Horizon scanning and analysis of techno-scientific trends

Scientific Foresight Study
Horizon scanning and analysis of techno-scientific trends

Scientific Foresight Study
July 2017

Abstract
This horizon scan has identified a selection of technology-related topics with a high presence in social media and news. Further, a scan was conducted to measure controversy on social media, which constituted an initial controversy ranking of 24 technology-related topics: controversy between different stakeholders could be a parameter for legitimising a topic’s relevance for investigation by STOA. Next, a set of topics with relatively high controversy between different stakeholders were selected, which have not yet been investigated in-depth by STOA. These topics are big data, gene technology, electric vehicles, autonomous cars and impact of algorithms.

Furthermore, other trending topics with a high STOA relevance with high potential impact on society were also selected for analysis: screen addiction, fake news and bioterrorism.

In total eight trending technology-related topics have been analysed, following detection of subtopics and sentiments in social media (Twitter) and news articles.

These eight technologies and social trends are areas for possible discussion amongst the STOA Panel members when considering new activities to be undertaken. The analyses of technology-related trending topics illustrate how data analytics using a combination of mind and machine methodologies could support the selection of emerging technologies and other trend areas for policy advisory purposes.
The STOA project ‘Horizon scanning and analysis of techno-scientific trends’ was carried out by Augmented Intelligence Institute at the request of the Science and Technology Options Assessment Panel, and managed by the Scientific Foresight Unit (STOA) within the Directorate-General for Parliamentary Research Services (DG EPRS) of the European Parliament.

AUTHORS

Michael Baumgartner, Bijan Farsijani (Augmented Intelligence Institute; http://www.augmento.ai)

STOA ADMINISTRATOR RESPONSIBLE

Lieve Van Woensel
Scientific Foresight Unit (STOA)
Directorate for Impact Assessment and European Added Value
Directorate-General for Parliamentary Research Services
Lieve.vanwoensel@ep.europa.eu

LINGUISTIC VERSION

Original: EN

ABOUT THE PUBLISHER

To contact STOA or to subscribe to its newsletter please write to STOA@ep.europa.eu
This document is available on the Internet at: http://www.ep.europa.eu/stoa/

Manuscript completed in July 2017
Brussels, © European Union, 2017

DISCLAIMER

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy.

PE 603.183
doi: 10.2861/244538
QA-01-17-579-EN-N
Table of Contents

EXECUTIVE SUMMARY ........................................................................................................................5

1. INTRODUCTION .............................................................................................................................10

2. METHODOLOGY ..............................................................................................................................13
   1.1. Target of analysis .......................................................................................................................13
   1.2. Data ........................................................................................................................................13
   1.3. Identification of trending topics ...............................................................................................14
   1.4. Stakeholder classification .........................................................................................................14
   1.5. Sentiment analysis ...................................................................................................................14
   1.6. Phrase extraction ......................................................................................................................14
   1.7. Phrase selection and clustering ...............................................................................................15
   1.8. Visualisation .............................................................................................................................15
   1.9. Provision of context (tweets, text snippets) ............................................................................15
   1.10. Controversy ranking ..............................................................................................................15

3. PRESENTATION OF THE FINDINGS ...............................................................................................16
   3.1. BIG DATA ...............................................................................................................................17
       3.1.1. Trending topics in news on Big Data ................................................................................17
       3.1.2. Trending topics of most positive and negative tweets on Big Data .................................23
       3.1.3. Stakeholder specific trending topics – Big Data ..............................................................26
       3.1.4. Big Data stakeholder activity ..........................................................................................28
   3.2. GENE TECHNOLOGY ...............................................................................................................29
       3.2.1. Trending topics in news on Gene Technology .................................................................29
       3.2.2. Trending topics of most positive and negative tweets on Gene Technology ..................32
       3.2.3. Mentions of CRISPR/Cas9 and CRISPR/Cas on social media .........................................34
       3.2.4. Stakeholder specific trending topics – Gene Technology .................................................35
       3.2.5. Gene Technology stakeholder activity .............................................................................37
   3.3. ELECTRIC VEHICLES ..............................................................................................................38
       3.3.1. Trending topics in news on Electric Vehicles .................................................................38
       3.3.2. Trending topics of most positive and negative tweets on Electric Vehicles .....................43
       3.3.3. Stakeholder specific trending topics – Electric Vehicles ..................................................45
       3.3.4. Electric Vehicles stakeholder activity ..............................................................................47
EXECUTIVE SUMMARY

Horizon scanning gives an overall view of techno-scientific topics to support the development of strategies for anticipating future advances.

A horizon scan was conducted to identify the most ‘trending’ techno-scientific topics and – amongst these – to single out the most controversial ones. Out of these, a set of STOA-relevant topics were selected, although these have not yet been investigated in depth within STOA. These topics are big data, gene technology, electric vehicles, autonomous cars and impact of algorithms.

Three further trending topics with high potential impact on society were selected for analysis: screen addiction, fake news and bioterrorism.

Within the eight selected topics, the analysis unraveled the complexity of the trending topics through a diverse range of facets. This analysis covers: Identification of areas of application, purposes and concerns related to these applications, detection and analysis of related technologies, detection and analysis of sentiments, as well as the geographical regions in which the respective topics were trending.

In this initial and exploratory study, Augmented Intelligence Institute – referred to hereafter as 'the experts' – used data analytics, i.e. several elements of their artificial intelligence toolkit, combined with a ‘journalistic’ approach (a combination of mind and machine-based data analytics) to collect and analyse 16 491 news articles and over 8.3 million tweets.

This blend of methodology made it possible to track the sentiments of citizens and various stakeholder groups such as politicians and companies. The most controversial trend phenomena have been identified. Further, the focus of this study to explain the WHY behind the data points by providing sample tweets and sample headlines (only available through the Augmento web app¹) combined with context information.

In a first pre-scanning phase, together with the STOA administrator, 24 technological and social trending phenomena were identified, the controversy of which should be evaluated.

While topics like Artificial Intelligence and Electric Vehicles ranked high when it comes to the amount of tweets (Figure 1), topics like Waste Management are discussed more controversially (Figure 2).

¹ To get access to this web platform, Members can request a username and password at stoa@europarl.europa.eu.
Figure 1: Social media activity per topic

Figure 2: Controversy detection graph
Figure 3 shows the controversy ranking, taking into account both the amount of tweets and sentiment divergence score.

**Figure 3: Controversy ranking**

The final selection of eight technological and social trends for this analysis results from a combination of five topics identified by the controversy analysis (Figure 3) with a high degree of potential societal impact and in which STOA has not conducted studies so far, and three additional trending topics which have recently been suggested by MEPs as important and relevant for uptake by STOA.

The eight analysed topics are:

1. Big Data;
2. Gene Technology;
3. Electric Vehicles;
4. Autonomous Cars;
5. Impact of Algorithms;
6. Screen Addiction;
7. Fake News, and
8. Bioterrorism.
For each topic, both news articles and social media data (Twitter) have been explored to investigate the trending topics in diverse aspects, covering possible application, technologies used, opportunities and concerns, etc.

The analysis of this selection of technology-related trends illustrates how a combination of human- and machine-based data methodologies could unravel a wide range of facets in each topic, as a result offering support in decisions over the suitability for investigation of these areas in a wider STOA study.

1. The analysis of news covering **Big Data** revealed that a major part of the coverage revolves around the link between Big Data and other technologies, and the application of Big Data in the private as well as in the public sector. *Business intelligence* and *health care* were the most significant examples in the dataset. The sentiment towards Big Data, measured by Twitter data, was mostly optimistic given the potential value various stakeholders expect from this trend. Negative discussions were mostly shaped by debates around Big Data’s use for campaigning. *Cambridge Analytica*, a company that uses data to change audience behaviour and *Robert Mercer*, co-CEO of Renaissance Technologies caught the most attention in this debate.

2. News on **Gene Technology** mostly focused on impact areas such as the influence of genomics on public health or the genome editing of human embryos. *CRISPR-Cas9* caught much attention through scientific experiments such as a genome surgery with *CRISPR-Cas9* to prevent blindness. Besides that, more negative articles raised ethical questions as well as warnings of potential dangers and unknown outcomes. Both positive and negative sentiments are well balanced on Twitter: Criticism was mainly directed towards the effects of GMOs on our food chain.

3. Much of the media discussion around **Electric Vehicles** is dominated by announcements of various car manufacturers. While traditional brands like *BMW* or *Nissan* do play a role, the undefeated champion in the race for attention remains *Tesla*. Price, speed and range turned out still to be the most important choice factors for going electric. **Electric Vehicles** are often seen as a *climate friendly* way of transport. They are seen as an opportunity to reduce emissions and improve air quality in cities. However, it has also been mentioned that if EVs are charged with fossil-fuel-generated energy the effect would be negative for the environment compared to conventional transport. That is why the topic of **Electric Vehicles** is closely related to *green energy*. On Twitter, positive sentiments towards electric vehicles clearly outweigh negative ones as tweets are mostly related to *green energy*, an optimistic outlook on the future and subsidy programmes like the *Norwegian incentive system*.

4. News on **Autonomous Cars** was shaped by future-oriented technology announcements and future impact areas, such as changing business models within the automotive industry and the potentially positive effect on traffic jams. *Uber, Google* and *Tesla* were getting considerably more attention than traditional car manufacturers. On social media the sentiment distribution was well balanced. Negative sentiments were mostly driven by the fear of *autonomous car crashes* and the accusation that *Uber* has stolen the design of *LIDAR* systems from Google. Comments containing positive sentiments were most often related to technical progress and the future impact of autonomous cars on safety issues. Several stakeholder groups discussed online whether
autonomous cars will become part of our daily life as early as the next decade.

5. The impact of **algorithms** is discussed very widely in the news, ranging from applications like *breast cancer* detection to improved user experience in software products. On social media, positive sentiments prevail, although topics like *fake news* or *bias of algorithm* raise criticism and negativity.

6. Trending topics around **Fake News** reflect a highly politicised debate in both online press and social media. The polarising nature of the topic is clearly reflected in the overwhelmingly negative sentiment score. The tendency towards harsh criticism and mistrust of the mainstream media is evident in the data.

7. **Screen Addiction** is discussed as a condition of the coming generation. News coverage mainly highlights the effects on the young generation as they adopt ever more immersive digital technologies. On social media, Screen Addiction is being compared to *drug and alcohol addictions*, indicating the extent of its harm as perceived by the public.

8. Media debate around **Bioterrorism** is shaped by public fear of terrorism from groups like *ISIL/Da'esh* and by discussion of potential counter-measures. Companies like *PharmAthene*, a provider of Anthrax vaccines, got much attention recently. *Bill Gates'* warning that tens of millions could be killed by Bioterrorism drove the discourse both in news as well as on social media.

The combination of horizon scanning and the analysis of technology-related trends, through a blend of machine- and mind-based data analytics methods, could support strategic decision-making and inspire thinking processes in planning and executing STOA activities.

---

**Note for the reader:**

The findings presented in this report are based upon data collected from January to April 2017.
1. INTRODUCTION

The STOA – Science and Technology Options Assessment (STOA) panel aims to ensure the European Parliament is well informed about techno-scientific advancements, and supporting the Members with evidence to anticipate possible future impact of these developments. This study was conducted to scan areas with a high relevance for uptake by STOA. First of all, the experts conducting this study were asked to look into technology-related trending topics about which there is much debate ongoing, topics with a high controversy in views from different stakeholders’ perspectives. In addition it was the aim to analyse some topics which already have been considered for investigation by STOA but not taken up so far.

This study as such consisted of two components: a first horizon scan of trendy techno-scientific topics combined with a controversy analysis between the stakeholders involved in these topics, and an analysis of eight trending topics, to get insight in the different perspectives, aspects and viewpoints from the diverse stakeholders involved in the technology, its applications and its consequences.

Horizon scanning gives an overall view of techno-scientific topics, to support the development of strategies for anticipating future advancements. The study is conducted on eight trending topics, of which five are based on the controversy ranking of horizon scan, and three other that are of potential particular interest to the STOA panel, being raised already as potential study areas, and/or with a high impact in case of manifestation.

These eight topics are analysed in depth, aiming to get a deeper insight (‘intelligence”) in the different facets, unraveling the complexity of the trending issue. This analysis investigates the stakeholders for which the topic is relevant, the related or embedded technologies, the target groups involved, the opportunities and challenges, the areas of impact, the sub-domains related to positive and negative sentiments and the geographical locations where these are key trends are an issue.

Hence, this horizon trend scan provides intelligence around social and technological phenomena in order to support the STOA panel in evaluating the trends’ potential societal impacts and provide options for future discussion and in-depth investigation. The combination of the horizon scanning and the analysis of technology related trends using a blend of machine- and mind-based data analytics methods is geared to handle big amounts of online media content and can enthuse thinking processes in programming STOA activities, as well as preparing the framework of a topic selected for a STOA activity when preparing execution.

First, the experts used a data analytics methodology, more precisely several elements of its artificial intelligence toolkit, combined with a ‘journalistic’ approach, to collect and analyse 16,491 news articles and over 8.3 million tweets. This means the study is based on a combination of mind- and machine-based data analytics, described in more detail in Chapter 2.

In the pre-scanning stage, 24 technological and social phenomena were analysed to gauge the controversy surrounding them in online discussions. For this analysis, a measure has been developed to track the activity and sentiment of online discussions with the aim of identifying the most controversially discussed socio-technological trends.

The following visualisation (Figure 4) shows the activity measured as total number of hits (y axis), sentiment divergence score based on stakeholder sentiment analysis (x axis) and the amount of unique users (bubble size):
As shown in Figures 4 and 5, topics like Artificial Intelligence rank highly when it comes to the amount of hits and unique users posting about the trend phenomenon, while topics like Waste Management are discussed most controversially.

In order to calculate a controversy ranking, as visualized in the following graph (Figure 5), both the controversy score based on sentiment divergence score and topic activity based on the amount of tweets have been used.
The final selection of trend phenomena results from a combination of a selection of topics resulting as relevant from the controversy identification (for instance these subjects lacking previous research by STOA) as well as topics with a potential high societal impact and/or having been recently suggested for uptake by STOA by MEPs.

Using a combination of machine based data analytics methodology (artificial intelligence tools) and mind-based journalistic methodology, (looking beyond raw data and to discover the stories behind it), big data sets related to the following phenomena have been analysed:

- Big Data,
- Gene Technology,
- Electric Vehicles,
- Autonomous Cars,
- Impact of Algorithms,
- Screen Addiction,
- Fake News and
- Bioterrorism.

Especially in environments characterised by exponential growth of online media content, data and knowledge, this combination of horizon scanning and in-depth analyses of technology-related trends via a mix of machine- and mind-based data analytics methods shows potential for inspiring future STOA activities.

The methodology and trends analysis are covered in this report. However, an alternative way to explore these is using Augmento’s webtool. For getting access to this web platform, Members can ask a username and password at stoa@europarl.europa.eu.
2. METHODOLOGY

The dataset used in this report consist of 16,491 news articles and over 8.3 million tweets, together with their associated metadata, collected from January to April 2017. In order to detect and map the topics across various sections of the data, the experts applied a data analytics methodology. The analysis modules performed to produce the report at hand employ stakeholder classification, sentiment classification, phrase extraction, and named entity recognition algorithms.

1.1. Target of analysis

During a topic pre-scanning phase, together with the STOA administrator the experts defined 24 technological and social trend topics, whose controversy should be evaluated. The topics were defined in an iterative process building upon both the experts’ ongoing qualitative scanning of news articles, as well as STOA’s thematic priority areas. The outcome, sorted by STOA priority areas:

Figure 6: Technological and social trend phenomena, sorted by STOA priority area

The experts developed a methodology for measuring controversy on social media and constituted an initial controversy ranking, published in the report “Stakeholder based controversy identification”.

In close collaboration with the STOA secretariat, the topics Big Data, Gene Technology, Electric Vehicles, Autonomous Cars, Impact of Algorithms, Screen Addiction, Fake News and Bioterrorism have been selected for an in-depth explorative study. Additional to performance on the controversy ranking, importance for the STOA panel and potential impact (as in the case of Bioterrorism) have been included as selection criteria.

1.2. Data

Building relevant search queries for each of the topics was done by applying keyword and hashtag tracking. The first step involved identifying the keyword/hashtag with highest relevance within each topic. Thereafter, dedicated tracking tools were applied to identify the set of keywords/hashtags that are most commonly associated with main keyword. The final keyword list for each topic is uniform across data sources, with the exception of Twitter, where hashtags corresponding to each keyword are included in the query.
1.3. Identification of trending topics

All news articles and tweets underlying this report’s findings were analysed by a trending topic algorithm.

Trending topics analysis is used to provide a representative picture of online text data. Regardless of the data source and the specific features underlying various text documents, it outputs most frequently occurring and relevant phrases, clustered according to their similarity.

In this report, data from news articles is partitioned into topics based on a list of keywords. Trending topics analysis for Twitter data is applied on specific subsets of tweets partitioned according to prevailing sentiments and the stakeholder categories. Thus, analysis of Twitter is performed on the following subsets:

- Most positive tweets
- Most negative tweets
- Across each stakeholder category (maximum of 10 stakeholder-subsets per topic, contingent on availability of data for various stakeholder categories).

Core tenets of sentiment and stakeholder classification used to generate the subsets are described below.

1.4. Stakeholder classification

Stakeholder classification is performed by a two-stage probabilistic classifier with a threshold. For each user, the classifier calculates the probability of belonging to a certain stakeholder category given the metadata (especially user description). At the second stage, the user is classified to the most probable category only if the associated probability exceeds the threshold; otherwise she is marked as Ambiguous. The threshold is set manually based on experimental results. The Ambiguous category is omitted from the stakeholder-specific trending topics analysis.

1.5. Sentiment analysis

Sentiment of each tweet is an ordered triple of real numbers, representing its negative, neutral and positive scores respectively. For each tweet, the classifier outputs the probabilistic sentiment distribution of a tweet (negative, neutral, positive). At the second stage, the tweet is classified as either positive or negative if the associated sentiment distribution exceeds the set thresholds; otherwise the tweet is marked as neutral and omitted from the sentiment-specific trending topics analysis. The threshold is set manually based on experimental results.

1.6. Phrase extraction

In the following step, each data subset is analysed for the most frequently occurring phrases, or n-grams, defined as a contiguous sequence of n words remaining after raw text processing (tokenization). We use commonplace tokenization steps such as stemming and stop word removal.

N-gram extraction method is based on a set of rules to identify those recurring sequences and can be divided in 3 parts: extraction of unigrams (a sequence of one word), bigrams (sequence of two words), and phrases (sequence of n-words). For both unigrams and bigrams the method breaks each sentence into a list of words and filters unigrams or bigrams that contains “nouns” or “adjectives”. The phrases extraction defines a syntax map and identifies sentences that follow a well-defined structure (for example a noun phrase followed by a
prepositional phrase).

1.7. Phrase selection and clustering

As a next step, phrases are first selected based on frequency and relevance, and are then clustered according to topic similarity. A mixed approach has been used to arrive at the final clusters. Named entities such as persons, locations, and companies were automatically classified using a named entity recognition algorithm. The remaining key phrase selections and resulting clusters are obtained using a journalistic approach.

1.8. Visualisation

The output is visualised in interactive graphs which allow for an intuitive interpretation of topic size and its underlying phrase components. Graphs report most common phrases and the amount of mentions accompanied by a representative sample tweet/text snippet associated with the phrase. The report contains two types of visualisations: bubble graphs and horizontal bar charts. Bubble graphs are used to illustrate trending topics consisting of multiple clusters, whereas bar charts show trending topics within non-clustered subsets.

1.9. Provision of context (tweets, text snippets)

Each phrase reported in the trending topics analysis is accompanied by sample text to provide context to the reader. Samples of Twitter data contain whole tweets, whereas phrases from articles use the title or a representative sentence drawn from the article body.

1.10. Controversy ranking

In addition to the methods used in the analysis above, the methodology of controversy ranking is presented. The ranking was among the key components in identifying the final list of techno-scientific topics analysed in this report.

We proposed a controversy score based on three factors: intra-topic controversy, inter-stakeholder controversy and topic activity.

**Intra-Topic Controversy (ITC)** of a topic with sentiment \((pos, neu, neg)\) is calculated using a formula.

ITC is high when both neutral score and the difference between positive and negative scores are low. Topic sentiment is the mean of sentiment scores of stakeholder groups it consists in.

**Inter-Stakeholder Controversy (ISC)** of a topic is the variance of sentiments of stakeholder groups it consists in. Again, stakeholder group’s sentiment is the mean of sentiments of tweets published by users classified to the given group. ISC is high when the stakeholder groups’ sentiments differ from each other a lot, and it is low when their sentiment about the topic is similar. For each trend phenomenon, the ISC was visualised using Stakeholder Sentiment Distance scatterplot, showing the sentiment distributions of stakeholder groups.
3. PRESENTATION OF THE FINDINGS

The findings of the analysis are presented in eight sections, covering one trend phenomenon each and following the same basic structure. Five of the section contain following results: general trending topics in news; and sentiment-specific trending topics, stakeholder-specific trending topics, and stakeholder activity derived from twitter data. The sections of Screen Addiction and Bioterrorism contain general trending topics in news and twitter data only due to a limited amount of data. The section of Fake News contains trending topics of most negative tweets only as positive tweet on that subject are negligible.

The bubble graphs show trending topics grouped into various categories. A trending topic is a phrase extracted from any given dataset and selected based on its relevance and activity, measured as the number of mentions. In the graph showing sentiment-specific trending topics, green bubbles contain topics from positive tweets, whereas red bubbles contain topics from negative tweets. The size of bubbles indicates the activity.

The bar charts visualise stakeholder-specific trending topics and stakeholder activity.

The methodology and trends analysis are covered in this report. However, an alternative way to explore these is using Augmento’s webtool. For getting access to this web platform, Members can ask a username and password at stoa@europarl.europa.eu.

Notice for the reader: The findings presented in this report are based upon data collected from January to April 2017.

---

2 With the exception of “Screen Addiction” and “Bioterrorism”, where the results include trending topics in news and trending topics in tweets, due to a limited amount of data.

3 See also the chapter Methodology.
3.1. BIG DATA

3.1.1. Trending topics in news on Big Data

Figure 12: Trending topics in news on Big Data

An analysis of 1140 articles related to Big Data illustrates the state of the art discourse around the trend. Big Data is discussed considering both commercial and public application, however commercial aspects appear to receive more media attention. Common topics for commercial application revolve around how Big Data can fuel efficiency (e.g. in supply chain management) and business intelligence. It is remarkable that Big Data seems to find its way also into the sphere of small businesses, being no longer a privilege of large corporations. However, also aspects which are newer to the business vocabulary have been recently linked to Big Data, such as data driven customer experience or how Big Data can transform whole business models.

Topics around public application include the interface of Big Data and Smart Cities, public space as well as Open Data.

Notwithstanding the above, Big Data is discussed as a facilitator for better decision-making and as a source for predictions and actionable insights.
Within **Big Data**, Health is an emerging topic in the recent discourse, and its development into a high impact area remains to be seen. The media discourse ranges from **Big Data** supported *image recognition of cancer cells* to more systemic topics such as the role of **Big Data** in *public health*.

*Europe* is a trending region in connection to **Big Data**, however, given the dataset at hand, it is less prominent in current discourse compared to *China, India, and the US*.

Internet giants like *Google* and *Facebook* are still getting the most attention in the field of **Big Data**, but are followed closely by *Apache Software Foundation*, an US non-profit corporation focusing on open source. Most notably *Cambridge Analytica* — known for its involvement in Donald Trump’s 2016 presidential campaign — has been a prominently discussed company in the field of **Big Data**. A weak but interesting signal is the mentioning of Californian electric vehicle start-up *Faraday Future* heralding the arrival of **Big Data** in the Automotive Industry.

Commonly mentioned technological aspects of **Big Data** are its linkage to *artificial intelligence* and the *internet of things*. The prominence of *open source* comes not without reason: Usually, when it comes to large business opportunities, the resources backing them would be expected to come with a heavy price tag. As an exception, **Big Data** has its roots and future in open source technologies driven by strong global tech-communities.

**Figure 13: Companies (Big Data)**
Figure 14: Issues (Big Data)

Figure 15: From Data to Action (Big Data)
Figure 16: Technology (Big Data)

Figure 17: Commercial Application (Big Data)
Figure 18: Places and Regions (Big Data)

Figure 19: Health (Big Data)
Figure 20: Public applications (Big Data)
3.1.2. Trending topics of most positive and negative tweets on Big Data

On Twitter, Big Data is discussed in a predominantly positive sentiment. The analysis estimates a general distribution of 82% positive and 18% negative sentiments (disregarding neutral comments).

Robert Mercer, an American computer scientist and co-CEO of Renaissance Technologies, was highly prominent in negative statements around Big Data. This is due to his mention in the Guardian article “Robert Mercer: the big data billionaire waging war on mainstream media” and his involvement in Cambridge Analytica.

The major part of the discussion on Social Media appears in positive sentiments towards Big Data. Trending topics of positive sentiments show a multifaceted range of discussed application areas such as Artificial Intelligence, the travel industry or homeland security.
Figure 22: Negative tweets on Big Data
Figure 23: Positive tweets on Big Data
3.1.3. Stakeholder specific trending topics – Big Data

This section presents most trending topics used by different stakeholder groups on Twitter when discussing Big Data.

This analysis is based on a dataset consisting of 2,575,713 tweets. Timeframe: 12/1/2017-24/4/2017.

Figure 24: Stakeholder specific trending topics – Big Data
### 3.1.4. Big Data stakeholder activity

**Figure 25: Big Data stakeholder activity**

3.2. GENE TECHNOLOGY

3.2.1. Trending topics in news on Gene Technology

Figure 26: Trending topics in news on Gene Technology

An analysis of 1239 news articles related to Gene Technology illustrates the state of the art discourse around the technology.

Impact areas of Gene Technology are diverse. The role of Gene Technology for genetic modification of humans is reflected in mentions of human embryos or designer babies.

Human health is a common theme, for example in the context of new cancer treatments or HIV vaccines.

Related technologies are most notably a genome editing technology known as CRISPR-Cas9 that allows permanent modification of genes within organisms as well as artificial intelligence.
Figure 27: Locations (Gene Technology)

Figure 28: Impact Areas (Gene Technology)
Figure 29: Issues (Gene Technology)

Figure 30: Technology (Gene Technology)
3.2.2. Trending topics of most positive and negative tweets on Gene Technology

This analysis estimates a general distribution of 55% positive and 45% negative sentiments (disregarding neutral comments) across all identified stakeholders.

Negative sentiments are generally related to farming and impact of GMO in the food chain combined with effects on human health. This is also reflected in company mentions like Monsanto as well as food and beverage companies like Coca Cola, accused for blocking GMO labeling laws. Minor, potentially emerging topics catering for negative sentiments are corporate legal fraud when dealing with gene technology and biological terror.

A big driver of positive tweets are food related lifestyle tweets—tweets of private users documenting their GMO free lifestyle. Hence, “GMO free” is often mentioned along with “soy free” and “fat free”. Also, scientific advances are discussed in mostly positive sentiments.
Figure 32: Negative tweets on Gene Technology

Figure 33: Positive tweets on Gene Technology
3.2.3. Mentions of CRISPR/Cas9 and CRISPR/Cas on social media

As an additional analysis, a guided keyword search was applied to track the activity around the terms **CRISPR/Cas9** and **CRISPR/Cas**, referring to the targeted genome editing technology.

Figure 34: Mentions of CRISPR/Cas9 and CRISPR/Cas on Social Media

3.2.4. Stakeholder specific trending topics – Gene Technology

This section presents most trending topics used by different stakeholder groups on Twitter when discussing Gene Technology.

This analysis is based on a dataset consisting of 590,802 tweets. Timeframe: 12/1/2017-24/4/2017.

Figure 35: Stakeholder specific trending topics — Gene Technology
STOA – Science and Technology Options Assessment

Media

Private persons

![Bar chart showing media and private persons' interest in topics related to gene technology.](chart)

**Individual experts**

![Bar chart showing individual experts' interest in topics related to gene technology.](chart)

3.2.5. Gene Technology stakeholder activity

Figure 36: Gene Technology stakeholder activity

3.3. ELECTRIC VEHICLES

3.3.1. Trending topics in news on Electric Vehicles

Figure 37: Trending topics in news on Electric Vehicles

An analysis of 2676 news articles illustrates the state of the art discourse around the technology.

News on Electric Vehicles is very much dominated by car brands and their announcements. We find established automotive matadors such as Nissan, Ford and BMW, who, along with production lines of combustion engines, now also target the EV market. However, the undefeated champion in the race for attention in the field of Electric Vehicles is Tesla.

Price, speed and range seem to be the most important choice factors for going electric but also tax benefits play an important role deliberating to purchase an EV.

Further, topics related to climate friendly transportation and green energy are on the news agenda. Electric Vehicles are mostly seen as an opportunity to reduce emissions and improve air quality in cities. However, it has also been mentioned that if EVs are charged with fossil fuel generated energy the effect would be negative for environment compared to conventional transportation.

Development of battery technology is a major technology related topic. Batteries are still one of the most problematic bottlenecks for the breakthrough of EV.
Figure 38: Companies (Electric Vehicles)

Figure 39: Locations (Electric Vehicles)
Figure 40: Climate and Ecology (Electric Vehicles)

![Climate and Ecology Diagram]

Figure 41: Choice Criteria (Electric Vehicles)

![Choice Criteria Diagram]
Figure 42: Tax (Electric Vehicles)

Figure 43: Car Models (Electric Vehicles)
Figure 44: Batteries (Electric Vehicles)
3.3.2. Trending topics of most positive and negative tweets on Electric Vehicles

The analysis estimates a general distribution of 74% positive and 26% negative sentiments (disregarding neutral comments) across all identified stakeholders.

Negative sentiments revolve around topics such as Bentley’s small electric SUV, low gas prices and Opec’s price war.

Dominating positive sentiments concern green energy, incentive programs like the one in Norway, safety and also Elon Musk’s “energy battle against Trump’s favorite fuels”. 
Figure 46: Negative tweets on Electric Vehicles

Figure 47: Positive tweets on Electric Vehicles
3.3.3. **Stakeholder specific trending topics—Electric Vehicles**

This section presents most trending topics used by different stakeholder groups on Twitter when discussing **Electric Vehicles**.

This analysis is based on a dataset consisting of 386,975 tweets. Timeframe: 12/1/2017-24/4/2017

**Figure 48: Stakeholder specific trending topics—Electric Vehicles**
3.3.4. Electric Vehicles stakeholder activity

Figure 49: Electric Vehicles stakeholder activity

3.4. AUTOMONOUS CARS

3.4.1. Trending topics in news on Autonomous Cars

Figure 50: Trending topics in news on Autonomous Cars

An analysis of 2283 articles related to Autonomous Cars illustrates the state of the art discourse around the technology.

It is highly remarkable that traditional car manufacturers like Toyota or Mercedes Benz make way for new players in the field of autonomous driving: Uber, Google and Tesla. Weak signals of interest are Amazon and the Chinese internet giant Baidu, which entered the global race for autonomous car domination. A central theme are technological advances such as speculations about when full autonomy can be reached and impact areas of autonomous driving such as security, traffic jams or changing business models.
Figure 51: Companies (Autonomous Cars)

Figure 52: Locations (Autonomous Cars)
Figure 53: Impact Areas (Autonomous Cars)

Figure 54: Technology (Autonomous Cars)
3.4.2. Trending topics of most positive and negative tweets on Autonomous Cars

The analysis estimates a general distribution of 44% positive and 56% negative sentiments (disregarding neutral comments) across all identified stakeholders.

Negative sentiments have recently been caused by the fear of autonomous car crashes and the accusation that Uber stole the design of laser-based LiDAR vision sensors from Google. Positive sentiments revolve around Goodyear’s AI tire concept as well as advances in security and adjacent areas such as artificial intelligence.
Figure 56: Positive tweets on Autonomous Cars

Figure 57: Negative tweets on Autonomous Cars
3.4.3. Stakeholder specific trending topics— Autonomous Cars

This section presents most trending topics used by different stakeholder groups on Twitter when discussing Autonomous Cars. This analysis is based on a dataset consisting of 816,751 tweets. Timeframe: 12/1/2017-24/4/2017.

**Figure 58: Stakeholder specific trending topics— Autonomous Cars**

**Media**
- Bosch team
- Autonomous car tests
- Cyber security
- Internal Uber documents
- Autonomous vehicle tech firm Argo AI
- Concept car

**Business representatives**
- Next decade
- Global insurance giant
- International Uber documents
- Autonomous
- Autonomous vehicle testing

**Companies**
- Cen2017
- Smart home
- Reliable mode of transportation
- Internal Uber documents
- Next decade

**Individual experts**
- Smart mobility
- Smart cities
- Data-driven smart mobility tech
- Next decade
- Deep learning
Private persons

- Arizona crash
- Semi autonomous
- Next decade
- Driving decisions
- Gerard Kruisheer
- Concept car
- Free ride in Pittsburgh

3.4.4. Autonomous Cars stakeholder activity

Figure 59: Autonomous Cars stakeholder activity

3.5. IMPACT OF ALGORITHMS

3.5.1. Trending topics in news on Algorithms

Figure 60: Trending topics in news on Algorithms

Trending topics in news on Algorithms draw a broad picture of related technologies, impact areas and issues. Closely related technologies are for example artificial intelligence, social media and search engines, where Algorithms fulfil a variety of context-specific purposes.

Recently mentioned impact areas range from Algorithms helping to detect breast and skin cancer to personalised user experiences. Algorithms are expected to outperform clinicians in predicting suicide risks, an endeavour Facebook also recently embarked on by identifying user posts indicating potential suicide risks.

Algorithms bring along several issues. Commonly discussed issues are that algorithms often behave like black boxes—which means that users cannot control outputs of Algorithms—algorithmic error rates, and social media “filter bubbles”. Recently, news coverage also dealt with reactions of SEO specialists who need to redesign mobile interstitials—pop-up banners on mobile phones—in order to avoid ranking punishment by one of Google’s algorithms.
Figure 61: Locations (Algorithms)

Figure 62: Companies (Algorithms)
Figure 63: People (Algorithms)

Figure 64: Technology (Algorithms)
Figure 65: Impact Areas (Algorithms)

Figure 66: Issues (Algorithms)
3.5.2. Trending topics of most positive and negative tweets on Algorithms

Figure 67: Trending topics of most positive and negative tweets on Algorithms

The analysis estimates a general distribution of 61% positive and 39% negative sentiments (disregarding neutral comments) across all identified stakeholders. Negative sentiments have recently been caused by discussions like algorithm errors, fake news and bias in algorithms. Positive sentiments revolve around applications of Algorithms like predicting terror attacks or support in drug development.
Figure 68: Positive tweets on Algorithms

Figure 69: Negative tweets on Algorithms
3.5.3. Stakeholder specific trending topics—Impact of Algorithms

This section presents most trending topics used by different stakeholder groups on Twitter when discussing Algorithms.

This analysis is based on a dataset consisting of 369,427 tweets. Timeframe: 3/3/2017-24/4/2017.

Figure 70: Stakeholder specific trending topics—Impact of Algorithms

Academia

- Big data
- Artificial intelligence
- Skin cancer
- Google DeepMind
- Racist algorithm

Business representatives

- Artificial intelligence
- Skin cancer
- Big data algorithm
- Search algorithm
- Facebook algorithm
- Racist algorithm
- Human intuition

Companies

- Algorithms update
- Google algorithm
- Artificial intelligence
- Big data
- Search algorithm
- Feed algorithm

Media

- Artificial intelligence
- Google algorithm
- Skin cancer
- Racist algorithm
3.5.4. Algorithms stakeholder activity

Figure 71: Algorithms stakeholder activity

3.6. SCREEN ADDICTION

3.6.1. Trending topics in news on Screen Addiction

Figure 72: Trending topics in news on Screen Addiction

Screen Addiction is closely related to the technologies it is caused by, especially mobile phones and intangible technologies such as social media networks and the rising technology of virtual reality.

The main demographic group mentioned in the context of Screen Addiction are children. A central question is how education in schools and screen time will go hand in hand in future.

Main impact area of Screen Addiction is health, whereas mental health is a main concern. Also smaller topics are covered such as phenomenons like phantom phone vibrations or the macro effect of Screen Addiction on public health.
Figure 73: Locations (Screen Addiction)

Figure 74: Technology (Screen Addiction)
Figure 75: Impact Areas (Screen Addiction)

Figure 76: Children and Youth (Screen Addiction)
Figure 77: Addictions (Screen Addiction)
3.6.2. Trending topics in tweets on Screen Addiction

Similar to the media discourse, Twitter conversations were shaped by health and mental health impact issues of Screen Addiction. Screen Addiction gets compared to other addictions such as alcohol or drug addiction to illustrate its harmfulness. Various ways to deal with Screen Addiction and cures such as “strategic unplugging” and “shock therapy” are being discussed.
Figure 79: Countries (Screen Addiction)

Figure 80: People (Screen Addiction)
Figure 81: Impact Areas (Screen Addiction)

Figure 82: Addictions (Screen Addiction)
3.7.  FAKE NEWS

3.7.1.  Trending topics in news on Fake News

Figure 83: Trending topics in news on Fake News

Trending topics around Fake News are deeply embedded into geopolitical debates reflecting tensions between the USA and Russia. The term Fake News is broadly used for vilifying political opponents. News about the inauguration crowd photo or the French presidential election were fueling the debate.
Figure 84: Locations (Fake News)

Figure 85: People (Fake News)
Figure 86: Law and Security (Fake News)

Figure 87: Policy Areas (Fake News)
Figure 88: Events (Fake News)
3.7.2. Trending topics of most negative tweets on Fake News

Figure 89: Trending topics of most negative tweets on Fake News

The analysis estimates a general distribution of 97.5% negative and 2.5% positive sentiments (disregarding neutral comments) across all identified stakeholders. Negative tweets mentioning Fake News illustrate the public divide concerning the legitimacy of mainstream media. The conversations on social media reflect growing scepticism and mistrust towards established news outlets such as CNN, BBC, NBC, Fox, or The Washington Post. Accusations of false facts pertain to a wide array of political debates - U.S. politics, international politics, refugee crisis, race- and identity politics.
Figure 90: People (Fake News)

Figure 91: News Stories (Fake News)
Figure 92: Left-Right Divide (Fake News)

Figure 93: Media (Fake News)
Figure 94: International Politics (Fake News)

Figure 95: Nationalism and minorities (Fake News)
3.7.3. Stakeholder specific trending topics—Fake News

This section presents most trending topics used by different stakeholder groups on Twitter when discussing Fake News.


**Figure 96: Stakeholder specific trending topics—Fake News**

- **Academia**
  - Denial of climate change
  - Donald trump
  - Angela merkel
  - Election poll
  - Evidence photos colluded with russia
  - #medialiteracy

- **Individual experts**
  - Donald trump
  - Denial of climate change
  - Angela merkel
  - James clooper

- **Media**
  - Donald trump
  - Angela merkel
  - Fbi director hearing
  - Denial of climate change
  - #vault7
  - Swedish integration minister
  - Fake news pandemic

- **Non corporate interests**
  - Donald trump
  - Fox news
  - Denial of climate change
  - Angela merkel
  - Bernie sanders
  - #resist
  - Fbi director
  - #vault7
  - #russiagate
Private persons

![Bar chart showing activity levels of different private persons.](image)

**3.7.4. Fake News stakeholder activity**

**Figure 97:** Fake News stakeholder activity


![Graph showing stakeholder activity.](image)
3.8. BIOTERRORISM

3.8.1. Trending topics in news about Bioterrorism

Figure 98: Trending topics in news about Bioterrorism


Media attention towards Bioterrorism has recently been fuelled by the increased threat of proliferation due to a combination of rise of ISIS, and technological advances that allow for low-cost acquisition of such weapons by terror groups. Stories like Bill Gates warning that biological weapons dwarf nuclear in their potential to cause damage provide additional momentum to the popular debate. The primary impact areas discussed in relation to Bioterrorism revolve around countermeasures highlighting aspects such as policy, funding, and vaccine development.
Figure 99: Companies (Bioterrorism)

Figure 100: People (Bioterrorism)
Figure 101: Impact Areas (Bioterrorism)

Figure 102: Terrorism (Bioterrorism)
Figure 103: Bio Warfare (Bioterrorism)

Figure 104: Pathogens and Therapeutics (Bioterrorism)
3.8.2. Trending topics in tweets on Bioterrorism

Figure 105: Trending topics in tweets on Bioterrorism


Bioterrorism is currently a niche topic on social media. Main activity around it is generated by discussions on various pathogens that could be used as potential Bioterrorism agents. Appearing often in the context of Bioterrorism is CRISPR. Conversations around CRISPR mainly warrant a new biodefense policy that properly addresses its threat as a new powerful tool to develop novel bioweapons.
4. CONCLUSIONS AND REFLECTIONS

This study covers an explorative horizon scan of technology-related topics which could change our lives, together with a controversy analysis and an in-depth analysis of a selection of eight trending topics. It shows the potential of data analytics, combining ‘machine-based methods’ using big data and artificial intelligence tools and mind-based methods, using journalists methods, for gaining insight for STOA in their strategic planning for the selection of the most strategic technology-related topics which could change our lives, and for setting the objectives and specifications of flagship studies, ensuring these encompass all facets relevant for the European society with relevance for EU policy.

Especially sentiment analysis and stakeholder analysis proved to be promising tools for observing opinion shifts not only by citizens but also by academia, industry experts, politicians and non-corporate interest groups like NGOs. Further, the combination of methodologies used allowed answering the WHY behind the data through journalistic research as an important part of the analysis.

This new way of creating knowledge by combining machine and human intelligence to analyse big amounts of text data can have significant impact in better understanding the interface between technology and society and therefore reveal new options for future projects and data driven discussions. This holds true for hitherto not assessed trend phenomena as well as for trends already explored by STOA, such as Artificial Intelligence which was partly covered by the study on the Ethics of cyber-physical systems, where an entirely new perspective could complement existing findings.

Reflections

The data explored for this study were taken from a limited time-period. This means that the outcomes give a picture of the situation at the time the data collection had been closed. It might be interesting to repeat such an analysis at regular time intervals, for instance a general horizon scan over technology-related discussions in social and mainstream media every year, complemented with a controversy analysis. An analysis using the combination of machine- and mind-based data analytics could be useful at the start of strategic STOA activities to ensure the full scope of a topic will be taken into account in a study.

The experts who conducted the study recommend to track sentiments such as fears and hopes of citizens and other stakeholders towards new technologies for an extended time period. Additionally, other tools, such as methods to identify and measure time horizon assessments of emerging technologies in online communication—could lead to insights driven by a future oriented perspective.

Overall conclusion

Summarizing, having also discussed the outcomes of this study during the STOA Panel meeting of 18 May 2017 in Strasbour, we could state that the combination of the horizon scanning and the analysis of technology related trends, via a blend of machine- and mind-based data analytics methods, could support strategic decision making and inspire thinking processes in planning and executing STOA activities.
This horizon scan has identified eight major technological trends relevant for STOA. First, a scan was conducted to measure controversy on social media, and this constituted an initial controversy ranking. After more detailed analysis of the main technology trends identified, a set of STOA-relevant areas were selected, which have not yet been investigated by STOA so far. These are big data, gene technology, electric vehicles, autonomous cars and impact of algorithms. A number of additional trend areas with high potential impact on society were identified for analysis: screen addiction, fake news and bioterrorism.

Within the eight topics selected for detailed analysis from the initial horizon scanning process, keywords, subtopics, and sentiments have been detected and analysed from social media and news articles. These eight technologies are areas for discussion amongst the STOA Panel members when considering new project activities to be undertaken.