive, approximately equal sized subsets (i.e., folds) (Kohavi 1995). The algorithm uses one of the k folds to evaluate the classifier by computing the accuracy and the other $k − 1$ folds to train it. This step is repeated k times until each fold has been used for training and evaluation. The result of the cross-validation is the mean accuracy of all k evaluations. Accuracy is defined as the overall number of correct classifications divided by the number of instances in the dataset.

Based on the cross-validations with varying of the C-parameter, we observed that higher values of C lead to higher classification accuracy. However, higher values of C also increase the compute time of training the SVM and improved the classification accuracy only slightly after C = 64. We determined C = 94 for the SVM, which has a very high accuracy of 0.925 considering results obtained in similar studies (e.g., Sriram et al. 2010). The SVM classification also leads to much better results than classification based on hand-picked keywords (Pang and Lee 2008).

For the second evaluation, we split the training data using a random selection of 80% ($N = 5984$) of the tweets for training the SVM and 20% ($N = 1496$) for evaluating the quality of the classification. We then compute several metrics based on the confusion matrix (Table 3).

The accuracy of the classification was 0.918, which is close to the average accuracy of the cross-validation. We computed the "no-information rate", the largest proportion of the observed classes, since the imbalance between the classes is large (Kuhn 2008). The no-information rate has a value of 0.807 and is significantly different ($p < 0.001$) from the accuracy. Additional metrics were computed according to Kuhn (2008) (Table 4).

Table 3 Confusion matrix of the SVM algorithm

	True class		
	Risk	Benefit	Neutral
Predicted class			
Risk	108	4	27
Benefit	3	101	15
Neutral	45	28	1165

Table 4 Metrics by class

Metric	Risk	Benefit	Neutral	Average
Sensitivity	0.69231	0.7594	0.9652	0.80564
Specificity	0.97687	0.98679	0.7474	0.90369
Pos. Pred. Value	0.77698	0.84874	0.941	0.85557
Neg. Pred. Value	0.96463	0.97676	0.8372	0.9262
Prevalence	0.10428	0.0889	0.8068	0.33333
Detection rate	0.07219	0.06751	0.7787	0.30613
Detection prevalence	0.09291	0.07955	0.8275	0.33332
Balanced accuracy	0.83459	0.8731	0.8563	0.85466

While accuracy shows very good values, we could identify issues of the classifier resulting from the imbalanced training set. For example, sensitivity for the classes "Risk" and "Benefit" and specificity for the class "Neutral" is relatively low, indicating that the classification is biased to assign the neutral class. However, for our analyses, we consider this result as a good compromise between sensitivity and specificity for all three classes.

4 Results

With an overall total of 642,033 tweets, we obtained 528,440 (82.3%) neutral tweets, 50,612 (7.88%) stated benefits (*BT*), and 62,981 (9.81%) stated risks about self-driving cars (*RT*). The risk ratio (RR) and benefit ratio (BR) were calculated as follows:

$$\text{RR} = \frac{\text{RT}}{\text{RT} + \text{BT}} = \frac{62981}{62981 + 50612} = 0.5544 \tag{1}$$

$$\text{BR} = \frac{\text{BT}}{\text{RT} + \text{BT}} = \frac{50612}{62981 + 50612} = 0.4456 = 1 - \text{RR} \tag{2}$$

The RR describes the ratio of tweets about risks of self-driving cars to the total number of tweets excluding the neutral tweets that neither contain information about risks nor benefits. BR is defined analogously and can also be calculated as $1 - \text{RR}$ when RR is already known as RR + BR equals to one.

Table 5 Number of tweets per year by class

Year	Neutral	Benefit	Risk	RR	BR
2015	407,179 83.0%	37,891 7.73%	45,214 9.22%	0.5441	0.4559
2016	121,261 79.9%	12,721 8.38%	17,767 11.7%	0.5828	0.4172
Overall	528,440 82.3%	50,612 7.88%	62,981 9.81%	0.5544	0.4456

In 2015, 407,179 (83.0%) of the tweets were neutral, 37,891 (7.73%) stated benefits, and 45,214 (9.22%) stated risks about self-driving cars (Table 5). The RR in 2015 equals to 0.5441 and BR equals to 0.4559. In 2016, 121,261 (79.9%) of the tweets were neutral, 12,721 (8.38%) stated benefits, and 17,767 (11.7%) stated risks about self-driving cars. Overall, we see only slightly changing ratios over the years in our analysis.

We found a decrease of the ratio of neutral tweets, which might indicate that the SVM classifier performs consistently over time. The classifier does not fail to detect risk and benefit tweets among the huge number of neutral tweets as time progresses and topics change. Additionally, the discussion about self-driving cars might also focus more on risks and benefits of the technology as they become more known to the public.

The fact that RR and BR did not change substantially from 2015 to 2016 shows that RR and BR might be a robust measure to quantify risk and benefit perception. Closer inspection of RR shows that it did change between the months and might be an important indicator of issues in risk and benefit perception. Figure 1 shows the development of RR and BR over time.

By analyzing the tweets in detail, we can track how Twitter users react to specific news reports, announcements, or other events and whether changing perceived risks or benefits is influenced by these events. We found that there is a relationship between occurrences related to self-driving cars and the content of tweets. This
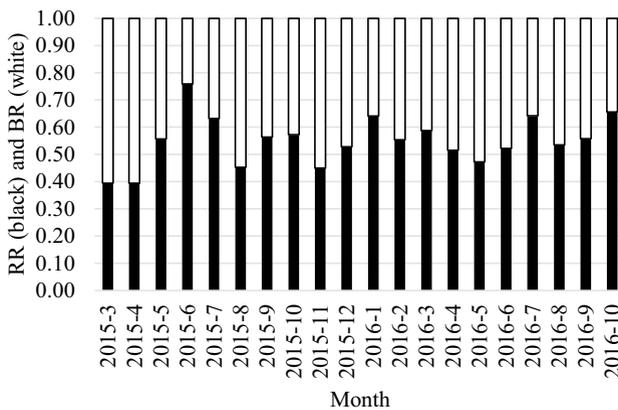

Fig. 1 RR and BR over time

finding allows us to state that, although being subjective, tweets regarding self-driving cars are related to facts and occurrences in the real world. In the following, we will give some examples.

In April 2015, we observed more tweets about benefits of self-driving cars than about risks as RR is smaller than 0.5 (Fig. 1). In this month, Google explained the use of solar energy and ridesharing for their self-driving car. Therefore, the Google Car has the potential to help protect the environment and ease congestion in major cities. Further, Google reported that their self-driving cars could reduce the number of people injured and killed in accidents. Tweets frequently discussed this report, which was one of the main reasons why in April 2015 RR was smaller than 0.5.

From May to July 2015, the tweets containing risks suddenly increased and outweighed the tweets containing benefits. The main reason for this is that Google released data showing that its self-driving cars had been involved in eleven minor accidents over the past 6 years. In June 2015, a website run by Google provided explanations for the accidents. In July 2015, Google reported its first accident with an injury. Although we have not explicitly filtered for Google-related posts, most tweets in this month were related to Google.

Evaluation of tweets from August 2015 revealed a spike in BR. We found that many tweets mentioned the announcement of self-driving crash trucks. These trucks are usually deployed on roadworks to protect the road workers from distracted drivers who would otherwise crash into them (Rubinkam 2015). Drivers of crash trucks are usually in a dangerous situation; replacing the driver with a driving automation system could save lives, which was obviously well-received by the public.

From September to October 2015, no noticeable changes in the RR occurred. In November 2015, Nevada approved regulations for self-driving cars, which lead to many tweets classified as benefit. The news that Google and Chrysler cooperated to produce self-driving cars was classified as benefit in April 2016. In June 2016, Volvo Cars & Ericsson released intelligent media streaming for self-driving cars, which lead to tweets classified as Neutral and, thus, did not influence RR.

In July 2016, the first fatal accident with the Tesla self-driving car occurred. We could detect this event in our data as it was often mentioned in tweets resulting in a spike in RR. In total, 631 risk tweets about self-driving cars were identified in our dataset compared to 352 benefit tweets.

In August 2016, Tesla announced a $2.6 Billion deal with Solar City to combine solar energy with self-driving cars. People also tweeted about self-driving tractors that promise to work by themselves. Also, Ford announced self-driving taxis for 2021 and people started tweeting about self-driving living rooms, so nobody has to leave the sofa. However, many tweets about accidents with self-driving cars were reported in this month, resulting in a RR value of 0.5348, which shows that tweets classified as risk ($N = 1269$) slightly prevail over tweets classified as benefit ($N = 1104$).

In October 2016, tweets often mentioned that a self-driving car hit a truck in Singapore. This lead to an increased risk rate. The last tweets in the analyzed dataset were from October 21, so the analyses do not reflect events after this date.

In addition to the RR, we also inspected the absolute numbers of tweets classified as risk (Fig. 2) and benefit (Fig. 3). Plotting the tweets over time, we could
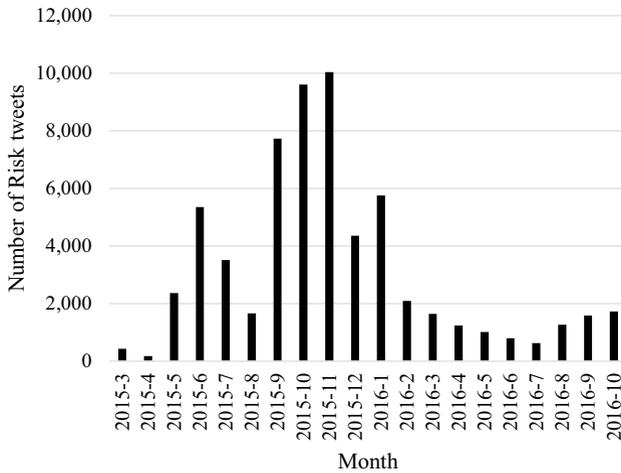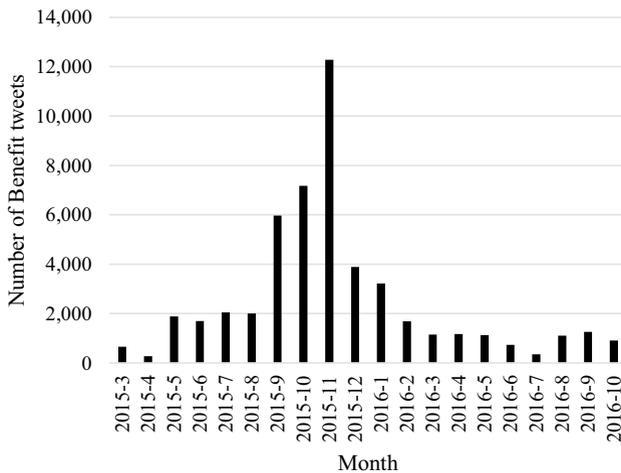
Fig. 2 Number of risk tweets per month

Fig. 3 Number of benefit tweets per month

observe several changes in the number of risk and benefit tweets. One characteristic month in our dataset is November 2015. A close inspection of the tweets shows that the annual *International Driverless Cars Conference* caused a general increase of tweets. This conference might not only have caused more tweets about self-driving cars but also led to an increased interest in reading tweets about self-driving cars, which were then retweeted—a self-reinforcing effect.

By analyzing the data, we recognized that the discussion about self-driving cars is dominated by the Google Car as found in previous research (Woisetschläger 2016). In contrast to currently available self-driving cars, the Google Car offers full self-driving capability (i.e., level 4 automation). So, the driver is not required to take
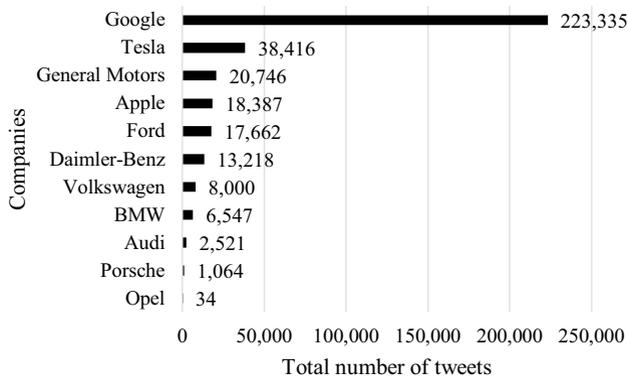
Fig. 4 Number of tweets about car manufacturers

over control of the car at any time. News such as a blind person driving the Google Car have been tweeted very often. Figure 4 shows the total number of tweets for each car manufacturer in the dataset. As described in Sect. 3.2, we specifically queried Twitter about tweets for those car manufacturers. The dominance of US-based companies is striking. The top five companies mentioned in tweets have their headquarters in the U.S. Furthermore, it is remarkable that IT companies are linked that strongly to this topic. Apple has not even officially announced a self-driving car but is only rumored to work on such a project, the Apple Car. However, we need to acknowledge that the dataset is biased for the U.S. market since we queried Twitter only for English tweets.

We also applied our analysis of the RR to the car manufacturer tweets. Again, we found interesting results as depicted in Fig. 5. The RR is lowest for five traditional car manufacturers. This might indicate that people tend to trust them to develop safe self-driving cars. However, we see Tesla is getting close to the traditional car
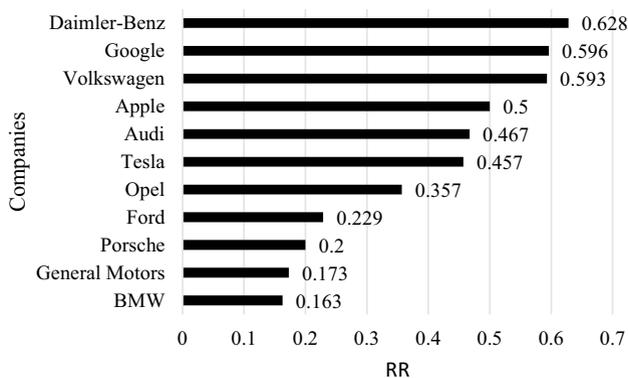

Fig. 5 RR of car manufacturers

manufacturers despite the accidents resulting from misuse of Tesla's Autopilot. It is also interesting that traditional German car manufacturers such as Daimler-Benz, Volkswagen, and Audi, which are considered the leading innovators of self-driving cars in Germany, have a high RR value. The existing driving automation system of Daimler-Benz, Intelligent Drive, is much more restrictive than the system of Tesla. Therefore, misuse of the driving automation system is much less likely than with the Tesla system and, thus, should be considered safer. However, users might perceive the additional restrictions as further loss of control, which is one of the major concerns about self-driving cars (Rödel et al. 2014; Langdon et al. 2018). In the case of Google, the users also seem to be afraid of not being able to drive anymore since one version of the Google Car had no steering wheel. Furthermore, many minor accidents of the Google Car have received much public attention. Again, we need to acknowledge that the results could be biased as the number of tweets differs very much between the car manufacturers. Our dataset contains over 6500 times more tweets about Google than about Opel. The RR makes the results comparable, but the readers should keep this difference in mind before drawing any fast conclusions.

5 Discussion

Before we discuss and interpret our results, we discuss their validity. First, we compare our findings to previous research. Fraedrich et al. (2016) found in a study conducted in Germany that 46% of their 1163 respondents had a positive connotation with self-driving cars. Schoettle and Sivak (2014) surveyed 1533 persons from the U.S., the U.K., and Australia for their opinions on self-driving vehicles and found that 56.8% of the respondents had at least a slightly positive opinion. Our BR ranges between 46% (2015) and 42% (2016), which is almost equal to the study of Fraedrich et al. (2016) and approximately 10% smaller than the rate of positive opinions from the study of Schoettle and Sivak (2014). Therefore, we are confident that our data resembles the public perceptions of risk and benefit perception and we found first support for our proposition *P1* that machine learning algorithms can be used to analyze the publics' risk and benefit perceptions regarding self-driving cars on Twitter. Second, our analysis shows that news and events were reflected in tweets as our in-depth analysis of tweets in the previous section has shown. Therefore, we also found support for our propositions *P2a* and *P2b*. In the following, we further discuss our findings and connect it to previous research on the acceptance of self-driving cars.

   Given the previous study results and the RR and BR values determined in this study, we conclude that people have specific reservations regarding self-driving cars. It could be surmised that if self-driving cars were released today, technology acceptance could not be guaranteed. Public concerns need to be addressed before the public is confronted with a rising number of increasingly automated cars or even full self-driving technology. We calculated the BR and RR values at various time points to analyze the tweets over time and found an increase in RR of 7.11% from 54.4% (2015) to 58.5% (2016). This might indicate that people's concerns have not been addressed. Furthermore, there might be a bias caused by a social amplification

of risk perceptions (Kasperson et al. 1988), i.e., people tend to talk more about risks than about benefits. While it is an interesting finding that social amplification of risk perceptions might exist in social media, the results are troubling. We see in many instances that social media increasingly leads to exaggerated risk perceptions that lead to irrational behaviors. The consequence might be that one single accident could severely reduce acceptance of self-driving cars of all manufacturers and providers.

We found many tweets that indicated a distorted perception of a risk, for example, "[…] Google's driverless cars have been involved in four car accidents" or "CAR CRASH Google Self Driving Cars to Decide if You Live or Die […]". However, rather than increasing the number of accidents, self-driving cars could significantly reduce them (Fagnant and Kockelman 2015). Other studies also found distorted perceptions, which might change as people become more familiar with self-driving cars (Woisetschläger 2016; Bansal et al. 2016).

People also expressed distrust towards the self-driving car manufacturing companies and conveyed their love for driving. For example: "Sorry @google not going to buy a self driving car I like driving and don't trust your technology." In this case, benefit perception is distorted. While driving can be enjoyable in certain situations, we find ourselves often confronted with less enjoyable aspects of driving such as traffic congestion, long monotonous highways with speed limitations or searching for a parking space in increasingly crowded cities. We find these concerns also in marketing research that predicts a lack of emotional attachment due to the loss of driving pleasure when a car drives itself (Olson 2017). First studies are addressing the user experience in self-driving cars (Rödel et al. 2014; Pettersson and Karlsson 2015; Niculescu et al. 2017). Other research proposes to adapt the driving style of self-driving cars to the driving style of their users to mitigate this issue (Kraus et al. 2009; Butakov and Ioannou 2015; Kuderer et al. 2015). To mitigate trust issues and to increase driving pleasure, previous research also suggests allowing drivers to take back control of their vehicles if they like (Yap et al. 2016).

People also displayed fear for their safety and privacy, for example, "[…] Can #driverless #cars be made safe from hackers?" expresses the fear of hackers that could take control of your vehicle. Hackers might even implement viruses that could spread from car to car, a risk that could prove to be real as hacker activities have been noted on current cars and due to the increased connectedness of self-driving cars (Lee et al. 2016). These hacking attacks could cause financial and physical harm and even death to car passengers and other road users, which is certainly more severe than having a personal computer hacked. Manufacturers of self-driving cars need to be aware of hackers and provide strategies to avoid hackers successfully attacking their vehicles. Previous research shows that hacker attacks have become one of the biggest concerns about self-driving cars (Zmud and Sener 2017).

Regarding the tweets classified as containing benefits of self-driving cars, many users liked the fact that they could save time by using self-driving vehicles. For example: "Sleepy time in the car for a in back seat. Wish I had a self driving car & I coulda joined em……". This tweet represents a case of distorted benefit

perception since only full self-driving automation allows sleeping while driving. The current level of self-driving cars is level 2, and it is likely to take some years until we arrive at level 3 or even level 4 automation. Meanwhile, drivers are misusing current self-driving cars, by leaving the driver's seat while driving on a public road using the Autopilot feature of a Tesla Model S (Krok 2015). Thereby, they, intentionally or unintentionally, risk their own and others' lives and potentially affect acceptance of self-driving cars by the public as the fatal accident of a self-driving Tesla has shown. Manufacturers should, however, be aware that people would like to sleep while being in a self-driving car to prevent it in cars with lower automation levels and offer a comfortable interior in full self-driving cars to enable it. Several studies mention sleeping while being in a self-driving car to be a popular activity (e.g., Cosh et al. 2017).

In general, people are impressed by the innovation put into the self-driving cars. For example: "[…] That hyper-futuristic driverless Mercedes has been spotted in San Fran – again […]". Most benefit tweets reflected that people were simply excited to try something new. For example: "[…] A perk of living near Google… We saw the self-driving car today on the highway!" Developers of self-driving cars have recognized that people are excited about this new technology and the benefits it could provide. Consequently, they are investing in the development of self-driving cars and already promise features that will only be implemented in several years. If communication strategies are not adjusted, this excitement could cause a misunderstanding of the potential benefits and exaggerated risk perceptions of self-driving cars. Focusing only on the benefits and even generating exaggerated benefit perceptions could also have adverse effects on public acceptance of self-driving cars (Nees 2016).

6 Conclusion

By analyzing 642,033 tweets, we could show that using machine learning to classify social media automatically is a promising approach to analyze acceptance of emerging technologies such as self-driving cars. Even if data from Twitter is prone to certain biases, our results are in line with previous research. Our approach mitigates some of the methodological shortcomings regarding data collection, time-consuming manual coding, and biases of online surveys for emerging technologies. Thereby, we follow the call by several researchers to provide new impetus for acceptance research (e.g., Benbasat and Barki 2007). Furthermore, our approach allows measuring the effect of certain events on public perception of emerging technologies. The identified perceived risks and benefits can be incorporated in traditional acceptance models for survey-based research.

Our research has several managerial implications. Based on our results, we identified the need for developers and manufacturers to listen to the voice of future potential customers. Even the objectively best solution or superior new technology development can fail if it does not appeal to the customer or does not create public acceptance. Therefore, an active management of the public acceptance is mandatory to reduce the failure rate of new technologies. In the case of self-driving cars,

companies need to rethink their communication strategies to address distorted perceptions of both the benefits and risks of self-driving cars (Kasperson and Kasperson 1996), already obvious with the first available level 2 automated cars. An overestimation of benefits might lead to a misuse of self-driving cars, the disappointment of initial users, and could have fatal consequences. An overestimation of risks by the public could lead to a resistance against self-driving cars before they even become widely available (Kleijnen et al. 2009; König and Neumayr 2017). Furthermore, practitioners should make sure to exploit the full potential of self-driving cars by implementing the benefits discussed in social media as described in this paper. First field studies show that people are more accepting of self-driving vehicles after having used prototypes (e.g., Alessandrini et al. 2011; Pendleton et al. 2015; Christie et al. 2016; Portouli et al. 2017; Madigan et al. 2017). Early personal experience with prototypes may lead to less susceptibility to distorted perceptions of self-driving cars and, therefore, should be made available more publicly by, e.g., self-driving car events of manufacturers, establishing additional model regions and test tracks, or creating driving experience centers for self-driving cars. We also found that user interface design will play an important role for the next generation of level 2 and level 3 driving automation systems as the risk perception of Daimler's Intelligent Drive system shows. Rather than the increased safety due to system restrictions, users primarily perceive the increased loss of control caused by them, which could reduce acceptance.

This research has some limitations. By using machine learning algorithms other than SVM and additional training data, analyses might be further improved. However, we do not expect major improvements since our analysis of the SVM classifier performance already showed good results and SVM are usually among the strongest performers for text classification (Socher et al. 2011). Further research should rather focus on more detailed risk and benefit categories. For example, Slovic (1987) describes risks as a combination of two factors: unknown risks, which are not observable and unknown to those exposed, and dread risks, which can be globally catastrophic and fatal. Hohenberger et al. (2017) divided the benefits of self-driving cars into the sub-categories of economic, time, and safety benefits. Further research could analyze risk and benefit perceptions using sub-categories of risks and benefits or identify new categories influencing technology acceptance based on data from social media.

For further research, it could be interesting to use topic modeling (Debortoli et al. 2016) to support the in-depth qualitative analysis of the tweets after classification. A next step could be to include other social media sources into the analyses such as Facebook, Reddit, or blogs. While our Twitter sample appears to be comparable to existing survey samples, extending the analyses to other platforms would allow analyzing differences in the explication and dissemination of perceptions between social media platforms. However, this would require extending the technical platform as the structures of other social media platforms differ from Twitter significantly. Besides analyzing opinions expressed in written text, VOC can also be extracted from other media such as recorded speech and videos, as mentioned by Brown (2017), who analyzed YouTube videos about experiences with driving automation. Including diverse sources may lead to a broader

overview of opinions and perceptions and helps to improve our understanding of the various aspects of emerging technologies, which are relevant for potential customers and society.

References

Alessandrini A, Holguin C, Parent M (2011) Advanced transport systems showcased in La Rochelle. In: Proceedings IEEE conference on intelligent transportation systems, ITSC. pp 896–900

Alhakami AS, Slovic P (1994) A psychological study of the inverse relationship between perceived risk and perceived benefit. Risk Anal 14:1085–1096. https://doi.org/10.1111/j.1539-6924.1994.tb00080.x

Bansal P, Kockelman KM, Singh A (2016) Assessing public opinions of and interest in new vehicle technologies: an Austin perspective. Transp Res Part C Emerg Technol 67:1–14. https://doi.org/10.1016/j.trc.2016.01.019

Bazilinskyy P, Kyriakidis M, de Winter J (2015) An international crowdsourcing study into people's statements on fully automated driving. Proc Manuf. https://doi.org/10.1016/j.promfg.2015.07.540

Benbasat I, Barki H (2007) Quo vadis TAM? J Assoc Inf Syst 8:211–218. http://aisel.aisnet.org/jais/vol8/iss4/7/

Blake ER (1995) Understanding outrage: how scientists can help bridge the risk perception gap. Environ Health Perspect 103:123–125. https://doi.org/10.2307/3432360

Bongaerts R, Kwiatkowski M, König T (2016) Disruption technology in mobility: customer acceptance and examples. In: Phantom Ex machina: digital disruption's role in business model transformation. Springer International Publishing, Switzerland, pp 119–135

Boshmaf Y, Muslukhov I, Beznosov K, Ripeanu M (2011) The socialbot network. In: Proceedings of the 27th annual computer security applications conference on—ACSAC'11. ACM Press, New York, USA, p 93

Brown B (2017) The social life of autonomous cars. Computer (Long Beach Calif) 50:92–96. https://doi.org/10.1109/MC.2017.59

Burns LD (2013) Sustainable mobility: a vision of our transport future. Nature 497:181–182. https://doi.org/10.1038/497181a

Butakov V, Ioannou P (2015) Driving autopilot with personalization feature for improved safety and comfort. In: 2015 IEEE 18th international conference on intelligent transportation systems. IEEE, Las Palmas, Spain, pp 387–393

Chang C, Lin C (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:1–39. https://doi.org/10.1145/1961189.1961199

Choi JK, Ji YG (2015) Investigating the importance of trust on adopting an autonomous vehicle. Int J Hum Comput Interact. https://doi.org/10.1080/10447318.2015.1070549

Christie D, Koymans A, Chanard T et al (2016) Pioneering driverless electric vehicles in Europe: the city automated transport system (CATS). Transp Res Proc 13:30–39. https://doi.org/10.1016/j.trpro.2016.05.004

Conover GD (1994) The eleven commandments for IVHS. In: Vehicle navigation and information systems conference proceedings. Yokohama, Japan, pp 503–506

Cosh K, Wordingham S, Ramingwong S (2017) Investigating public opinion regarding autonomous vehicles: a perspective from Chiang Mai, Thailand. Lect Notes Electr Eng 450:3–10

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 13:319–340

Davis FD, Bagozzi RP, Warshaw PR (1989) User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 35:982–1003. https://doi.org/10.1287/mnsc.35.8.982

Debortoli S, Müller O, Junglas I, Vom Brocke J (2016) Text mining for information systems researchers: an annotated topic modeling tutorial. Commun Assoc Inf Syst 39:110–135

Earle TC, Cvetkovich G (1995) Social trust: toward a cosmopolitan society. Praeger Publishers, Westport

Eckoldt K, Knobel M, Hassenzahl M, Schumann J (2012) An experiential perspective on advanced driver assistance systems. Inf Technol 54:165–171. https://doi.org/10.1524/itit.2012.0678

Fagnant DJ, Kockelman K (2015) Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transp Res Part A Policy Pract 77:167–181. https://doi.org/10.1016/j.tra.2015.04.003

Feinerer I, Hornik K, Meyer D (2008) Text mining infrastructure in R. J Stat Softw 25:1–54

Feldman R (2013) Techniques and applications for sentiment analysis. Commun ACM 56:82. https://doi.org/10.1145/2436256.2436274

Fishbein M, Ajzen I (1975) Belief, attitude, intention and behavior: an introduction to theory and research. Addison-Wesley, Reading

Fraedrich E, Lenz B (2014) Automated driving: individual and societal aspects. Transp Res Rec 2416:64–72

Fraedrich E, Cyganski R, Wolf I, Lenz B (2016) User perspectives on autonomous driving. In: Arbeitsberichte 187. Geographisches Institut, Humboldt-Universität, Berlin

Gasser TM, Westhoff D (2012) BASt-study: definitions of automation and legal issues in Germany. In: Proceedings of the 2012 road vehicle automation workshop

Gogoll J, Müller JF (2017) Autonomous cars. in favor of a mandatory ethics setting. Sci Eng Ethics 23:681–700. https://doi.org/10.1007/s11948-016-9806-x

Gold C, Dambock D, Lorenz L, Bengler K (2013) "Take over!" How long does it take to get the driver back into the loop? In: Proceedings of the human factors and ergonomics society annual meeting. pp 1938–1942

Griffin A, Hauser JR (1993) The voice of the customer. Mark Sci 12:1–27. https://doi.org/10.1287/mksc.12.1.1

Gupta N, Fischer ARH, Frewer LJ (2012) Socio-psychological determinants of public acceptance of technologies: a review. Public Underst Sci 21:782–795. https://doi.org/10.1177/0963662510392485

Gupta N, Fischer ARH, Frewer LJ (2015) Ethics, risk and benefits associated with different applications of nanotechnology: a comparison of expert and consumer perceptions of drivers of societal acceptance. Nanoethics 9:93–108. https://doi.org/10.1007/s11569-015-0222-5

Haboucha CJ, Ishaq R, Shiftan Y (2017) User preferences regarding autonomous vehicles. Transp Res Part C Emerg Technol. https://doi.org/10.1016/j.trc.2017.01.010

Haustein S, Bowman TD, Holmberg K et al (2016) Tweets as impact indicators: examining the implications of automated "bot" accounts on Twitter. J Assoc Inf Sci Technol 67:232–238. https://doi.org/10.1002/asi.23456

Hohenberger C, Spörrle M, Welpe IM (2016) How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups. Transp Res Part A Policy Pract 94:374–385. https://doi.org/10.1016/j.tra.2016.09.022

Hohenberger C, Spörrle M, Welpe IM (2017) Not fearless, but self-enhanced: the effects of anxiety on the willingness to use autonomous cars depend on individual levels of self-enhancement. Technol Forecast Soc Change 116:40–52. https://doi.org/10.1016/j.techfore.2016.11.011

Hopkins DJ, King G (2010) A method of automated nonparametric content analysis for social science. Am J Pol Sci 54:229–247. https://doi.org/10.1111/j.1540-5907.2009.00428.x

Hsu C-W, Chang C-C, Lin C-J (2016) A practical guide to support vector classification. In: Natl. Taiwan Univ. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. Accessed 25 Oct 2016

Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: Nédellec C, Rouveirol C (eds) Machine learning: ECML-98. Springer, Berlin, pp 137–142

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47:263. https://doi.org/10.2307/1914185

Kasperson RE, Kasperson JX (1996) The social amplification and attenuation of risk. Ann Am Acad Pol Soc Sci 545:95–105. https://doi.org/10.1177/0002716296545001010

Kasperson RE, Renn O, Slovic P et al (1988) The social amplification of risk: a conceptual framework. Risk Anal 8:177–187. https://doi.org/10.1111/j.1539-6924.1988.tb01168.x

Kauer M, Franz B, Schreiber M et al (2012) User acceptance of cooperative maneuverbased driving—a summary of three studies. Work 41:4258–4264. https://doi.org/10.3233/WOR-2012-0720-4258

Kleijnen M, Lee N, Wetzels M (2009) An exploration of consumer resistance to innovation and its antecedents. J Econ Psychol 30:344–357. https://doi.org/10.1016/j.joep.2009.02.004

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Int Jt Conf Artif Intell 14:1137–1143. https://doi.org/10.1067/mod.2000.109031

König M, Neumayr L (2017) Users' resistance towards radical innovations: the case of the self-driving car. Transp Res Part F Traffic Psychol Behav 44:42–52. https://doi.org/10.1016/j.trf.2016.10.013

Körber M, Gold C, Lechner D, Bengler K (2016) The influence of age on the take-over of vehicle control in highly automated driving. Transp Res Part F Traffic Psychol Behav. https://doi.org/10.1016/j.trf.2016.03.002

Kranz M, Roalter L, Michahelles F (2010) Things that twitter: social networks and the internet of things. In: What can the internet of things do for the citizen (CIoT) workshop at the eighth international conference on pervasive computing (Pervasive 2010)

Kraus S, Althoff M, Heißing B, Buss M (2009) Cognition and emotion in autonomous cars. In: IEEE Intelligent vehicles symposium, proceedings. pp 635–640

Krok A (2015) This is the stupidest misuse of Tesla's autopilot yet. http://www.cnet.com/news/this-is-the-stupidest-misuse-of-teslas-autopilot-yet/. Accessed 25 Nov 2015

Kuderer M, Gulati S, Burgard W (2015) Learning driving styles for autonomous vehicles from demonstration. In: 2015 IEEE International conference on robotics and automation (ICRA). IEEE, Seattle, pp 2641–2646

Kuhn M (2008) Building predictive models in R using the caret package. J Stat Softw 28:1–26. https://doi.org/10.1053/j.sodo.2009.03.002

Kyriakidis M, Happee R, de Winter JCF (2015) Public opinion on automated driving: results of an international questionnaire among 5000 respondents. Transp Res Part F Traffic Psychol Behav 32:127–140. https://doi.org/10.1016/j.trf.2015.04.014

Lancelot Miltgen C, Popovič A, Oliveira T (2013) Determinants of end-user acceptance of biometrics: integrating the "Big 3" of technology acceptance with privacy context. Decis Support Syst 56:103–114. https://doi.org/10.1016/j.dss.2013.05.010

Langdon P, Politis I, Bradley M, et al (2018) Obtaining design requirements from the public understanding of driverless technology. In: Advances in intelligent systems and computing. pp 749–759

Lee E-K, Gerla M, Pau G et al (2016) Internet of vehicles: from intelligent grid to autonomous cars and vehicular fogs. Int J Distrib Sens Netw 12:241–246. https://doi.org/10.1177/1550147716665500

Madigan R, Louw T, Wilbrink M et al (2017) What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems. Transp Res Part F Traffic Psychol Behav 50:55–64. https://doi.org/10.1016/j.trf.2017.07.007

Markwalter B (2015) The path to driverless cars. IEEE Consum Electron Mag 6:125–126. https://doi.org/10.1109/MCE.2016.2640625

Martins C, Oliveira T, Popovič A (2014) Understanding the Internet banking adoption: a unified theory of acceptance and use of technology and perceived risk application. Int J Inf Manage 34:1–13. https://doi.org/10.1016/j.ijinfomgt.2013.06.002

McCorkindale T (2010) Can you see the writing on my wall? A content analysis of the Fortune 50's Facebook social networking sites. Public Relat J 4:1–14. https://doi.org/10.1017/CBO9781107415324.004

Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5:1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Nees MA (2016) Acceptance of self-driving cars: an examination of idealized versus realistic portrayals with a self-driving car acceptance scale. Proc Hum Factors Ergon Soc Annu Meet. https://doi.org/10.1177/1541931213601332

Neuendorf K (2016) The content analysis guidebook. SAGE Publications, London

NHTSA (2013) National highway traffic safety administration preliminary statement of policy concerning automated vehicles. http://www.nhtsa.gov/About+NHTSA/Press+Releases/U.S.+Department +of+Transportation+Releases+Policy+on+Automated+Vehicle+Development. Accessed 1 Sep 2016

Niculescu AI, Dix A, Yeo KH (2017) Are you ready for a drive? User perspectives on autonomous vehicles. In: Conference on human factors in computing systems—proceedings. pp 2810–2817

Okazaki S, Diaz-Martin AM, Rozano M, Menendez-Benito H (2014) How to mine brand tweets procedural guidelines and pretest. Int J Mark Res 56:467–489. https://doi.org/10.2501/IJMR-2014-008

Olson EL (2017) Will songs be written about autonomous cars? The implications of self-driving vehicle technology on consumer brand equity and relationships. Int J Technol Mark 12:23. https://doi. org/10.1504/IJTMKT.2017.081506

Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the seventh conference on international language resources and evaluation. pp 1320–1326

Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2:1–135. https://doi. org/10.1561/1500000001

Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing—EMNLP'02. pp 79–86

Payre W, Cestac J, Delhomme P (2014) Intention to use a fully automated car: attitudes and a priori acceptability. Transp Res Part F Traffic Psychol Behav 27:252–263. https://doi.org/10.1016/j. trf.2014.04.009

Pendleton S, Uthaicharoenpong T, Chong ZJ, et al (2015) Autonomous golf cars for public trial of mobility-on-demand service. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 1164–1171

Pettersson I, Karlsson ICM (2015) Setting the stage for autonomous cars: a pilot study of future autonomous driving experiences. IET Intell Transp Syst 9:694–701. https://doi.org/10.1049/ iet-its.2014.0168

Piao J, McDonald M, Hounsell N, et al (2016) Public views towards implementation of automated vehicles in urban areas. In: Transportation research procedia

Portouli E, Karaseitanidis G, Lytrivis P, et al (2017) Public attitudes towards autonomous mini buses operating in real conditions in a Hellenic city. In: 2017 IEEE intelligent vehicles symposium (IV). IEEE, pp 571–576

Ramos J (2003) Using TF-IDF to determine word relevance in document queries. In: Proceedings of the first instructional conference on machine learning. pp 133–142

Rödel C, Stadler S, Meschtscherjakov A, Tscheligi M (2014) Towards autonomous cars: the effect of autonomy levels on acceptance and user experience. In: AutomotiveUI 2014—6th international conference on automotive user interfaces and interactive vehicular applications, in cooperation with ACM SIGCHI—Proceedings

Rogers EM (2003) Diffusion of innovations, 5th edn. Free Press, New York

Rubinkam M (2015) Driverless truck meant to improve safety in work zones. In: Yahoo! news. https:// www.yahoo.com/news/driverless-truck-meant-improve-safety-zones-202055180.html. Accessed 1 Aug 2016

Ruths D, Pfeffer J (2014) Social media for large studies of behavior. Science 346(80):1063–1064. https:// doi.org/10.1126/science.346.6213.1063

SAE International (2014) Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. In: On-Road Autom. Veh. Stand. Comi. http://standards.sae.org/j3016_201401/. Accessed 24 Oct 2016

Schoettle B, Sivak M (2014) A survey of public opinion about autonomous and self-driving vehicles in the U.S., the U.K., and Australia. https://deepblue.lib.umich.edu/handle/2027.42/108384

Sharma R, Yetton P, Crawford J (2009) Estimating the effect of common method variance: the method-method pair technique with an illustration from TAM research. MIS Q 33:473–490

Siegrist M (2000) The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. Risk Anal 20:195–204. https://doi.org/10.1111/0272-4332.202020

Slovic P (1987) Perception of risk. Science 236(80):280–285. https://doi.org/10.1126/science.3563507

Socher R, Lin CC-Y, Ng AY, Manning CD (2011) Parsing natural scenes and natural language with recursive neural networks. In: ICML'11 Proceedings of the 28th international conference on international conference on machine learning. pp 129–136

Spinrad N (2014) Google car takes the test. Nature 514:528. https://doi.org/10.1038/514528a

Sriram B, Fuhry D, Demir E, et al (2010) Short text classification in twitter to improve information filtering. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval—SIGIR'10. pp 841–842

St Louis C, Zorlu G (2012) Can twitter predict disease outbreaks? BMJ 344:e2353. https://doi.org/10.1136/bmj.e2353

Tan A-H (1999) Text mining: the state of the art and the challenges. In: Proceedings of the PAKDD 1999 workshop on knowledge disocovery from advanced databases. pp 65–70

Treat JR, Tumbas NS, McDonald ST, et al (1977) Tri-level study of the causes of traffic accidents: final report. https://trid.trb.org/view.aspx?id=144150. Accessed 1 Aug 2016

Tumasjan A, Sprenger TO, Sandner PG, Welpe IM (2010) Predicting elections with twitter: what 140 characters reveal about political sentiment. In: Proceedings of the fourth international AAAI conference on weblogs and social media. pp 178–185

Twitter (2016a) The search API. https://dev.twitter.com/rest/public/search. Accessed 1 Aug 2016

Twitter (2016b) Public API: GET search/tweets. https://dev.twitter.com/rest/reference/get/search/tweets. Accessed 1 Aug 2016

Twitter (2016c) Help center: the basics. https://support.twitter.com/articles/131209. Accessed 1 Aug 2016

Twitter (2017) Selected company metrics and financials. https://investor.twitterinc.com/. Accessed 29 Sep 2017

Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27:425–478. https://doi.org/10.2307/30036540

Venkatesh V, Davis FD, Morris MG (2007) Dead or alive? The development, trajectory and future of technology adoption research. J Assoc Inf Syst 8:267–286. https://doi.org/10.1016/j.wneu.2011.04.002

Venkatesh V, Thong JYL, Xu X (2016) Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 17:328–376

Ward C, Raue M, Lee C, et al (2017) Acceptance of automated driving across generations: The role of risk and benefit perception, knowledge, and trust. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10271:254–266

Woisetschläger DM (2016) Consumer perceptions of automated driving technologies: an examination of use cases and branding strategies. In: Autonomous driving: technical, legal and social aspects. pp 687–706

Yadron D, Tynan D (2016) Tesla driver dies in first fatal crash while using autopilot mode. In: Guard. https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk

Yap MD, Correia G, van Arem B (2016) Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips. Transp Res Part A Policy Pract. https://doi.org/10.1016/j.tra.2016.09.003

Zmud JP, Sener IN (2017) Towards an understanding of the travel behavior impact of autonomous vehicles. Transp Res Proc 25:2500–2519. https://doi.org/10.1016/j.trpro.2017.05.281

Zmud JP, Sener IN, Wagner J (2016) Self-driving vehicles. Transp Res Rec J Transp Res Board 2565:57–64. https://doi.org/10.3141/2565-07