

Eight goals for a human-centric internet

D1.9: NGI Topic guides and evaluation report I

Work package	WP 1: Topic Identification
Task	1.7 NGI Topics Filtering & deep dives
Due date	31/03/2020
Submission date	31/03/2020
Deliverable lead	DATALAB, Aarhus University
Dissemination level	Public
Nature	Report
Authors	Lynge Asbjørn Møller, DATALAB, AU Anja Bechmann, DATALAB, AU Kristóf Gyódi, DELab, UW Michał Paliński, DELab, UW Łukasz Nawaro, DELab, UW
Version	1.0
Reviewers	Heshani Jayaratne, Nesta Katja Bego, Nesta Louis Stupple-Harris, Nesta
Status	Final

Disclaimer: The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained herein.

Acknowledgement: This Report is part of a project that has received funding from the **European Union's Horizon 2020 research and innovation programme under grant agreement N°825652**

Table of Content

1 Introduction	2
1.1 Purpose and scope	2
2 Methodology behind topic selection	4
2.1 Identification of emerging technologies and issues	4
2.1.1 Keyword identification	5
2.1.2 Selection of initial umbrella topics	6
2.2 Expert workshop	7
2.3 Topic selection	7
2.4 Further analysis	9
3 Topic guides	10
3.1 Trustworthy Information Flows	10
3.2 Decentralised Power on the Internet	11
3.3 Personal Data Control	12
3.4 Sustainable and Climate-friendly Internet	13
3.5 Safer Online Environments	14
3.6 An Inclusive Internet	15
3.7 Competitive European Ecosystems	16
3.8 Ethical Internet Technology	17
4 References	19

1 Introduction

In recent decades, there has been a revolution in the development of internet technologies across a wide range of fields, and all indications are that the technological progress is continuing at a rapid pace. These breakthroughs undoubtedly have a profound impact on society, and while they present significant opportunities, there are also complex dilemmas and challenges emerging around these new technologies.

Currently, the development of the internet technologies of the future is centralised around a few internet giants in monopoly-like positions on the global data market and, without an adequate regulatory response, humans risk losing control to data-driven, non-human-centric business models.

Progressive development of internet technologies and policy is needed in order to support a more human-centric evolution of the Internet, and in this process insights into emerging technologies and their corresponding challenges and opportunities can be of great value for European policy-makers. Understanding these emerging challenge areas will allow policy-makers to become involved in shaping their development early on to embed the more human-centric values central to the European Commission's *Next Generation Internet* initiative.

Employing a mixed method approach that includes computational social science methods and expert workshops, we identify a set of key topics that will be central in developing a more democratic, inclusive and resilient *Next Generation Internet*.

1.1 Purpose and scope

The purpose of this report is to present the first eight key NGI topics identified as part of the NGI Forward project. The topics were identified through an evaluation process involving a combination of quantitative and qualitative approaches. This report details the evaluation criteria and selection process and presents the list of topics along with individual topic guides.

Building on the findings from the expert workshop (see details in section 2.2), we have chosen to focus primarily on societal and human issues and values related to internet development rather than on technology as an end in itself. The rapid technological development in the recent decades demonstrates that focusing on specific tools and technology may render topics obsolete within just a few years, while societal challenges are more likely to remain relevant and allow the EU to focus on a wider range of solutions beyond a predetermined technology (see further rationale behind topic selection in section 2.3).

The following eight topics were selected as part of the first wave of key NGI topics:

- [Trustworthy Information Flows](#)

- [Decentralised Power on the Internet](#)
- [Personal Data Control](#)
- [Sustainable and Climate-friendly Internet](#)
- [Safer Online Environments](#)
- [An Inclusive Internet](#)
- [Competitive European Ecosystems](#)
- [Ethical Internet Technology](#)

These topics will play a central role in the NGI, forming a clear focus through the initiative, informing NGI's policy and technology research agenda, fuelling discussion on the NGI Forward Exchange platform, and shaping the thematic focus of events and other stakeholder engagement activities.

2 Methodology behind topic selection

To identify the key NGI topics, we employed a mixed method approach to validate and triangulate the veracity of the qualitative and quantitative data (Greene et al., 1989) - a methodological framework already tested in the *NGI Engineroom* project (see Bego, 2018).

In the first phase, we collected qualitative data from technology news articles and academic working papers to identify trending keywords related to the Internet in the broader public and research community respectively. The identification of trending keywords is based on growth in interest and debate over time, the centrality of the keywords within clusters of other keywords, and the keyword's connection to key emerging technology areas.

In the second phase, we organised an expert workshop with leading stakeholders in the internet research community to help narrow down our areas of focus and verify or adjust our results.

In the third phase, we synthesised the results to select eight key NGI topics that we then further illustrated in the news and research article-based dataset through deep dives using co-occurrence analysis and *topic modelling*.

All three phases are detailed in the following sections. Detailed descriptions on methods applied in the computational and workshop approaches respectively can be found in the NGI Forward deliverables *D1.2: Visualisations of key emerging technologies and social issues* (Gyódi et al., 2019) and *D1.13: Value-driven future internet: A social science perspective I* (Møller & Bechmann, 2019).

2.1 Identification of emerging technologies and issues

By taking a data-driven approach to kick off the topic selection process, we are able to cast a wider net and take into consideration a large corpus of articles and working papers. The data sources include technology news articles and academic working papers, as they complement each other well by representing different layers of information. While news articles cover the most current affairs and can be used to identify general trends, academic research papers provide more detailed insight into narrow fields and can help pinpoint the most important aspects and areas of the trends.

Working papers were chosen over published academic articles, as they have several important advantages: compared to the long publishing times of peer-reviewed journals, working papers are posted early in the research process, and they thus present the newest results and often propose more radical ideas that would likely be rejected during the peer review. Two repositories were included: ArXiv (STEM sciences) and SSRN (social sciences).

More than 213.000 news articles from 14 major English-language technology websites from the US, EU and Australia, and more than 139.000 working papers from 2 working paper

repositories covering both social and STEM sciences were collected (see Gyódi, Nawaro, Paliński & Wilamowski, 2019).

2.1.1 Keyword identification

The first step of the text-mining analysis is the identification of the most trending keywords based on observing changes in frequencies of the keywords over time. This method highlights the most trending technology-related keywords without any prior assumptions or short-lists, facilitating the discovery of unexpected but highly relevant areas.

Initially, we identified trending keywords for the different datasets separately to allow us to gain insight about which keywords are particularly popular among academics and the news media, and how the keywords evolve differently depending on the source. Unsurprisingly, news media seem to provide earlier signals than academic research, when it comes to social issues, while the academic papers provide more detail about the technologies in question.

We reviewed the top 1000 trending keywords across sources and chose the most relevant keywords for further analysis (172 unigrams and bigrams) that we then organised into umbrella topics. The highest trending NGI-related keywords are presented and organised in umbrella topics in Figure 1 and 2 below.

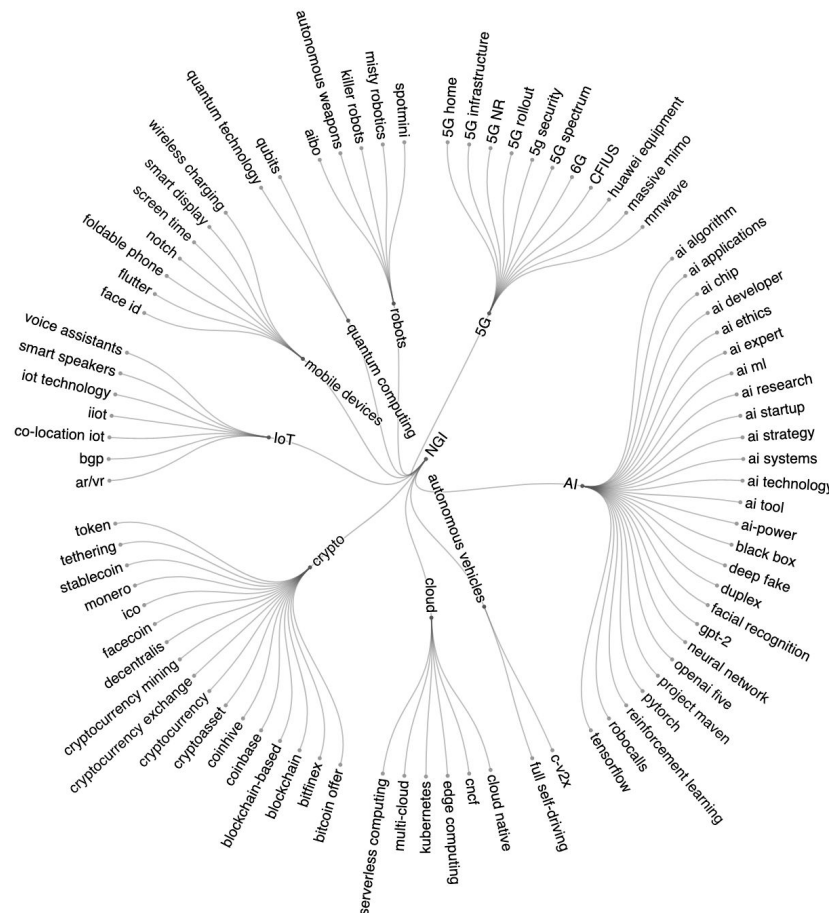


Figure 1. Umbrella topics and identified keywords: Emerging technologies

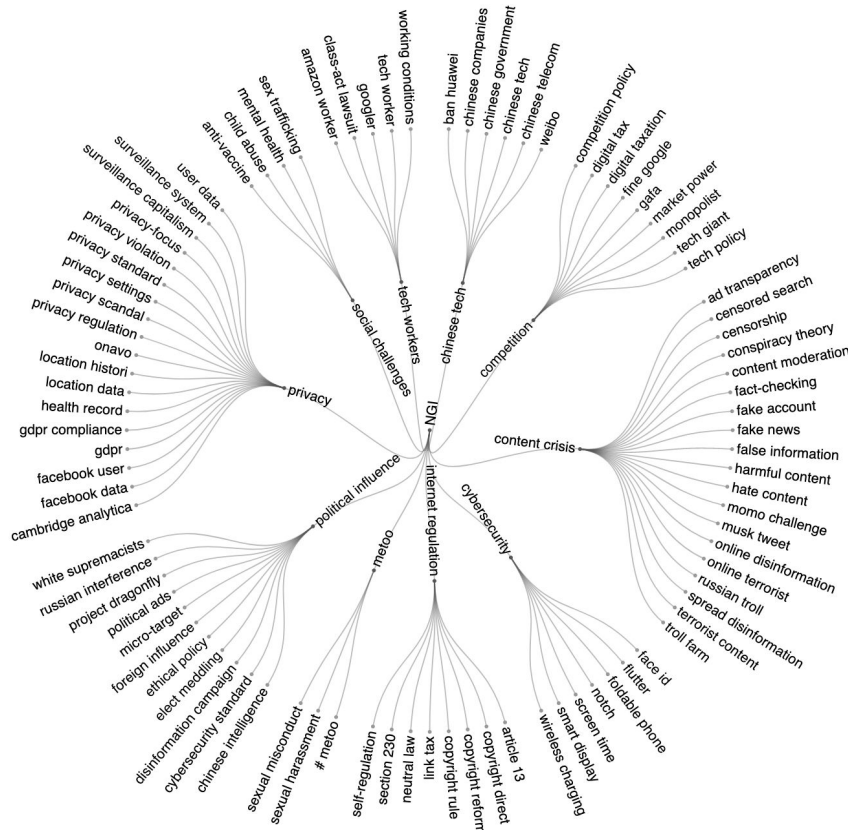


Figure 2. Umbrella topics and identified keywords: Relevant social challenges

2.1.2 Selection of initial umbrella topics

Based on the identification of trending keywords and umbrella topics, we then chose eight of the umbrella topics for further analysis in the expert workshops. This was largely a manual process, where we grouped keywords based on our expertise and insights from the cluster analysis and chose the most important umbrella topics where the social relevance is high, EU-level regulations are only partially in place, and ICT technology plays a crucial role.

The eight initial umbrella topics were:

- Artificial intelligence and machine learning
- Internet of Things
- Blockchain and cryptocurrencies
- Quantum computing
- Internet regulation
- Social media and content crisis
- Market competition
- Chinese tech sector

Further details on the methodology behind the identification of trending keywords can be found in the NGI Forward deliverable *D1.2: Visualisations of key emerging technologies and social issues* (Gyódi et al., 2019).

2.2 Expert workshop

With the aim of narrowing down our areas of focus, verifying or adapting the initial results and gathering expert insights, we organised the first of a series of three expert workshops with leading Internet research stakeholders at the annual conference for the *Association of Internet Researchers (AoIR)* on October 2nd, 2019. AoIR is one of the largest organisations in the world devoted exclusively to Internet research, and it prides itself in taking a leading role in advocating ethical and socially responsible approaches to Internet research (*About – AoIR*, n.d.). The participants in the workshop consisted mainly of social science researchers with computer scientists and engineers also present.

The workshop was conceptualised in the *future workshop* format (Jungk & Müllert, 1987) that is a useful method for groups of people to develop new ideas or solutions to complex problems that require the involvement of different stakeholders, particularly as a method to carry out participatory research aimed at bringing about social change (Alminde & Warming, 2019).

The workshop was split into three phases: a **critique phase**, where participants were asked to brainstorm on internet-related societal issues, technologies and values; a **fantasy phase**, where participants were asked to come up with the best possible ways to solve an issue identified earlier; and a **implementation phase**, where participants were asked them to reprioritise, rename or replace the list of umbrella topics identified in D1.2 to better reflect the ideas, values and solutions discussed in the two previous phases.

The qualitative approach to and main outcomes of the workshop are detailed in the NGI Forward deliverable *D1.13: Value-driven future internet: A social science perspective I* (Møller & Bechmann, 2019).

2.3 Topic selection

As a final step in the topic selection process, we synthesised the data-driven insights together with the workshop results to select the eight key NGI topics that differ from the initial eight umbrella topics based on critique and insights from the workshop. The rationale behind the selection of each key NGI topic is found below, while the actual topic guides are found in section 3.0.

Generally, the discussions that took place during the workshop were mainly focused on issues and values related to technological development rather than the technologies themselves. Participants argued for a larger focus on technology-agnostic solutions to human problems, where technology is not an end in itself, but a means to an end available

to make a positive societal impact. In the critique of the umbrella topics identified in D1.2, the technologies were deemed meaningless without societal or sector-specific context.

Consequently, we decided to define a set of broad key NGI topics focused towards different important goals in achieving a more human-centric internet. The chosen key NGI topics are technology-agnostic to prevent them falling out of relevance in the coming years and broadly interpretable and solution-agnostic to prevent predetermining specific solutions. Focusing on specific technologies or tools in the key topics would be risky, as a tool or technology that might be emerging in importance today may completely fall out of relevance in the near future due to new developments. An example that comes to mind is cryptocurrencies that were initially very hyped but quickly fell out relevance.

This was also discussed at the workshop, where the umbrella topic *Blockchain and cryptocurrencies* identified in D1.2 was criticised for being irrelevant in the broader picture in terms of actual adoption. Instead, focus was put on the potential of decentralisation beyond cryptocurrencies as a means to break up the monopolisation and consolidation on the Internet. These discussions are covered under the selected key topic *Decentralised Power on the Internet*.

Disinformation, misinformation, deep fakes, content moderation, and content censorship were significant talking points at the workshop, while these and similar keywords are heavily featured in datasets of trending keywords in news media and working papers. We selected the key topic *Trustworthy Information Flows* to encompass these issues.

The rights to online privacy and data control were identified as basic digital rights at the workshop. We selected the key topic *Personal Data Control* on the basis of these and similar discussions covering data sovereignty, data management, data infrastructures, data rights and privacy.

A significant point of discussion at the expert workshop was sustainability. The researchers called for more focus on environmental sustainability in the discourse revolving around a next generation future Internet - both in terms of the actual environmental impact of the Internet and how internet technologies can help combat environmental issues - and they argued for sustainability as a stand-alone key NGI topic: *Sustainable and Climate-friendly Internet*.

When asked to prioritise human and societal issues on the Internet, workshop participants consistently prioritised issues such as online harm and online violence. Based on these discussions, we selected the key topic *Safer Online Environments*.

Subjects like the digital divide, access, literacy and education were consistently prioritised highly in the workshop exercises, and they all relate to access and inclusion on the Internet, producing the key topic *An Inclusive Internet*.

The initial umbrella topic *Chinese tech sector*, selected based on trending keywords in D1.2, was criticised at the workshop for being too narrow as a stand-alone key topic and having a pro-Western bias. Instead, we selected the key topic *Competitive European Ecosystems* to encompass support for a competing European internet narrative over the Chinese- and US-centric narratives.

Although most discussions were focused on issues, solutions and values, workshop participants did discuss technologies in relation to the human-centric aspects of technology and ethical implementation and use of technology. To encompass these debates, we selected the key topic *Ethical Internet Technology*.

The topic selection process produced the following eight topics as the first wave of key NGI topics:

- Trustworthy Information Flows
- Decentralised Power on the Internet
- Personal Data Control
- Sustainable and Climate-friendly Internet
- Safer Online Environments
- An Inclusive Internet
- Competitive European Ecosystems
- Ethical Internet Technology

2.4 Further analysis

For each key NGI topic, we manually identified all associated keywords from the most trending NGI-related keywords to inform the key topics and serve as input for further analysis.

To reveal more details about each key topic and enable us to find which emerging keywords were most often mentioned together in the dataset, we secondly conducted a co-occurrence analysis for the keywords in each key topic. These reveal the frequently co-occurring words for each keyword and can be used to identify the areas where a technology is applied, or connections to regulatory issues.

The key NGI topics were also explored through *topic modelling* analysis using the Latent Dirichlet Allocation (LDA) model, a probabilistic *topic model* using Bayesian formulation to reveal hidden *topics* in a given text corpus (Blei et al., 2003). To avoid mix-ups between the academic term *topic* related to *topic modelling* and the actual eight key NGI topics, the academic term is written in *italics* throughout the report.

For the *topic modelling* analysis, we used the associated keywords as input to automatically discover latent themes present in the dataset, providing us with the most important clusters of keywords related to each key NGI topic. In each topic guide in the next section, the six largest of these clusters of keywords (called *topics*) are presented in tables. These serve as explanatory contexts of the discussions revolving around the eight key NGI topics. The size



of these *topics* is determined by the number of tokens in each *topic*, i.e. the number of instances or occurrences of words within the document (e.g., the phrase, "duck duck goose" has three tokens, but only two words).

3 Topic guides

The following section includes a brief description of each of the eight selected key topics and the associated trending keywords. In each topic guide, we briefly explore the co-occurrences and *topic modelling* results related particularly to Europe, and you can explore the co-occurrences further online (<http://46.101.233.156:65442/mandala2.html>) and the largest *topics* from the *topic modelling* analysis in the tables in each topic guide.

3.1 Trustworthy Information Flows

It is widely recognised that trustworthy information flows are essential for healthy democracies, but with social media and the Internet, content can spread much faster and in less moderated ways, challenging traditional information flows. The problem of online mis- and disinformation - often referred to as fake news - has evolved from a journalistic concern to one of the most urgent democratic issues in recent years. Despite major attention from the media, academia and governments, an effective solution is still not available. Coupled with other issues such as governmental censorship and large-scale content moderation by online platforms, information flows are changing rapidly, and further research is needed to explore different solutions that are sustainable and consider often conflicting values.

The co-occurrences reveal various cases in which information flows have changed online and threatened democracy. This includes examples such as election meddling, with ‘European elections’ being mentioned in co-occurrence with ‘disinformation’, ‘misinformation’ and ‘political advertising’, and the vaccination debate, with ‘vaccines’ present in co-occurrence with ‘conspiracy theories’, indicating that misinformation is highly associated with the recent anti-vaccination movement. The *topic modelling* also suggests connections between European elections and disinformation in *Topic 4*, while *Topic 2* suggests that there is a growing consensus among European regulators on the need to regulate to protect the flow of accurate information.

Trending keywords: content crisis, musk tweet, fake news, disinformation, conspiracy theories, political ads, infowars, fact-checking, fake account, anti-vaccine, disinformation campaigns, false information, russian troll, russian interference, deep fake, troll farm, election meddling, influence campaign, content moderation, ad transparency, online disinformation, spread disinformation, censored search, censorship.

Topic modelling (top 10 keywords in the six largest topics)

<i>Topic 1 (41.9% of tokens)</i>	<i>Topic 2 (23.5% of tokens)</i>	<i>Topic 3 (11.9% of tokens)</i>
whatsapp china google app instagram	eu google commission regulation uk	vaccine cambridge_analytica data jones infowar

account india twitter platform 2018	russia measles committee government law	zuckerberg alex_jones privacy gdpr research
<i>Topic 4</i> (11.6% of tokens)	<i>Topic 5</i> (2.4% of tokens)	<i>Topic 6</i> (0.9% of tokens)
disinformation twitter ira european_parliament dorsey research amazon google european_election tweet	google brazilian far-right crypto twitter blockchain brazil edge french nix	collin jack memo amazon zuck youtube china damian_collin cideo ban

3.2 Decentralised Power on the Internet

The Internet was originally designed to be open and decentralised. But the de facto internet of today is controlled by a handful of giant companies with virtual monopoly control, acting as gatekeepers by enforcing policies on their users. However, visions for a more decentralised Internet are gaining traction - an Internet where humans can communicate without relying on big companies that collect data for profit. Some concepts for a decentralised Internet utilize distributed web and blockchain technologies to yield a more open and accessible Internet, while others focus on empowering people to publish and own content on the web outside centralised social media platforms. More research is needed into these solutions, both technical and socio-technical.

The co-occurrences provide insights into existing European attempts to decentralise power on the Internet, such as ‘GAFA’ (abbreviation of Google, Amazon, Facebook, and Apple) being mentioned alongside French President Emmanuel Macron and French Minister of Economy and Finance Bruno Le Maire. This might be pointing towards the announced French “digital tax” that was set to hit some of America’s biggest tech companies but was delayed to avoid a potential trade war with the US. The *topic modelling* points towards a heavy focus on cryptocurrency in the news media’s coverage of decentralisation (*Topic 3* especially), but as *Topic 6* suggests there has also been an interest in competitive start-up ecosystems.

Trending keywords: decentralised computing, crypto, cryptocurrency, blockchain, tech giants, ico, tech workers, blockchain-based, networking giant, gafa, monopolist, coinhive, reward based crowdfunding, hoc network, distributed ledger.

Topic modelling (top 10 keywords in the six largest topics)

<i>Topic 1 (50.9% of tokens)</i>	<i>Topic 2 (18.4% of tokens)</i>	<i>Topic 3 (13.5% of tokens)</i>
ai data iot attack google device security user research network	facebook app apple coinbase huawei user instagram google mine sk_telecom	cryptocurrency token bitcoin ico investor fund trade regulation exchange invest
<i>Topic 4 (3.2% of tokens)</i>	<i>Topic 5 (1.8% of tokens)</i>	<i>Topic 6 (1.5% of tokens)</i>
5g malware ericsson infect trojan amsterdam ransomware attack huawei oracle	au t-mobile tax sprint blockchain supply_chain cba coinbase gdpr bank	aw startup_battlefield vmware vr startup_alley yc quantum_computing kubernet start-up cloud

3.3 Personal Data Control

Recent revelations including the Cambridge Analytica scandal have made clear the lack of control we have over our own data, and the sheer amount of data collected online has created a major privacy concern. New approaches to privacy and data rights are needed in order in order to realise the societal and environmental potential of big data to connect diverse information and conduct rapid analysis - such as data sovereignty, data portability, and collective data rights. Achieving this will require research into the ways policymakers can fit these new concepts into existing data regulation frameworks in a way that offers individuals better control and authority, and builds public trust and engagement.

Among other things, the co-occurrences suggest privacy concerns with facial recognition technology, and especially the use of Chinese technology with ‘Huawei’ being mentioned alongside ‘facial recognition’. The co-occurrences also reveal that the GDPR and the Cambridge Analytica scandal have been the centre of attention in these debates. The *topic modelling* also suggest a major focus in the media on regulation in *Topic 3* and *Topic 5* in different regions including the GDPR in the EU, the Information Commissioner's Office (ICO) in the UK, and the American Civil Liberties Union (ACLU) and the Federal Trade Commission (FTC) in the US. *Topic 1* reveals that the tech giants, including Facebook, Google, Huawei and Apple, have been at the centre of attention in the debate on data protection.

Trending keywords: cybersecurity, privacy, cambridge analytica, gdpr, facial recognition, data breach, user data, equifax, health record, data brokers, privacy settings, privacy scandal, cybersecurity standards, surveillance capitalism, privacy standard, duckduckgo, gdpr compliance, location data, privacy regulation, location history, privacy violation, surveillance system, onavo, privacy-focus, facebook user, facebook data, multi-factor authentication, aadhaar, speech recognition, privacy risk, differential privacy, cloud storage, govern data, digital citizen, cybersecurity law.

Topic modelling (top 10 keywords in the six largest topics)

<i>Topic 1 (59.2% of tokens)</i>	<i>Topic 2 (21.2% of tokens)</i>	<i>Topic 3 (7.8% of tokens)</i>
facebook app user google huawei apple zuckerberg account report people	ai cybersecurity attack au malware cloud singapore organisation 2019 business	bill uk law committee government legislation regulation ico ftc gdpr
<i>Topic 4 (3.9% of tokens)</i>	<i>Topic 5 (3.4% of tokens)</i>	<i>Topic 6 (1.3% of tokens)</i>
5g windows 10 microsoft patient chrome alexa huawei nhs drone device	eu court aclu law rule u. data_protection commission european border	kaspersky dhs whatsapp russian bounty bug_bounty see_also kaspersky_lab hackeron russia

3.4 Sustainable and Climate-friendly Internet

The environmental impact of the Internet is enormous and growing rapidly. Each activity online comes with a small price in terms of carbon emissions and with over half the global population now online, this adds up. According to some estimates, the global carbon footprint of the Internet and the systems supporting it amounts to about 3.7 percent of the total carbon emissions, similar to the amount produced by the airline industry globally (Griffiths, 2020). As the Internet expands into new territory, it is estimated that the carbon footprint of the global internet technologies will double by 2025 (ibid.). Indeed, sustainability should be a bigger priority, and further insights are needed into how emissions could be

controlled, how awareness of the environmental impact of the Internet can be raised, and how internet technologies can be utilized in the fight against climate change.

The co-occurrences reveal ‘data centres’ and ‘e-scooters’ at the centre of attention in news media’s coverage of the environmental impact of internet technologies. Both of these are also present in the *topic modelling*, in *Topic 1* and *Topic 5* respectively. Interestingly, ‘quantum computing’ is mentioned alongside ‘global warming’ in the co-occurrences and in *Topic 5* alongside keywords such ‘renewable energy’, ‘power’ and ‘solar’. Quantum computing requires remarkably little power, while the technology can be utilized for energy system optimization, with ExxonMobil and IBM in 2019 investing heavily in developing next-generation energy technologies using quantum computing.

Trending keywords: climate, blockchain, climate change, global warming, cryptocurrency mining, circular economy, greenhouse gas, paris agreement, ipcc.

Topic modelling (top 10 keywords in the six largest topics)

<i>Topic 1</i> (44.9% of tokens)	<i>Topic 2</i> (15.9% of tokens)	<i>Topic 3</i> (11.9% of tokens)
amazon speci company facebook data water human food technology ai	2018 robot earth record warm temperature 5g atmosphere model planat	facebook plastic mar eu apple vote user company 2019 country
<i>Topic 4</i> (10% of tokens)	<i>Topic 5</i> (4% of tokens)	<i>Topic 6</i> (2.1% of tokens)
eu car fire coal 2019 vehicle china poland paris_agreement bloc	renewable_energy quantum_computing electricity solar energy ai bill student renewable power	hydrogen co2 norway emission scooter fuel plant startup_battlefield carbon_dioxide apple

3.5 Safer Online Environments

People increasingly experience the internet as a hostile space. Cyberviolence in many shapes and forms is a growing concern, and it has a significant impact on an increasing number of people, LGBTQ+, ethnic minorities, women and children in particular. It will be

vital for a more human-centric Internet to build safe online environments. For this to happen, a range of issues needs to be taken into consideration, including the role of social media providers and the protection of free expression. At the same time, solutions need to be investigated, such as effective moderation or containment procedures, creating useful aid for victims of cyberviolence and enabling law enforcement to take action against offenders.

The co-occurrences also point towards the role and responsibilities of the social media providers with keywords such as ‘content moderation’ and ‘tech giants’ co-occurring with ‘hate speech’ and ‘Infowars’, the controversial site that YouTube and others have censored. The far-right and specifically 8chan, an anonymous message board linked to racism, extremism and terrorist attacks, are mentioned alongside ‘racism’ and ‘white supremacism’, indicating these as sources of hate content. The *topic modelling* also suggests an interest in cyber racism and white supremacy online in relation to internet the infrastructure company Cloudflare that took the rare step of pulling the plug on its customer 8chan (*Topic 6*), linking to the debate on free speech versus hate speech.

Trending keywords: health, mental health, #metoo, sexual harassment, hate speech, white supremacist, anti-semitic, sexual misconduct, terrorist content, child abuse, sex trafficking, momo challenge, online terrorist, harmful content, hate content.

Topic modelling (top 10 keywords in the six largest topics)

<i>Topic 1</i> (55.9% of tokens)	<i>Topic 2</i> (15.6% of tokens)	<i>Topic 3</i> (13.5% of tokens)
facebook google user twitter data app amazon report platform apple	ai start-up invest apple cloud alexa singapore 5g business samsung	patient robot ai research study 2018 vaccine human food hospital
<i>Topic 4</i> (1.9% of tokens)	<i>Topic 5</i> (1.5% of tokens)	<i>Topic 6</i> (0.9% of tokens)
chrome dog multi-cloud opioid fertility app browser azur ml android	au nhs cybersecurity australia breach patient eu ransomware record organisation	blockchain bezo tc ipo alibaba hong_kong white_supremacist bee cloudflare ride-hail

3.6 An Inclusive Internet

The Internet offers a potential for inclusiveness in a global and diverse community, but if access is not evenly distributed, the Internet will deepen inequality. Half of the population of the world is still offline, urban areas are better connected than rural, and those that are connected in advanced ways may not be in a position to realise the full potential of the Internet to improve their lives and mitigate against critical issues. Many disabled people also are excluded from using online information and services, so inclusive infrastructures and tools are needed to remove barriers and create an inclusive and accessible Internet for all.

The co-occurrences within this key topic point towards different technologies that enable disabled people to use internet technology such as smart speakers and voice assistants, while both the co-occurrences and *topic modelling* suggest an interest in 5G technologies in this context. While 5G can potentially expand access to technologies like augmented reality, the Internet of Things, autonomous vehicles, 3D printing and wearable tech, the initial adoption of 5G networks may exacerbate the current digital divide, as developing countries are likely to take longer to implement the networks than the developed countries.

Trending keywords: democracy, 5g networks, voice assistant, edge devices, media literacy, netizen, 5g infrastructure.

Topic modelling (top 10 keywords in the six largest topics)

Topic 1 (37.2% of tokens)	Topic 2 (19.6% of tokens)	Topic 3 (15.8% of tokens)
backdoor vulnerability spokesperson huawei boardroom royal found misrepresent china security	dell_emc appliance price_tag lenovo network_equipment screen vodafone samsung edge sd-wan	huawei inclusive 1,500 italian 4gb depiction plastic strive carv religion
Topic 4 (11.5% of tokens)	Topic 5 (2.8% of tokens)	Topic 6 (2% of tokens)
cio diverse inclusive gartner organisation team senior_management tend represent	motorola speculation fcc verizon net_neutrality unfold pocket broadband nbn	implant bbc telecom_giant chinese_government cnet huawei mistake meng ai

female	pai	microsoft
--------	-----	-----------

3.7 Competitive European Ecosystems

Today, the Internet is dominated by two narratives that give little agency to users: the American model, ruled by capitalist market powers with internet giants harvesting massive amounts of personal data to shape human behaviour, and the Chinese model characterised by mass surveillance and government control of the internet. These narratives cannot go unchallenged, and growth and innovation in the European tech industry without acquisitions from the U.S. and China-based companies is needed to support a competing narrative adhering to European values. This requires further research into possible policy and regulatory initiatives that can increase Europe’s competitiveness in the technology sector.

The co-occurrences for this key topic reveal several European attempts at regulating and taxing the tech business, with the French Minister of the Economy and Finance Bruno le Maire being mentioned in relation to France’s “digital tax” and the new President of the European Commission Ursula von der Leyen who has emphasised the need to strengthen the EU’s ability to compete with the US and China on digital technology. European regulation is also central in *Topic 5* of the *topic modelling*, while *Topic 4* suggest an interest in the European start-up ecosystems.

Trending keywords: digital tax, market power, fine google, tech policy, competition policy, paid prioritization, digital taxation, neutrality law, neutrality repeal, competition, Huawei equipment, serverless computing, multicloud, AR/VR, wireless charging, quantum technology, qubits, transform leadership, market disruption, AI start-up, AI expert, AI strategy, facial recognition, openai five, reinforcement learning.

Topic modelling (top 10 keywords in the six largest topics)

<i>Topic 1</i> (39.3% of tokens)	<i>Topic 2</i> (28.1% of tokens)	<i>Topic 3</i> (6.8% of tokens)
facebook google data user people twitter app content report policy	5g huawei apple cloud microsoft amazon custom device ai product	tesla fcc net_neutrality car t-mobile billion vehicle verizon au amazon
<i>Topic 4</i> (5% of tokens)	<i>Topic 5</i> (4.7% of tokens)	<i>Topic 6</i> (1.3% of tokens)
eu brexit climate_change	eu tax google	uber driver lyft

startup_battlefield start-up europe city european student digital	apple law commission fine rule court facebook	mozilla openai gmail browser chrome didi ride-hail
---	---	--

3.8 Ethical Internet Technology

Recent examples, such as Google's censored search engine developed for the Chinese market ('Project Dragonfly'), instances of algorithmic bias in criminal cases, racially targeted ads and "differential" pricing, and the use of Facebook data for voter manipulation, have shown that the Silicon Valley attitude of 'moving fast and breaking things' has failed miserably. With the rapid development of new technologies in the Internet of Things, Artificial Intelligence and Machine Learning, further research is needed in order to develop targeted ethical frameworks for the development and implementation of new technologies.

The co-occurrences here reveal the media discourse around many of the recent scandals associated with ethical technology development, including the mentioned 'Project Dragonfly' and Google's other controversial project in collaboration with the Pentagon, 'Project Maven'. The *topic modelling* also suggests a focus on 'Project Maven' in *Topic 2* and *Topic 5*, while *Topic 6* suggests focus on the ethical aspects of the CRISPR gene editing technology.

Trending keywords: project dragonfly, killer robots, autonomous weapons, misty robotics, aibo, AI ethics, AI ml, blackbox, project maven, full self-driving, c-v2x, cryptocurrency mining, facecoin, smart speakers, face id, screen time, algorithm governance, algorithmic decision making.

Topic modelling (top 10 keywords in the six largest topics)

<i>Topic 1</i> (40.8% of tokens)	<i>Topic 2</i> (18.7% of tokens)	<i>Topic 3</i> (14.1% of tokens)
google facebook ai china data microsoft app platform user employee	ai facebook robot google zuckerberg worker human data military algorithm	game battery product_purchase influence_editorial though_vox via_affiliate affiliate_partnership may_earn samsung vox_media
<i>Topic 4</i> (11.7% of tokens)	<i>Topic 5</i> (5.6% of tokens)	<i>Topic 6</i> (2.5% of tokens)
phone	ai	crispr



headphone camera app device screen apple music ipad iphone	eu pentagon google europe government committee singapore dod cybersecurity	gene-edit car baby embryo vaccine dna vehicle children gm
--	--	---

4 References

- About – AoIR*. (n.d.). Retrieved 1 November 2019, from <https://aoir.org/about/>
- Alminde, S., & Warming, H. (2019). Future workshops as a means to democratic, inclusive and empowering research with children, young people and others. *Qualitative Research*, 146879411986316. <https://doi.org/10.1177/1468794119863165>
- Bego, K. (2018, October 22). Ten Challenges for The Internet. *Next Generation Internet*. <https://ngi.eu/news/2018/10/22/ten-challenges-for-the-internet/>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274. JSTOR. <https://doi.org/10.2307/1163620>
- Griffiths, S. (2020, March 6). Why your internet habits are not as clean as you think. *BBC Future*. <https://www.bbc.com/future/article/20200305-why-your-internet-habits-are-not-as-clean-as-you-think>
- Gyódi, K., Nawaro, Ł., Paliński, M., & Wilamowski, M. (2019). *D1.2: Visualisations of key emerging technologies and social issues*. The European Commission.
- Jungk, R., & Müllert, N. (1987). *Future Workshops: How to Create Desirable Futures*. Institute for Social Inventions.
- Møller, L. A., & Bechmann, A. (2019). *D1.13: Value-driven future internet: A social science perspective I*. The European Commission.