4

The Voice of the Crowd—An Innovation Mining Study on Autonomous Driving

Michael Bartl and Juan Rosenzweig

4.1 Introduction

There is little doubt that the Internet has changed the way consumers communicate. An increasing number of users actively gather together online and communicate in web forums, blogs, and various kinds of user-generated content (UGC) platforms. They exchange personal experiences and opinions about products and their usage and talk about opportunities for solving product-related problems. Some of them even develop product modifications and innovations, which they post online and share with other community members. This turns online communities into powerful sources of innovation (Füller et al. 2006; Bartl et al. 2012; Bilgram et al. 2008). Within this context organizations are experimenting with a variety

M. Bartl (✉)
HYVE AG, München, Germany
e-mail: michael.bartl@hyve.net

J. Rosenzweig
HYVE AG, Wien, Oesterreich

A. Brem, E. Viardot (eds.), *Revolution of Innovation Management*,
DOI 10.1057/978-1-137-57475-6_4

of new and modified innovation research approaches promoting the role of consumers as valuable cocreators of products and services (von Hippel 2005; Chesbrough 2003; Prahalad and Ramaswamy 2000; Cui and Wu 2016; Gemser and Perks 2015). One example is the concept of crowdsourcing with the underlying idea of taking tasks traditionally performed by companies and outsourcing them to an undefined, generally large group of people in the form of an open call (Howe 2006). Other advancements are made in developing further qualitative research approaches with netnography as a prominent example (Bartl et al. 2016b; Kozinets 2002; Brem and Bilgram 2015; Wiles et al. 2013; Zhang et al. 2013). Evolved from ethnographic research, the core idea of netnography is to gain unbiased, unobtrusive consumer insights by "listening in" the user conversation. The advantage of the researcher's in-depth qualitative analysis of consumer quotes is the strength of netnography and, at the same time, its limitation. In order to manage the exponentially growing data volumes of UGC, new quantitative approaches relying on automation in text analysis of software-based information retrieval are on the rise. The aim of this chapter is to introduce innovation mining as a new powerful quantitative research technique and systematic procedure to identify, select, and analyze large volumes of user conversations on the Internet and make them usable for innovation challenges. Sections 4.2 and 4.3 describe the field of autonomous driving as a disruptive field of innovation which is chosen to showcase the innovation mining method. Section 4.4 describes the five methodological steps of innovation mining. Section 4.5 summarizes the study results followed by a concluding outlook in Sect. 4.6.

4.2 The Innovation Path of Autonomous Driving

Childhood dreams from the televisions series of Knight Rider are coming true. The Knight Industries Two Thousand (KITT) was a self-driving Pontiac Firebird Trans Am packed with lots of artificial intelligence supporting Michael Knight and the Foundation for Law and Government to fight down numerous villains. In the 1980s the self-driving KITT was a science fiction scenario for the audience. Today integrated camera, radar, laser, infrared, and ultrasonic technologies make it possible to record and interpret all relevant data from the car's surroundings. Then, a control unit backed with lots of computing power can take over and drive the car without any human intervention. The autonomous car is definitely not a gadget for a few enthusiasts, it will be the most impactful and disruptive innovation in the history of the automobile with enormous social and economic implications. Moreover, it is exciting that we are right now experiencing the birth of this innovation that will be around for the next centuries.

First estimations state that autonomous cars can contribute $1.3 trillion in annual savings to the US economy alone, with global savings estimated at over $5.6 trillion (Morgan Stanley 2013). There are many drivers for the overall savings as illustrated in Fig. 4.1.

For example, improvements in fuel consumption can be achieved when driving smoothly or using cruise control compared to manual breaking and throttling. Furthermore, self-driving cars could prevent 90 per cent of road traffic accidents, which are mainly caused by human error. Over

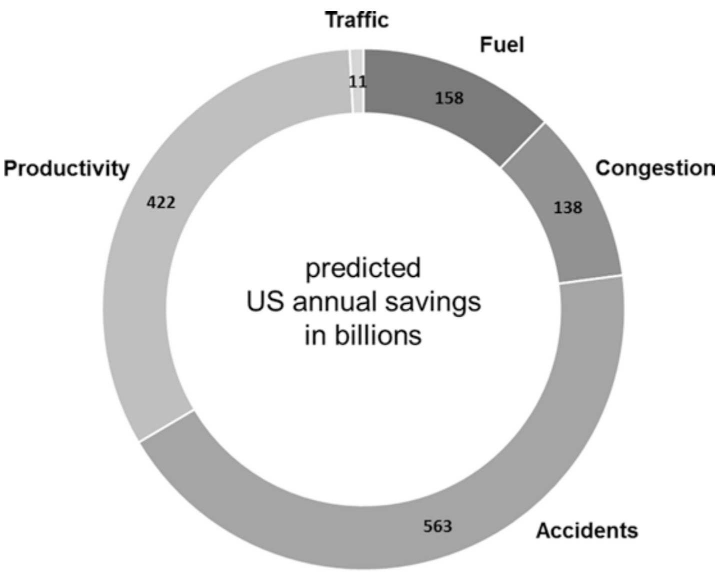


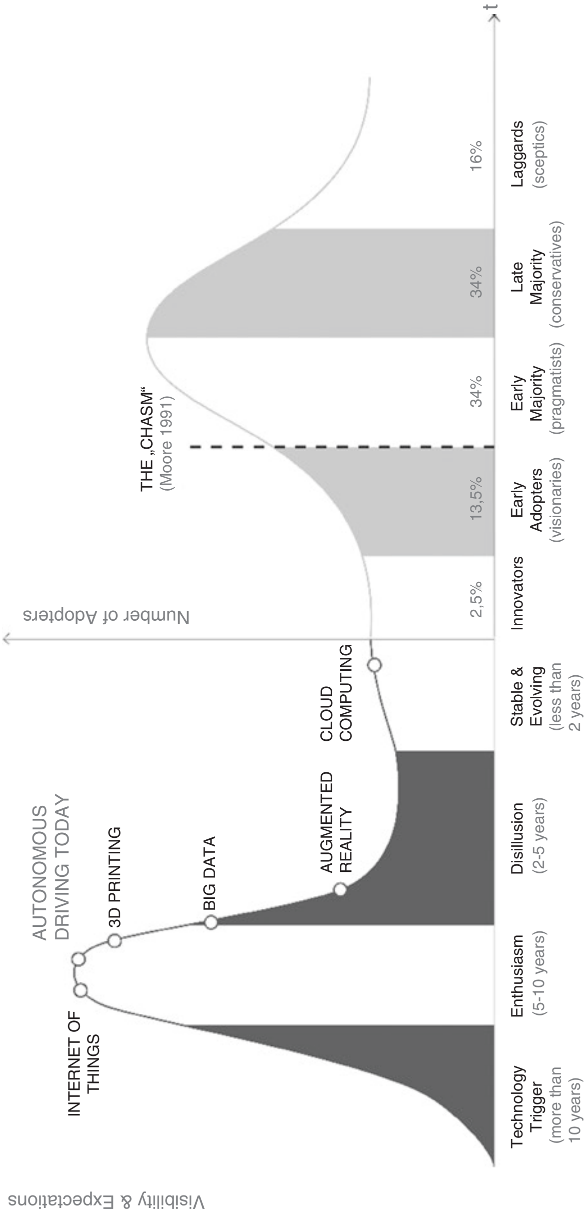
Fig. 4.1 US annual savings of $1.3 trillion (adapted from: Morgan Stanley 2013)

40 per cent of fatal crashes involve alcohol, distraction, drug involvement, and/or fatigue. When you take into account the loss of earnings, household production, medical and emergency services costs, travel delay, and administration costs, this adds up to a massive amount of money. There are also huge productivity gains as occupants do not have to drive anymore and can use their new free time. Better traffic management on roads with connected and autonomous cars will lead to less congestion. These saving predictions will, however, only apply in a world of fully autonomous cars. This is level 4 of the National Highway Traffic Safety Administration (NHTSA) model of technology penetration shown in Fig. 4.2.

Beside some analysts' estimations of the expected economic impact the authors want to include an additional perspective on the innovation path of autonomous driving (AD) using Gartner's hype cycle combined with Roger's diffusion model of innovation. The hype cycle offers a suitable tool to evaluate the current stage and relative maturity of the technology in the early phases of its life cycle (Fenn 2012). The model can be used to indicate consumer attitudes towards technology and can serve as a basis to analyze opportunities and investment risks regarding a certain technology (Linden and Fenn 2003). The shape of the hype cycle curve in Fig. 4.3 illustrates the media overenthusiasm through the period of disillusion to an eventual understanding of the technology's relevance and role in the market (De

No-Automation Level 0 (now)	The driver is in complete and sole control of the primary vehicle controls – brake, steering, throttle, and motive power – at all times.
Function-specific Automation Level 1 (now)	Automation at this level involves one or more specific control functions. Examples include electronic stability control or pre-charged brakes, where the vehicle automatically assists with braking to enable the driver to regain control of the vehicle or stop faster than possible by acting alon.
Combined Function Automation Level 2 (now)	This level involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. An example of combined functions enabling a Level 2 system is adaptive cruise control in combination with lane centering.
Limited Self-Driving Automation Level 3 (2020+)	Vehicles at this level of automation enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control. The driver is expected to be available for occasional control, but with sufficiently comfortable transition time.
Full Self-Driving Automation Level 4 (2025+)	The vehicle is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip. Such a design anticipates that the driver will provide destination or navigation input, but is not expected to be available for control at any time during the trip. This includes both occupied and unoccupied vehicles.

Fig. 4.2 Modified from NHTSA autonomous driving classification system



Source: Adapted figure based on Gartner's Hype Cycle (Gartner, August 2015) and Rogers Diffusion Curve.

Fig. 4.3 The innovation path of autonomous driving (adapted from: HYVE Science Labs 2015)

Marez Lieven and Gino 2004). AD is positioned right now at the peak of the curve within the enthusiasm phase, which is characterized by rapidly increasing content on the topic offered through various media channels such as TV, newspaper, magazines, and especially social media. The current peak of inflated expectations will be followed by a trough of disillusionment.

At this stage we will experience a rather low maturity of AD in the user domain. This is due, on the one hand, to the missing case scenarios of reasonable use and, on the other hand, to the lack of triability (e.g. test drives) of the new technology. Both factors represent mandatory requirements for customers' willingness to accept self-driving cars. Furthermore, according to the diffusion of the innovation model (Rogers 2003) the relative advantage over existing solutions, the compatibility with existing values, the relative complexity, and the observability will determine the pace of user acceptance and the course of the traditional *adoption curve* starting with the innovators and early adopters. A decisive point in Fig. 4.3 will be the entering stage of the diffusion curve. This is a familiar exercise for auto manufacturers when it comes to the introduction of a new car model. However, in this case self-driving cars cannot be treated simply as a new series. The innovation is too disruptive in all dimensions to do so. It has to be treated rather as the next wave of technology and a new S-curve companies need to jump on. Right now it seems that many automobile original equipment manufacturers (OEMs) are thinking of overcoming the entering stage of the diffusion model by simply continuing current car model strategies and at the same time scaling up advanced driver assistance systems until they arrive at a fully autonomous version. This intended seamless transition to AD may be attractive in preserving existing business models but won't be adequate for the degree of disruptiveness self-driving cars offer for new business opportunities.

4.3 Adding the Voice of the Crowd to Autonomous Driving

In order to understand the development of research in AD in the last years, it is important to take a look at the existing literature. The literature review of Rosenzweig and Bartl (2015) led to 399 peer-reviewed academic contributions identified form various academic literature databases. The search is based on specific terms such as "autonomous driving", "self-driving car", and "driverless car" either in the title, keywords, or abstract. The findings show a continuous increase of publications over time. In the last five years more articles have been published than in the whole two decades before. A journal count analysis shows that the IEEE, the world's largest professional association for the advancement of technology, contributed a large share of published contributions on the topic. This corresponds with the findings of a conducted topic analysis which reveals that more than 90 per cent of all publications focus on technology development of robotics, autonomous systems, vehicular technologies, and so on, while only 1 per cent of the published work has a research focus on user acceptance of AD. This lack of knowledge on the user perspective and their acceptance of the technology is currently the most pressing research gap and makes self-driving cars a showcase for a technology push innovation. This research study intends to add the user perspective by analyzing the largest existing user data set on autonomous driving which is formed by several hundreds of thousands of consumer statements in social media.

The development of the web and social media content has led customers to discuss their thoughts, opinions, experiences, and feelings online, creating a massive, publicly available data source (Egger and Lang 2013). This immense data source can help us identify sentiment, affect, subjectivity, and other emotional states in online text leading to new thrilling opportunities to understand the general public and consumers in almost every topic (Pang and Lee 2008). Social media analysis as a foresight method can detect emerging consumer needs long before the general public recognizes them (Chan and Franklin 2011; Keller and von der Gracht 2014; Olson et al. 2012). Web-monitoring methods are particularly appreciated for their holistic analyses, earliness and forwardness, and future orientation while including present aspects as well as the current context and exibility (Landwehr 2007). Particularly, social media content on a big data scale is appreciated as a useful source of information because it is the social web where critical discussions develop their own dynamics faster and on a broader reach than other forms of media (Francisco 2008; van Liere 2010). Thereby, the content of social media posts delivers nonredundant and diverse knowledge and information

(Hecht and Gergle 2010; Rodan 2010). Thus, social media monitoring is a tool for generating foresight during the emergence of an issue, trend, or topic such as autonomous driving. This vague information in its early phase of development is described as a weak signal (Ansoff 1975, 1980). Social media analysis can serve as an instrument to detect weak signals to which companies respond and upon which they base their decisions (Keller and von der Gracht 2014).

4.4 The Method of Innovation Mining

Innovation mining is a particular form of social media analysis that focuses on innovation-related topics (Bartl 2015; Bartl et al. 2016a). Whereas common web-monitoring techniques are mainly used to gather insights about brand perception or media impact (Croll and Power 2009; Egger and Lütters 2013), innovation mining aims to match technologies and product attributes with user applications and adoption behaviour. Considered mainly a quantitative method it relies heavily on key technologies such as artificial intelligence, automatic web information retrieval, and natural language processing for tracking and analyzing Internet content in search for patterns, trends, and valence (Pang and Lee 2008; Kruse et al. 2013). Web monitoring in general is not only considered one of the fastest-growing forms of media, but is also regarded as a scientifically well-grounded analysis of UGC (Egger and Lang 2013; Gensler et al. 2010). UGC is perceived as being impartial and unbiased, while it offers the chance to understand the needs and doubts of the potential customers as well as the used language within a certain topic (Egger and Lütters 2013). Characterized by extensive volunteering effort, lack of central control, and freedom of expression (Rheingold 1993), it creates a basis for identifying and understanding opinions, desires, tastes, needs, and decision-making influences of customers in a passive nonintrusive manner (Kozinets 2002). A vast part of UGC develops in online communities, which are considered as thematically focused platforms where knowledge is exchanged regarding specific product domains. Such communities work as meeting places for users to discuss new product ideas, opportunities, and product improvements (Kozinets 1999), where continuous discussion regarding opinions, attitudes, needs, and discontent concerning all kinds of topics, products, brands, and companies is expressed (Bartl et al. 2012; Egger and Lang 2013; Füller et al. 2006).

Figure 4.4 shows the five-step approach of innovation mining. Whereas the first and the last steps are specifically aligned to the context of innovation, steps two to four represent the commonly applied core process of social media analysis (Egger and Lang 2013).

The search and collection process focuses on the gathering of textual content, available in sites open for public reading access and based on Information Retrieval (IR) (Robertson 1981). In order to gather all the possible UGC regarding AD, the web-monitoring tool InMap was used for the study at hand, provided by the technology company Insius and developed in cooperation with the University of Cologne. Like other IR systems available to the public, such as Google, Bing, Yahoo, Twitter, and Facebook, the search tool uses Boolean keyword combination-based query language. In order to analyze the field of AD a general search including "autonomous driving" OR "self-driving car" OR "self-driving cars" OR "driverless car" OR "driverless cars" OR "autonomous vehicle" OR "automated driving" OR "piloted driving" was conducted. The search led to around 471,000 documents including one or more of the searched terms within user posts. The search concentrated on the English language without any specific geographical limitation with a focus on

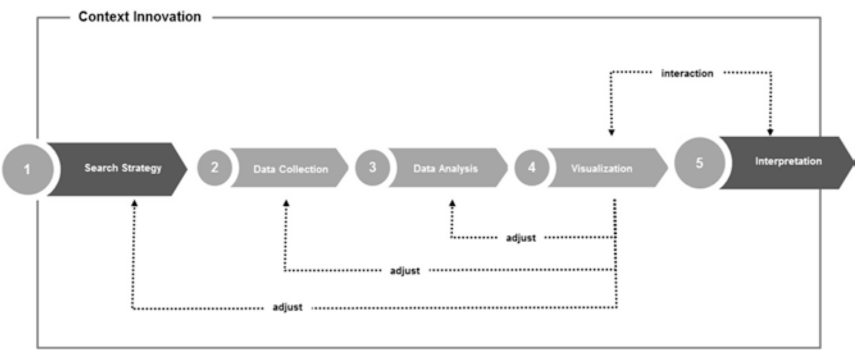


Fig. 4.4 The five-step approach of innovation mining

including technology affine UGC sources. The social chat on Facebook and Twitter was not included in this data set.

As the search results include content published by editors as well content created by users a further sub-step was required referred to as "clean-up". In this sub-step a manual selection process of irrelevant websites where UGC cannot be found such as patent sites, research sites, and so on are eliminated, which led to the reduction of the results to around 381,000 documents. Furthermore, as in the available sources UGC and non-UGC are still combined, the remaining results are analyzed to see if they fulfil the three main aspects of UGC by the Organisation for Economic Co-operation and Development (OECD) (publication requirement, creative effort, and creation outside professional routines and practices; OECD 2007), while also eliminating duplicates. Finally, machine-learning techniques were utilized. For this process 1529 results were manually analyzed, classified, and used to train a classifier to order the rest of the unclassified texts. This final sub-step led to a total of 106,305 documents defined as UGC with a precision of 99.5 per cent, and a recall of 91.5 per cent (see formulas presented in Fig. 4.5).

$$\text{precision} = \frac{\left|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}\right|}{\left|\{\text{retrieved documents}\}\right|}$$

$$\text{recall} = \frac{\left|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}\right|}{\left|\{\text{relevant documents}\}\right|}$$

After the search and collection procedure, the in-depth analysis of the 106,305 retrieved documents followed. The process of the analysis is described in this section while actual results are presented in the following chapter. As a first step documents are analyzed to be classified on the overall sentiment, based on document sets where the general sentiment is

$$precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|}$$

$$recall = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{relevant\ documents\}|}$$

Fig. 4.5 Precision and recall formulas known due to a star system ranking, manual classifications, or sentiment ratings. Also, unsupervised machine-learning approaches can be trained to classify with the available data either in a binary approach (positive or negative) or continuous between two bounds. The second step focuses on a more detailed analysis, breaking the documents into smaller textual entities first, such as sentences, phrases, or words. Using tokenization techniques, sets of text can be divided according to specific rules in order to identify particular words, word phrases, sentences, or passages taking into consideration special exemptions such as abbreviations or enumerations. In addition, in this step the bag-of-words method can be utilized along with word frequency distributions to eliminate information that is not valuable such as articles ("a", "an"), conjunctions ("and", "or"), and direct speech ("I", "you", "me", etc.) if necessary.

Furthermore, sentence detection techniques have to be applied to break down documents into coherent sentences, which can be further analyzed in a highly simplified explanation through aspects that are represented by nouns (e.g. car) and sentiments that are represented by adjectives or adverbs (e.g. good, bad, poor, etc.). To identify the nouns, adjectives, and adverbs, Part-of-Speech Taggers (POS-Tag) from natural language processing can be used where each word is classified into its respective category (nouns, adverbs, adjectives, conjunctions, etc.) and then mapped. This classification then leads to several possible sets of combinations such as "Noun-conjunction-noun-verb-adverb-adjective" that would match "Design and quality was very good" (Egger and Lang 2013), reflecting that sentences including both nouns and adjectives are regarded as candidates for bearing customer opinions. Other aspects such as distance between different words are also analyzed. Further rules apply to understand the sentiment of a sentence such as polarization, meaning the inclusion of a negation (e.g. "not good" or endings such as "n't") where the sentiment although having a positive adjective (good) changes to a negative perception. The basic idea is the analysis and understanding of different word combinations or POS-Tags as complete sets of words. After the determination of the opinion, the candidate's further normalization, aggregation, and pruning steps are performed for summarization. Finally, the results are also analyzed on a world level, through the analysis of word frequency distributions within and across documents, smaller text entities that prove helpful to gain insights into topics, most relevant vocabulary within a topic, as well as specific vocabulary such as automotive brands names.

Finally, after the analysis of the results is concluded, proper visualization techniques to understand the gathered data such as tag clouds, network representation, pie and bar charts, line graphs, data series plots, and bubble charts are used. These give the ability to define selected information for a thorough analysis through specified time frames, within certain phrases, through geographical regions, to differentiate among sources, and so on. The results of the search are presented in detail in the following chapter.

4.5 Study Results

In order to find out how users actually refer to autonomous driving a frequency analysis of the most used terms was conducted as presented in Fig. 4.6. Despite the fact that the term "autonomous driving" is more accepted in the academic literature on the topic, also encapsulating a broader technical perspective, the results showed that from the users' perspective "driverless car" or "self-driving car" are much more used and accepted terms. "Driverless car" can be observed to be the most popular term with 22,383 mentions throughout the data set. Such insight is highly valuable for companies in order to align product naming, communication strategy, and market introduction activities.

Understanding the most influential social media sources is valuable information in order to recognize where users talk about AD in the web, where the most engaged types of customers can be found, and how influence structures work. The vast majority of quotes are distributed in many diverse sources (see Fig. 4.7). A high volume of discussion regarding AD takes place in Reddit.com, which is the most impactful single source on the topic with around 10 per cent of the total discussion. It contains ten times more customer quotes than the second-placed UGC source quora.com with nearly 1 per cent, followed by arstechnica.com, with similar impact on the topic as ask.fm. Reddit is a leading social news aggregator on the Internet. Arstechnica is a technology, news, and information website

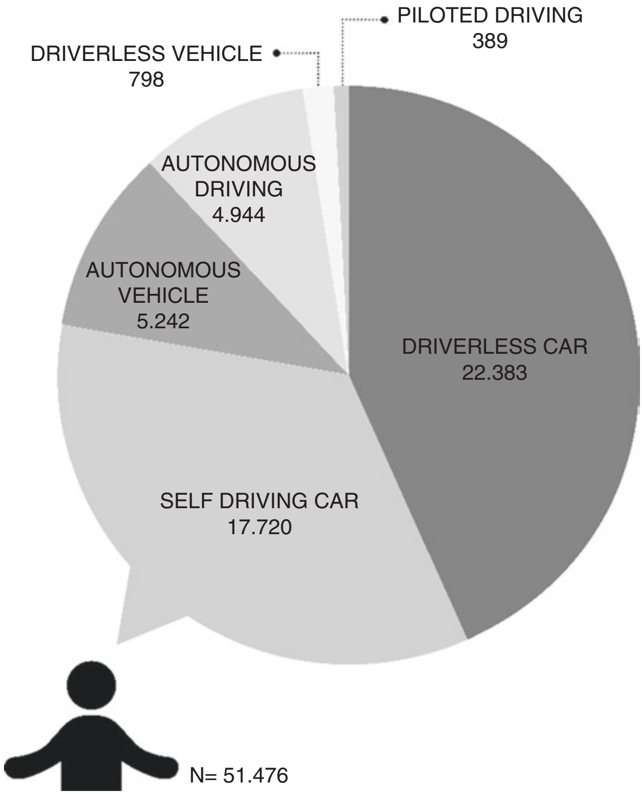


Fig. 4.6 Nomenclature analysis (adapted from: HYVE Science Labs 2015)

that publishes news, reviews, and guides where the writers are mainly postgraduates and research institution workers. They are both considered as technology-focused media where early adopters and most knowledgeable users of technology can be found while also being regarded as top technology discussion sites. Quora, on the other hand, is a question-and-answer website where questions are asked, answered, edited, and organized by its community of users. Recently valued at nearly $1 billion, it is among the top 200 websites globally.

Moreover, analyzing the topic evolution over time, it can be observed that there has been a strong and rapid growth since 2010. Compared to the literature review analysis (Rosenzweig and Bartl 2015), which has a

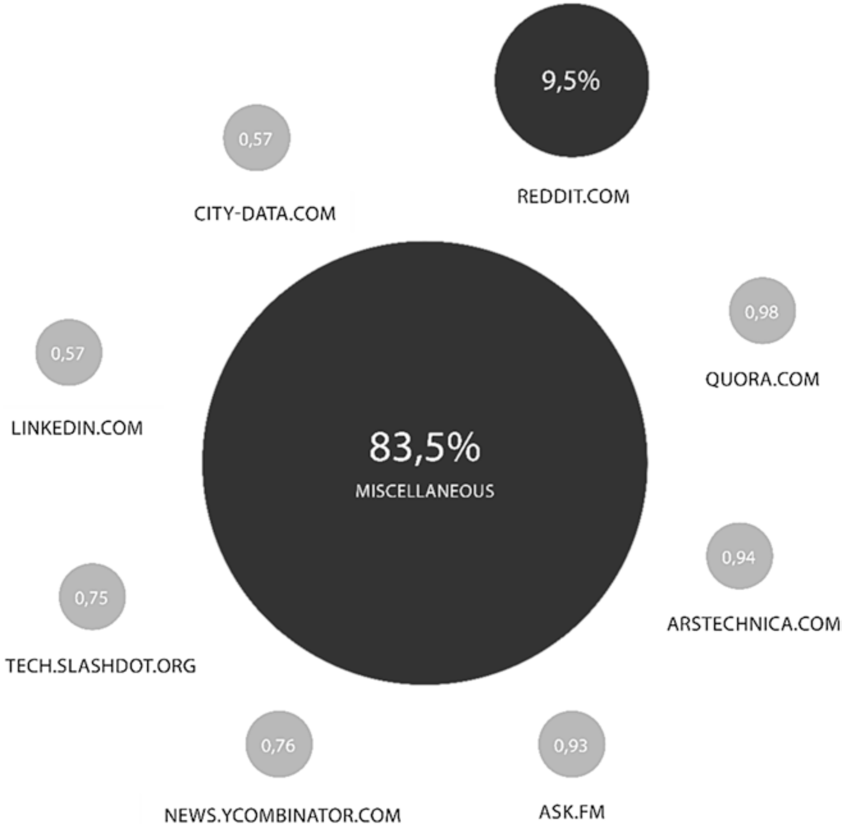


Fig. 4.7 UGC data sources (adapted from: HYVE Science Labs 2015)

steadier and constant growth of publications over the years, this can be due to the fact that posts and news have much more sudden reactions compared to the formality and reaction time of the academic literature on the topic. The interest in the topic can also be related to the official Google driverless-car debut in 2010. Since then the volume of the AD discussion has been doubling every year. In a more detailed monthly view shown in Fig. 4.8, two peaks, the first in May 2014 and the second in March 2015, can be found. Relating this high UGC volume to the main sources, the most impactful single discussions can be identified. The highest volume in customer discussion can be observed in March

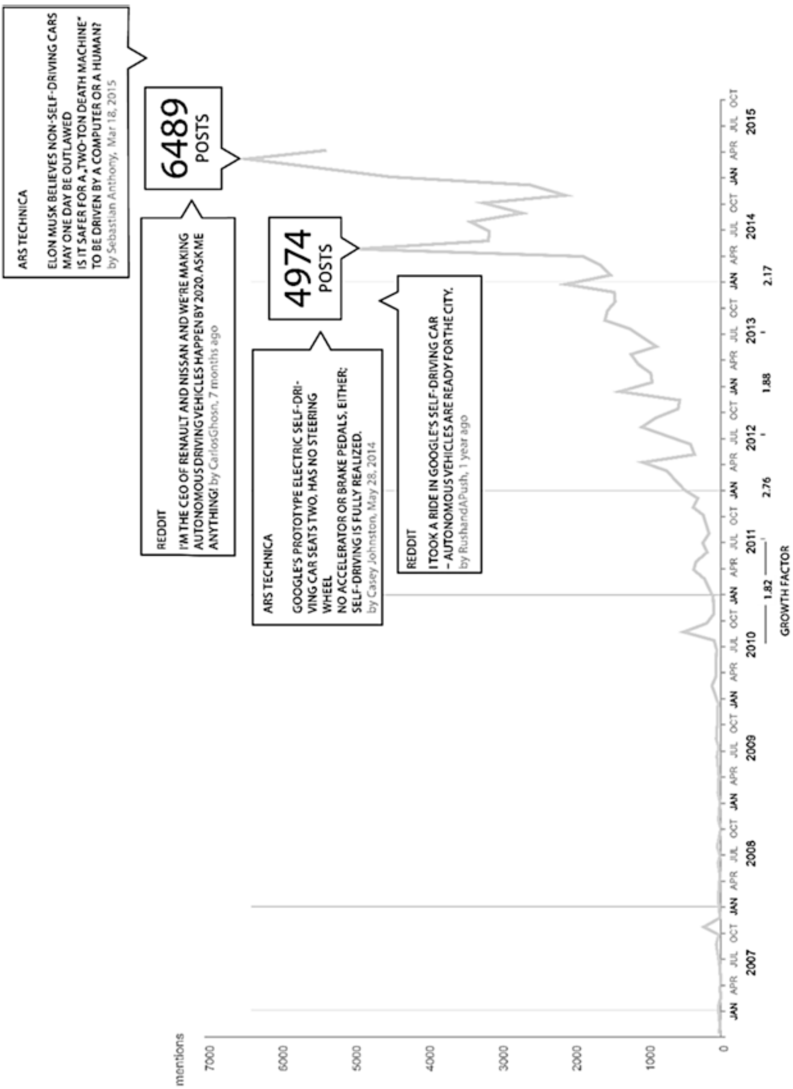


Fig. 4.8 Consumer conversation over time (adapted from: HYVE Science Labs 2015)

2015 with 6489 posts. This was triggered partly by Elon Musk's (CEO of Tesla Motors) comment that non-self-driving cars would someday be outlawed. In the same month the most impactful discussion was when Carlos Ghosn (CEO of Renault and Nissan) opened a forum of open questions regarding the topic and promoted it through Twitter. The second month with the highest volume of discussion is May 2014 with 4974 posts when Google revealed its prototype Google Car for the first time. Another impactful discussion in that month includes an article from a reporter that "took a ride in Google's self-driving car" and was debated in Reddit.

In order to understand how people talk about AD in more detail, the inherent concepts of the 106,305 documents were further analyzed with the help of a network map. "Concepts" are defined as nouns which carry a polarization, being either positive, negative, or neutral, depending on the adjectives or adverbs (sentiment carriers) with which they are mentioned in a sentence. The most mentioned "concepts" within the concept map (see Fig. 4.9) are positioned nearer to the centre while the importance decreases towards the outside. The number in the circles represents the recurrence of each concept in percentage. The polarization of the "concepts" is reflected through colours. Yellow are neutral, red are negative, and green are positive concepts. Therefore it can be observed that the most important concepts in the AD field are discussed with a neutral position and are without surprise "car", which is mentioned in 76 per cent of the comments, followed by "vehicle" in 20 per cent of the comments, and "technology" in 10 per cent of the comments. An interesting fact to state is that positive Internet discussions occur twice as much as negative ones. This is in contrast to many of the traditionally conducted and survey-based market research studies, which show a much more sceptical picture of consumer attitudes towards AD. This may be explained by the fact that UGC is contributed by (lead) users who generally like to deal with innovative and advanced topics (Jeppesen and Laursen 2009). They face needs and requirements months or even years before the bulk of the marketplace encounters them.

Deeper understanding on the positive, negative, and neutral concepts is achieved through contextual and background information given to each concept by the concept drivers. Drivers are defined as the polarized words

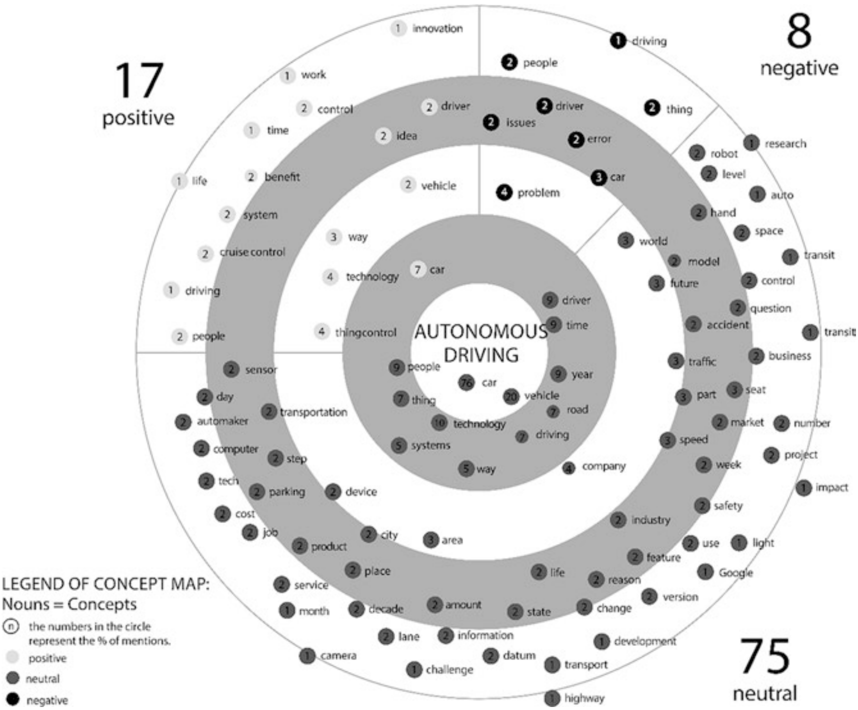


Fig. 4.9 Concept network map including the 100 most important concepts of AD (adapted from: HYVE Science Labs 2015)

co-occurring with the respective concept. These drivers act as indicators of readiness of the technology from the user perspective. The results of the driver analysis are shown in Fig. 4.10. Within the top positive concepts the main drivers that make this concept positive are "fully autonomous", "smart", "safe", "modern", and so on. On the other side the appearance of the driver "expensive" among the top negative concepts reflects the current user perspective of the still high prices of technology. The main other negative drivers are "less", "average", "inevitable", and "dangerous". The driver "electric" appears highly ranked in the top neutral concepts signalizing the continuous mentioning of electric gears in the AD context.

To analyze the competitive field of AD it is of high importance to understand how traditional automotive industry stands versus digital players. A frequency analysis of the competing brands developing AD

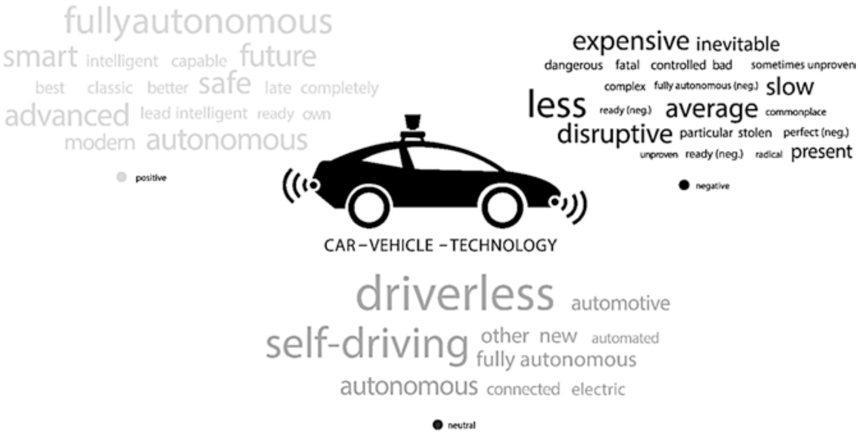


Fig. 4.10 Positive, negative, and neutral drivers for the concepts "car", "vehicle", and "technology" (adapted from: HYVE Science Labs 2015)

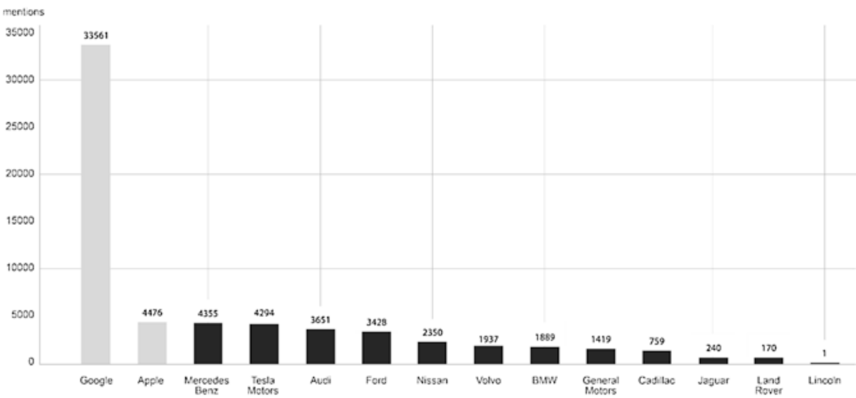


Fig. 4.11 Brand name mentions within posts on AD (adapted from: HYVE Science Labs 2015)

technology was conducted in order to understand the presence of each of the main brands in the public online discussion. Figure 4.11 reflects the number of times each brand is mentioned throughout all the documents, either with a sentiment or without a sentiment in the sentence.

The data clearly reflect the intense efforts of Google to be the first to offer the technology and to be successful in communicating its advancements to the public, as well as positioning itself as the technological leader in the users' perception. Google manages to be mentioned more frequently than all other brands together with a total of 33,561 occurrences. Furthermore, it is important to notice that Apple, despite not having officially made any announcements regarding AD and being supported only by rumours, comes in second place with 4476 mentions. This makes the two biggest digital players the leading brands in the AD discussion. Tesla, a third player from the Silicon Valley, is positioned in the top five followed by the majority of the traditional automotive players of which only Mercedes and Audi are positioned in the top ranks.

A concluding analysis was dedicated to the positive concept "time" with its most influential driver "free time". This driver demonstrates the consciousness of the customer of one of the main benefits that AD technologies bring: to engage in different activities rather than driving. The potential of regained time in the car is huge, that is the average US driver spends around 465 hours per year driving a car (Frazzoli 2014). Hence, the question is what users would do with their new free time? The most frequently mentioned activities by users reveal the word "Internet" and "Email", which are mentioned more than 8000 times (see Fig. 4.12). The concepts "TV" and

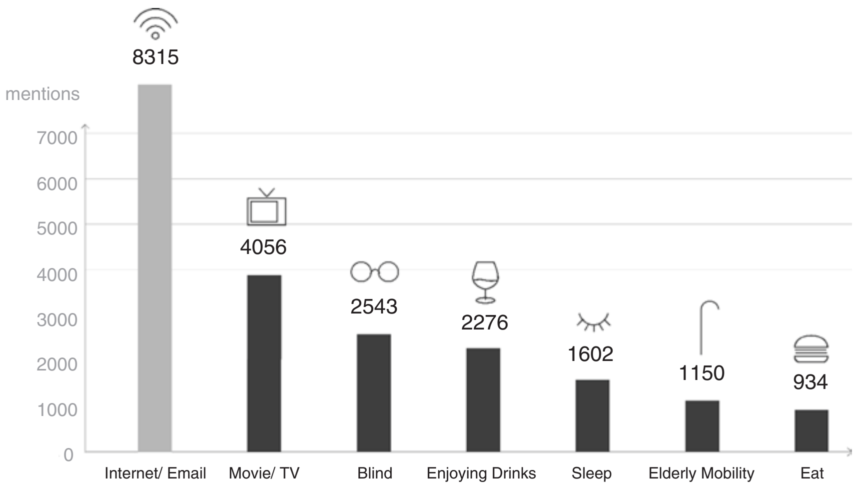


Fig. 4.12 Activity mentions related to AD (adapted from: HYVE Science Labs 2015)

"Movie" reflect another main category of activities the users discuss. The extensive chat regarding blind people driving cars can be directly related to Google's successful campaign showing a blind man taking a ride to a Taco Bell in a self-driving car. Surprising is the high volume of discussion on the wish of enjoying a glass of beer or wine ("Enjoying Drinks") and enabling elderly people to win back some of their lost freedom of movement. Hence, automated mobility seems to become an important catalyst of future social interaction. The categories of sleeping and eating complete the more than 20,000 mentions on possible activities in a self-driving car.

4.6 Outlook

The field of automotive innovation is notably radical because it is embedded in a whole ecosystem where traditional players and nontraditional players are involved. Some known players in the game are OEMs, suppliers, logistics, and passenger and freight transportation. In addition to the advantage of fewer accidents one major outcome of autonomous driving will be that the occupant will have a substantial and increasing free time budget within the car. The competition to get a share of the consumer's time budget has already begun. Just as a thought experiment, think about what known industries could do to serve their customers in the setting of fully autonomous cars. Warner Brothers, Barnes & Nobel, EA Games, and publishing houses want to bring their entertainment offerings and offered content to the car to reach a new audience. As shown in Fig. 4.11 companies like Google and Apple have already great prospects of bringing their services into the self-driving car. Car rentals or companies such as Uber see new opportunities for passenger transportation. Insurances develop new offerings. Designers and manufacturers of furnishings could think of new concepts to revolutionize the interior of cars. Fast food restaurants or food delivery services could utilize autonomous distribution services. The same applies to logistics companies such as DHL, Fedex, and UPS. The energy providers could think of movable charging stations which can load, store, and release energy as part of a revolutionary decentralized energy system. Steel and carbon companies can think of totally new coachwork concepts. Nestlé and

Unilever can invent new healthy snacking concepts for break time. Travel Agents such as Thomas Cook and TUI will have a huge playground to offer amazing experience routes or traveling arrangements where one is driven overnight to the holiday destination of one's choice. A provider of office accommodation can develop fleets of mobile offices to add them to their mortar and brick office offerings. When you think about all this through experiment and the hypothetical scenarios derived, it becomes obvious that AD cannot be mastered by a single industry. There will be new partnerships, collective efforts, business models, and joint developments and ventures to profit from the new time budgets consumers can spend on competing activities such as working, relaxing, travelling, being entertained, eating, and so on. Based on these thoughts the authors want to term AD the master class of open innovation. One needs to utilize a collective brain and collective problem solving to handle all the new dimension of the demand side. Right now it really seems to be a field predominantly driven by technology push. Strongly missing is the consumer perspective as well as thoughts on use case scenarios and business model generation. This innovation mining study adds the perspective of the crowd to address the existing research gap.

An important future challenge in innovation management will be to find the right role for social media in the process. One key to answering this challenge will be to pick the best available mix of approaches that fits each company individually. Generally, innovating companies have two basic options. First, they can actively involve and engage customers using cocreation and crowdsourcing techniques. Second, they can utilize already existing user-generated data pools. Both netnography and innovation mining belong to this second group of exploration methods with the aim of analyzing, interpreting, and integrating the voice of the crowd in innovation activities. The two techniques form a symbiosis with netnography emphasizing the qualitative dimension of generating in-depth consumer understandings based on smaller data samples and intensive manual work of the researcher. Innovation mining is emphasizing the quantitative dimension, especially regarding data volume and processing speed. Using a metaphor, netnography is the tool which could be characterized as a microscope whereas innovation mining represents the telescope. In times where the amount of accessible user data is rapidly increasing year by year up to an inconceivable volume of data, automatized approaches to utilize the voice of the crowd will naturally gain attention. In the context of innovation management the first studies on innovation mining in the field of autonomous driving, augmented reality, and chemicals proved to have very promising results in detecting weak signals and exploiting information arbitrage to support foresight and a user-centric view on technology acceptance (Stockinger 2015; Bartl 2015; Francisco 2008). However, there are still some inherent limitations in a process of automated web crawling based on natural language-processing algorithms. The aggregated format still cannot always reflect human intelligence and language skills to read "between the lines" and interpret consumer insights. There will be an equilibrium of quantitative and qualitative research to fully utilize social media for innovation with a huge need to catch up with quantitative methods in times where data is called the new oil in a digital economy.

References

Ansoff, Harry I. 1975. Managing Strategic Surprise by Response to Weak Signals. *California Management Review* 18(2): 21–33.

Ansoff, Harry I. 1980. Strategic Issue Management. *Strategic Management Journal* 1(2): 131–148.

Bartl, Michael. 2015. Innovation Mining. In *Conference Proceedings XXVI ISPIM Conference 2015*. Shaping the Frontiers of Innovation Management, Budapest, Hungary.

Bartl, Michael, Johann Füller, Hans Mühlbacher, and Holger Ernst. 2012. A Manager's Perspective on Virtual Customer Integration for New Product Development. *Journal of Product Innovation Management* 29(6): 1031–1046.

Bartl, Michael, Oliver Gluth, Rainer Rieger, and Harald Schmidt. 2016a. Innovation Mining. *Zeitschrift für Ideen- und Innovationsmanagement*, February 16.

Bartl, Michael, Vijai Kumar, and Hanna Stockinger. 2016b. A Review and Analysis of Literature on Netnography Research. *International Journal of Technology Marketing* 11(2): 165–196.

Bilgram, Volker, Alexander Brem, and Kai-Ingo Voigt. 2008. User-Centric Innovation in New Product Development—Systematic Identification of

Lead Users Harnessing Interactive and Collaborative Online-Tools. *International Journal of Innovation Management* 12(3): 419–458.

Brem, Alexander, and Volker Bilgram. 2015. The Search for Innovative Partners in Co-creation: Identifying Lead Users in Social Media Through Netnography and Crowdsourcing. *Journal of Engineering and Technology Management* 37: 40–51.

Chan, Samuel W.K., and James Franklin. 2011. A Text-Based Decision Support System for Financial Sequence Prediction. *Decision Support Systems* 52(1): 189–198.

Chesbrough, Henry. 2003. The Era of Open Innovation. *MIT Sloan Management Review* 44: 35–41.

Croll, Alistair, and Sean Power. 2009. *Complete Web Monitoring*. Sebastopool: O'Reilly Media.

Cui, Anna S., and Fang Wu. 2016. Utilizing Customer Knowledge in Innovation: Antecedents and Impact of Customer Involvement on New Product Performance. *Journal of the Academy of Marketing Science* 1–23.

De Marez Lieven, S.B., and Verleye Gino. 2004. ICT-Innovations Today: Making Traditional Diffusion Patterns Obsolete, and Preliminary Insights of Increased Importance. *Telematics and Informatics* 21(3): 235–260.

Egger, Marc, and André Lang. 2013. A Brief Tutorial on How to Extract Information from User-Generated Content (UGC). *KI-Künstliche Intelligenz* 27(1): 53–60.

Egger, Marc, and Holger Lütters. 2013. Listening Is the New Asking: Social Media-Analyse in der Marktforschung. *Transfer Werbeforschung & Praxis* 59(4): 35–41.

Fenn, Jackie. 2012. Hype Cycle for Emerging Technologies. *Gartner Research Retrieved* 7(24).

Francisco, Phil. 2008. Information Arbitrage: Gaining Competitive Advantage Through Data Analytics. *Information Management*, December 22.

Frazzoli, Emilio. 2014. Can We Put a Price on Autonomous Driving? *MIT Technology Review*, March.

Füller, Johann, Michael Bartl, Holger Ernst, and Hans Mühlbacher. 2006. Community Based Innovation: How to Integrate Members of Virtual Communities into New Product Development. *Electronic Commerce Research* 6(1): 57–73.

Gemser, Gerda, and Helen Perks. 2015. Co-Creation with Customers: An Evolving Innovation Research Field. *Journal of Product Innovation Management* 32(5): 660–665.

Gensler, Sonja, Franziska Völckner, Marc Egger, Kai Fischbach, and Detlef Schoder. 2010. Listen to Your Customers! Using Consumer-Generated Content to Elicit Brand Image. In *Proceeding of the 39th EMAC Conference*.

Hecht, Brent, and Darren Gergle. 2010. The Tower of Babel Meets Web 2.0: User-Generated Content and Its Applications in a Multilingual Context. In *CHI'10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 291–300.

Howe, Jeff. 2006. The Rise of Crowdsourcing. *Wired Magazine* 14(6): 1–4.

HYVE Science Labs. 2015. Autonomous Driving—The User Perspective. October.

Jeppesen, Lars B., and Keld Laursen. 2009. The Role of Lead Users in Knowledge Sharing. *Research Policy* 38(10): 1582–1589.

Keller, Jonas, and Heiko A. von der Gracht. 2014. The Influence of Information and Communication Technology (ICT) on Future Foresight Processes: Results from a Delphi Survey. *Technological Forecasting and Social Change* 85: 81–92.

Kozinets, Robert. 1999. E-Tribalized Marketing? The Strategic Implications of Virtual Communities of Consumption. *European Management Journal* 17(3): 252–264.

Kozinets, Robert V. 2002. The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities. *Journal of Marketing Research* 39(1): 61–72.

Kruse, Paul, Andreas Schieber, Eric Schoop, and Andreas Hilbert. 2013. Idea Mining—Text Mining Supported Knowledge Management for Innovation Purposes. In *Proceedings of the Nineteenth Americas Conference on Information Systems*.

Landwehr, Katja. 2007. *Strategische Technologiefrühaufklärung: Grundlagen, Systematik und Methoden*. Saarbrücken: VDM Verlag Dr. Müller.

Linden, Alexander, and Jackie Fenn. 2003. Understanding Gartner's Hype Cycles. *Strategic Analysis Report No. R-20-1971*. Stamford: Gartner Inc.

Moore, Geoffrey A. 1991. *Crossing the Chasm—Marketing and Selling High-Tech Products to Mainstream Customers*. New York: Harper Business Press.

National Highway Traffic Safety Administration. 2013. Preliminary Statement of Policy Concerning Automated Vehicles 2013. Accessed December 2015. http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf

OECD. 2007. Participative Web: User-Created Content: Web 2.0, Wikis and Social Networking. Directorate for Science, Technology and Industry, Committee for Information, Computer and Communications Policy, October 2007.

Olson, David L., Dursun Delen, and Yanyan Meng. 2012. Comparative Analysis of Data Mining Methods for Bankruptcy Prediction. *Decision Support Systems* 52(2): 464–473.

Pang, Bo, and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 2(1–2): 1–135.

Prahalad, C.K., and Venkat Ramaswamy. 2000. Co-Opting Customer Competence. *Harvard Business Review* 78: 79–91.

Rheingold, Howard. 1993. *The Virtual Community: Homesteading on the Electronic Frontier*. Boston: MIT Press.

Robertson, Stephen E. 1981. The Methodology of Information Retrieval Experiment. *Information Retrieval Experiment* 9–31.

Rodan, Simon. 2010. Structural Holes and Managerial Performance: Identifying the Underlying Mechanisms. *Social Networks* 32(3): 168–179.

Rogers, Everett. 2003. *Diffusion of Innovation*. New York: Simon and Schuster.

Rosenzweig, Juan, and Michael Bartl. 2015. A Review and Analysis of Literature on Autonomous Driving. *E-Journal Making-of Innovation*, October.

Stanley, Morgan. 2013. Autonomous Cars: Self-Driving the New Auto Industry Paradigm, November 2013.

Stockinger, Hanna. 2015. Consumers' Perception of Augmented Reality as an Emerging end User Technology: Social Media Monitoring Applied. *Künstliche Intelligenz* 29(4): 419–439.

van Liere, Diederik. 2010. How Far Does a Tweet Travel? Information Brokers in the Twitterverse. In *MSM'10 Proceedings of the International Workshop on Modeling Social Media*, 1–4.

von Hippel, Eric. 2005. *Democratizing Innovation*. Cambridge: MIT Press.

Wiles, Rose, Andrew Bengry-Howell, Graham Crow, and Melanie Nind. 2013. But is it innovation? The Development of Novel Methodological Approaches in Qualitative Research. *Methodological Innovations Online* 8(1): 18–33.

Zhang, Yi, Ying Guo, Xuefeng Wang, Donghua Zhu, and Alan L. Porter. 2013. A Hybrid Visualisation Model for Technology Roadmapping: Bibliometrics, Qualitative Methodology and Empirical Study. *Technology Analysis & Strategic Management* 25(6): 707–724.