UNIVERSITY OF TARTU
DEPARTMENT OF ENGLISH STUDIES

NEWS MEDIA COVERAGE AND PUBLIC PERCEPTION OF ARTIFICIAL INTELLIGENCE: A CORPUS-BASED STUDY OF TWO BRITISH ONLINE MEDIA SOURCES

BA thesis

MARIA JOHANNA MILDER
SUPERVISOR: ÜLLE TÜRK, M.A.

TARTU
2022

ABSTRACT

The United Kingdom is one of the countries highly indexed for artificial intelligence readiness. Artificial intelligence is an ever-growing technology which, due to its broad term, is often not understood too well by the public. Online news media is one source which has the potential power to influence public knowledge and opinion on certain topics. The aim of this thesis is to examine the coverage of AI in two British online newspapers, *The Guardian* and *MailOnline*, during the period 2019-2021, and to compare the findings to the general public's perception of AI.

To achieve the aim, an overview of AI in the UK, as well as of the news media in the UK is provided. Additionally, Chapter One includes an overview of topic-relevant studies in relation to AI in the news media, with particular emphasis on British news media, and public perception of the technology. In Chapter Two, the data collection method, the corpus software Sketch Engine used for the corpus-based analysis, findings as to how AI is presented, what dissimilarities in the different news media occurred during 2019-2021, and whether public perception of AI matches that of the news media coverage are presented. The conclusion summarizes the main findings of the analysis.

TABLE OF CONTENTS

ABSTRACT 2
INTRODUCTION 4
1. NEWS MEDIA COVERAGE AND PUBLIC PERCEPTION OF ARTIFICIAL INTELLIGENCE 10
1.1 Artificial intelligence application and recent developments 10
1.2 Media Coverage of AI 13
1.3 Public understanding of AI 16
2. CORPUS-BASED STUDY OF THE COVERAGE OF ARTIFICIAL INTELLIGENCE IN *THE GUARDIAN* AND *MAILONLINE* DURING 2019-2021 22
2.1 Method 22
2.2 Findings and Discussion 27
CONCLUSION 39
REFERENCES 41
APPENDICES 55
Appendix 1. Public perception of AI according to research; results which were only present in American surveys are marked in blue, and results which had a high number of results between 2019-2021 are marked in green. 55
Appendix 2. Top 100 keywords and multi-word terms in the AIC, TG and MO, and top 50 in 2019-2021 sub-corpora. 56
RESÜMEE 60

INTRODUCTION

In 1997, reigning world champion Gary Kasparov was beaten by Deep Blue, IBM's chess-playing computer program, which astounded many and was largely the introduction of artificial intelligence (hereafter AI) to the general public due to extensive publicity (IBM n.d.). Today, artificial intelligence systems are being deployed in all areas imaginable with immense potential for change and impact on society. The term, AI, is used quite liberally to describe anything from routine data analysis to complex deep neural networks which do not need constant human intervention (Committee on Standards in Public Life 2020: 12; Gov.uk 2019: para 11). Though a broad term, it has been well explained by Nilsson (2010: 13), a leading AI researcher: "Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.". Thanks to this data-enabled foresight, AI will be able to perform increasingly complex tasks (Committee on Standards in Public Life 2020: 12). This thesis will focus on AI as a type of technology based on self-learning algorithms that can potentially benefit the development of companies and society. Thereof, AI will be used as an umbrella term, compiling different AI technologies (i.e., machine learning, automation and robotics).

The AI frontier is not fixed but constantly advancing and, thus, countries that do not continuously develop their technologies could fall behind. McKinsey (2019: 4) reports that the United States and China are the leading countries in AI research and employment, whereas in Europe, one of the countries highly indexed for AI readiness is the United Kingdom. Specifically, the UK is in the top quartile globally for research, start-up investment, digital absorption, innovation foundation and ICT (information communication technology) connectedness. It does, however, rank lower on automation potential and human capital

(McKinsey 2019: 4). UK's Committee on Standards in Public Life (2019: 15) also revealed that actual adoption of AI in the public sector remains limited, thus revealing that the British public might not be highly aware of AI and its utilization. However, an area which the UK is already recognised in is technology for healthcare (McKinsey 2019: 1; Committee on Standards in Public Life 2020: 15) and as having the largest software industry in Europe. Additionally, the UK ranked third in the world for private investment into AI companies in 2020, behind only the US and China (Office for Artificial Intelligence 2021: 10) with the US investing 50 times more and China investing eight times more into AI than the UK. Even promising UK start-ups (including DeepMind, SwiftKey and Magic Pony) are being increasingly acquired by large US companies (Google, Microsoft and Twitter – for €410 million, €200 million and €120 million, respectively) before they can mature, limiting the UK's ability to keep and advance its position as a leader (Gerner 2020). As for the future potential, there is high investment into AI, with Grech, CEO of Tech Nation, pointing out that UK firms that were adopting or creating AI-based technologies had received €2.1 billion in funding in 2020, compared to for example €464.7 million raised in Germany, who is also in the top quartile of AI countries and ahead of the UK in some aspects (Gov.uk 2021; McKinsey 2019). Evidently, the UK government has adopted the aim of increasing its involvement in AI. This claim is further supported by the fact that the UK was among the top three countries that passed the highest number of AI-related bills in 2021 (Zhang et al 2022: 12). The reason for this could be that AI could add approximately an incremental 22% (McKinsey 2020: 2) or €279 billion (Committee on Standards in Public Life 2020: 12) to UK's GDP by 2030. This is seen as a paramount growth potential as from 2010 to 2015, productivity grew at only 0.2% a year (McKinsey 2019: 2). The effect that the UK

government estimates AI will have, has increased as recently as from 2019 to 2021: from 5% of GDP (Gov.uk 2019) to a predicted 10% of GDP by 2030 (Gov.uk 2021b).

The UK's substantial involvement in regard to AI largely started in 2017 when the Alan Turing Institute was made the national institute for AI and data science by the government, which examines AI as well as its impact on society. In the same year, the government published its 2017 Industrial Strategy that identified AI and data as one of four 'Grand Challenges' to modernise the UK economy (Gov.uk 2021). The €1.1 billion AI Sector Deal followed in 2018 in which the UK government set a clear goal of becoming a global leader in AI (Gov.uk 2019). This led to the creation of three new bodies: a government Office for AI; an industry-led AI Council; and the Centre for Data Ethics and Innovation (CDEI), which focuses on identifying the measures needed to make sure the development of AI is safe and ethical; currently reviewing public concern on issues such as data bias and online targeting (Committee on Standards in Public Life 2020: 13-14; Gov.uk 2019; Office for Artificial Intelligence 2021: 13). As recently as in September 2021, the government introduced a new ten-year plan called the National AI Strategy in which they outlined how the UK can continue to strengthen its position through "recognising the power of AI to increase resilience, productivity, growth and innovation across the private and public sectors." (Office for Artificial Intelligence 2021: 7). Some of the activities include publishing different frameworks and roadmaps, ensuring that AI programmes are accessible for schools, new visa regimes to attract the world's best AI talent, and monitoring diversity and safety. One of the proposed outcomes was also that public trust in AI would be improved. (Office for Artificial Intelligence 2021: 7-14)

Currently, the AI gap between Europe and the present leaders in AI (the US and China) has increased by 20% from 2017 to 2019 (McKinsey 2019: 3), which is substantial given that

reports agree that AI is going to have a profound effect on societies and economies (Perrault et al 2019; Crawford et al 2016). This is highlighted also by Ouchchy (et al 2020) who observes that the use and application of artificial intelligence has been significantly increasing over the past years. The UK is considered an AI superpower, with particular strengths in research, investment and innovation and, therefore, it is important that public trust is increased. Without it, organizations lack the confidence to invest in data-driven solutions, while citizens are hesitant to participate in data-sharing schemes and adopt data-driven technologies (Spiro et al 2022). Developers and AI companies already fear that the public is mainly met with negative depictions of AI in the media (SCAI 2018) by creating an image of an AI which is still distant (Naughton 2018). One of the main reasons for this is the aim to attract traffic by being sensationalist (Brennen et al 2018; SCAI 2018). Stone (et al 2016: 5) argues that if society approaches AI with fear and suspicion, "missteps that slow AI's development or drive it underground will result, impeding important work on ensuring the safety and reliability of AI technologies". Therefore, citizens need to acquire a minimum understanding of this technology to formulate their opinion on its uses. As the public mostly receives information from the news media, their awareness, knowledge, opinions and even behaviours are influenced by the way the news media cover the topic of science and technology (Chuan et al 2019; Dudo et al 2011). Therefore, it is important that the topic is covered extensively, or at least informatively. Siapera (2017: 29) argues that, although there is much debate over the precise nature of the relationship between the media and society, the increasing significance of the news media is still undeniable. Therefore, the articles published in the news media offer an opportunity to examine how a particular topic is presented to the general public.

In relation to the UK, the Select Committee on Artificial Intelligence (SCAI 2018) states that the British public is not fully aware and often has a distorted view of the AI topic. According to research by communications regulator Ofcom (2021: 2), 73% of adults consume news through the Internet. Two of the most widely read digital titles in the UK are *The Guardian* and *MailOnline* (Ofcom 2021; Newman et al 2021: 62). *The Guardian* is a left-leaning British daily quality newspaper, and *MailOnline* is the online version of *The Daily Mail*, a right-leaning British daily middle-market tabloid newspaper. The former has been praised for its investigative journalism, its objective discussion of issues, and criticism (Britannica 2021), whereas *The Daily Mail* has historically been known for its coverage of foreign news (Britannica 2019). The distinction between broadsheets or 'quality' newspapers and 'middle-market' newspapers can be made in terms of their content and language: the former has a higher news content and generally more in-depth articles with longer sentences and paragraphs; the latter is a mix of entertainment as well as coverage of significant events; thus neither is sensationalist in content (BBC n.d.; Chandler and Munday 2011).

For the reason that the UK is considered to be one of the best countries for AI growth, it is important that the British news media cover the topic informatively. Public perception can fuel what direction development takes; therefore, the future of AI in the UK also depends on how the media covers it and, thus, what the public will read out of it. The topic of public understanding of and opinion towards AI has been more prominent in academic writing, whereas less has been written about how the news media covers AI. Of the few, Brennen et al (2018) looks at an eight-month period in the UK news media, Duberry and Hamidi (2021) concentrates on AI in relation to the coronavirus pandemic, and Sun et al (2020) conducts an analysis of 40 years of American newspapers.

The aim of this thesis is to examine the coverage of AI in two newspapers representing the British news media during 2019-2021 and to compare these to the general public's perception of AI. Specific research questions include:

RQ1: What keywords are most relevant in British news media about AI?

RQ2: How do *MailOnline* and *The Guardian* differ in their coverage of AI?

RQ3: How have AI keywords changed between 2019-2021?

RQ4: Do different survey results of British public perception of AI match the ideas portrayed by the news media?

To achieve the aim, an overview of AI, its application and everyday examples is provided. Additionally, the first chapter includes an overview of topic-relevant studies in relation to AI in the news media and public perception of the technology. In the second chapter, the data collection method, the corpus software Sketch Engine used for the corpus-based analysis, findings and a discussion are presented. The conclusion summarizes the main findings of the analysis.

1. NEWS MEDIA COVERAGE AND PUBLIC PERCEPTION OF ARTIFICIAL INTELLIGENCE

This chapter serves as the background to the study with an overview of AI, its applications, risks and benefits, as well as everyday examples of its use in order to introduce the technology. This is followed by an outline of previous research done in regard to AI in the news media, with a particular emphasis on the British news media. After that, the topic of public perception of AI is discussed, as the analysis in Chapter Two will include a discussion of how public perception matches news media coverage of AI.

1.1 Artificial intelligence application and recent developments

AI is handled as a branch of computer science which is based on self-learning algorithms. It has the capability to mimic and potentially extend human intelligence by automatically extracting data, understanding its patterns and correlations, learning from said patterns, making data-driven decisions, and solving problems (Deloitte 2017: 2-4; Deng 2018: 180; Committee on Standards in Public Life 2020: 12). Other cognitive capabilities include, for example, motion (robotics), creativity and dialogue. The latter is made possible by natural language processing (NLP), which enables AI to read, comprehend and generate human languages (Deng 2018: 180). It is important to note that "fully functioning AI systems do not exist yet, and it has been estimated that they will be with us anywhere between 2029 and the end of the century" (Everitt et al 2018; Makridakis 2017: 52). It means that the AI of 2022 is still considered 'weak AI' or 'narrow AI'. The theoretical form of AI which has advanced to human-like intelligence levels, is referred to as Artificial General Intelligence (AGI) or, as it is commonly called, 'strong AI' or 'true AI' (Insider 2020; IBM Cloud Education 2020). While weak AI focuses on performing a specific task, such as answering questions based on user input or playing chess, and relies on human interference (data input, parameter defining), strong AI

can perform a variety of functions, eventually teaching itself to solve problems (IBM Cloud Education 2020), thus replicating the full depth and breadth of human skills and cognition.

The main purpose of AI is to execute tasks that are complicated for humans or, if continuously performed by humans, lead to a decrease in effectiveness. Yet, although AI offers massive gains in efficiency and performance to practically all industry sectors, not only has its application increased but also concerns regarding its uses. Figure 1 presents the lists of the risks and benefits of AI as highlighted by industry experts and researchers compiled by the author. The lists do not include long-term risks and opportunities (i.e. existential risks) that are foreseen by some (Ord 2020: 37) as these are associated with a super-intelligent AI, which is still considered distant (Naughton 2018; Makridakis 2018: 52) and such views are even met with criticism (Dafoe 2020: para 9).

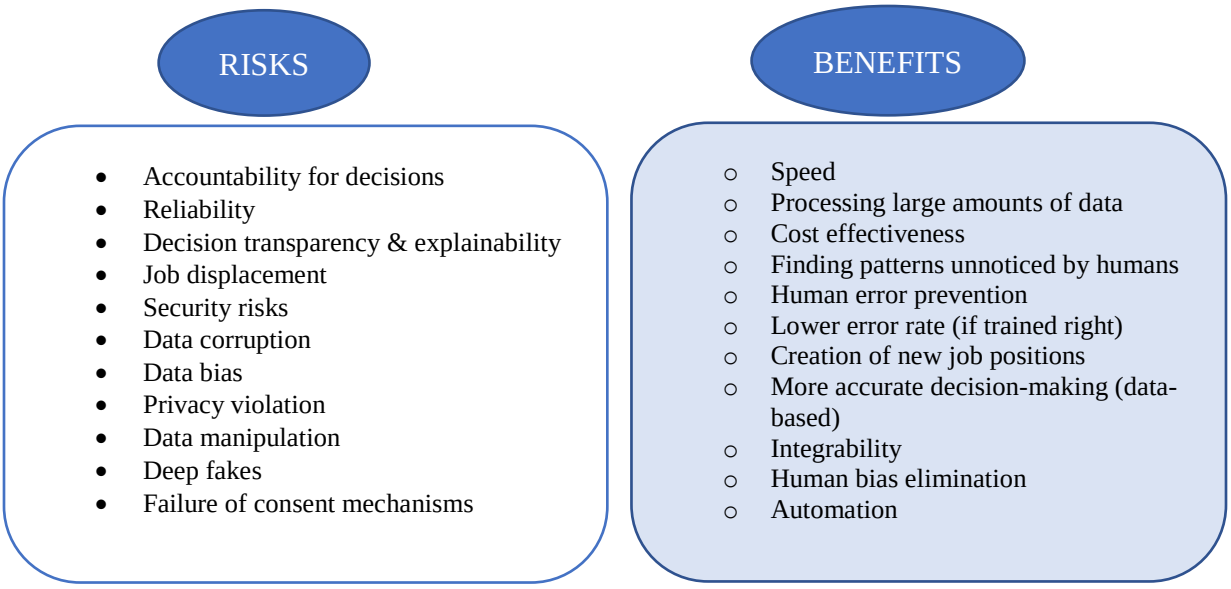

Figure 1. Overview of the risks and benefits of modern AI.
Source: compiled by the author, based on: Brundage et al 2018; CDEI 2020; CDEI 2021; Committee on Standards in Public Life 2020; Dafoe 2018; Deng 2018; Everitt et al 2018; Gov.uk 2019; Hall et al 2017; Informatics Europe & EUACM 2018; Nadimpalli 2017.

The risk of bias is one of the more prevalent discussion points although the Committee on Standards in Public Life (2020: 27) argues that AI systems will be no more biased than the

human processes they are replacing; there is also the potential benefit of *human* bias elimination. Additionally, some existing systems are designed in a way that makes it impossible to measure bias but AI can, in some cases, successfully identify and reduce said bias, or at least help measure it more statistically. Nevertheless, data scientists are sceptical when it comes to the idea that any AI system could be completely free of bias. Due to this, research on AI fairness and transparency has exploded since 2014 (Zhang et al 2022: 11). Other risks include transparency (i.e., whether it is explained exactly what and how will be done with the collected data) and security risks.

On the benefits side, AI can help find patterns unnoticed by humans, reduce human error, create more precise analytics, and turn data collecting devices into powerful diagnostic tools. One example of this is wearable devices such as smartwatches and fitness trackers.

In 1950, British mathematician Alan Turing was the first to envision AI (computers, in his case) as algorithms that are able to emulate human intelligence. Five years later, AI was formally initiated after it was coined by computer scientist John McCarthy for a workshop at Dartmouth University (Deloitte 2017: 2; Nilsson 2010: 42). From there on, the field has experienced exponential growth of data and connected devices, algorithms, and faster processing through the use of cloud computing (Deloitte 2017: 2; Deng 2018: 180). An important advance in AI has been machine learning, which is present in technologies from machine translation to self-driving cars and is the most widely-used form of AI (Gov.uk 2019; Hutson 2017).

AI is used in virtually all businesses, across a range of industries from banking and finance, to healthcare and transportation; in fact, most humans interact with it in some capacity on a daily basis. This brief section will showcase some everyday examples of AI: media

suggestions (e.g. the audio-streaming service Spotify's "Discover Weekly" playlists for each individual user), product recommendations, spam filters and navigation apps, self-driving cars, chatbots (customer service 24 hours a day), web search, robot vacuum-cleaners, noise-cancelling headphones, search engines, face-recognition services (face ID), payment transaction validation and verification, and IoT devices (wearables, i.e. smartwatches) (Deloitte 2017: 2, 5, 16; Deng 2018: 174; Nilsson 2010: 616; Bar Standards Board 2019: 3; Brundage et al 2018: 9; Insider 2020; Office for Artificial Intelligence 2021: 16). It can be observed that AI technologies are becoming embedded in people's everyday lives.

1.2 Media Coverage of AI

The year 2016 was described by Microsoft CEO Satya Nadella as "the year of the bots" (Bruner 2016). There were advances in speech interfaces using NLP algorithms, as well as growth in investments in AI start-ups. This, combined with warnings about the dangers of AI from well-known scientists and innovators such as Stephen Hawking, Bill Gates and Elon Musk, as well as unrealistic predictions from the news media, led to an increased hype around the topic (Deloitte 2017: 5).

Yet, there have been few studies conducted analysing the reporting of AI in the news media. In fact, the author was able to find only one solely UK-based research. Brennen et al (2018) examined at an eight-month period across six mainstream news outlets using a mixed-method analysis approach. Their key findings include: about 60% of news articles are indexed to industry products, initiatives, or announcements; right-leaning outlets highlight issues of economics and geopolitics (automation, national security, investment), whereas left-leaning outlets highlight issues of ethics of AI (discrimination, data bias, privacy).

Sun et al (2020) focused only on American newspapers, reasoning that the US is the leader in AI development with the methods of content analysis and media framing theory. This led to them coding fourteen topics. The topic with the most articles was robot/humanoid, followed by brain/life science and regulation/policy. Some examples of the keywords manually selected to represent each theme, respectively, include: *robot, movie, android*; *brain, scientist, science*; and *government, regulate, military*. Additionally, they looked at potential argumentation patterns. The most prevalent were labelled as pragmatic patterns (a variety of applications; everyday utility - robots), relativizing patterns (AI utility is overestimated) and economically optimistic patterns (AI boosts the economy).

Duberry and Hamidi (2021) examined the portrayal of AI specifically in the context of the COVID-19 pandemic during the month of April 2021 in French and English-speaking mainstream media. They found that no news article provided a definition of AI (thus relying on the readers' knowledge) and that most articles highlighted the technology in the context of information processing and adaptability. Additionally, that American newspaper coverage is more optimistic than pessimistic, whereas European newspapers (Switzerland, France, the UK) present a more balanced perspective of the risks and benefits associated with AI.

Similarly, Bunz and Braghieri (2022) conducted an analysis of AI applications in the healthcare sector with an emphasis on AI systems replacing or outperforming medical experts and/or being personified (i.e. discussed more as 'having intentions' and being a 'boss' than as a system). Here, Bunz and Braghieri took both the US and the UK into consideration (The Wall Street Journal, The Daily Telegraph and The Guardian). The results indicate that instead of replacing humans, AI systems are mostly linked to outperforming them. Bunz highlights this as an issue: that outperforming human expertise places AI above critique. The most frequent word

in the titles between 2015 and 2019 is ‘technology’ or its variant ‘tech’, followed by ‘AI’, ‘intelligence’ and ‘health’. Notably, words ‘robot’, ‘doctor’, ‘future’ and ‘medicine’ are less frequent. They observed that the framing of AI systems in the healthcare sector is technology-related, not healthcare-related.

In regard to stakeholders, Brennen et al (2018) concludes that 33% of unique sources are from the industry sector (i.e., from company executives, researchers, employees), 17% from academia (including researcher employed by a university), and about 5% from government sources. This is supported by Bunz and Braghieri (2022) who found that Google is by far the most cited stakeholder, even in the healthcare sector, almost double the times of Facebook. Overall, by far the most cited entities or people are from the tech industry or business-related (Google Facebook, Microsoft, IBM) (Sun et al 2020; Bunz and Braghieri 2022), followed by public institutions (e.g., the American Food and Drug Administration, the European Union and the World Health Organization) and universities (e.g., MIT, Oxford University and University of Pennsylvania) (Bunz and Braghieri 2022). The most referenced individual is Elon Musk (Brennen et al 2018). Sun et al (2020) explored the topic further and expanded it to include nations and found that China, the US and Japan are the three nations mentioned most frequently.

To sum up, Bunz and Braghieri (2022) and Duberry and Hamidi (2021) focused on AI in relation to the healthcare industry, with the latter specifically on the coronavirus pandemic. Regarding the time period examined, Brennen et al (2018) concentrated on eight months in 2018, Bunz and Braghieri (2022) on one month in 2021; Sun et al (2020) included articles from 1977 to 2018, Duberry and Hamidi (2021) from 1980 to 2019. Sun et al (2020) is US-based, Duberry and Hamidi (2021) Europe-based with one British news media source, and Bunz and Braghieri (2022) is both US- and UK-based. Taking the limitation in previous research into

consideration, this study will examine the representation of AI across industries; secondly, include articles beginning from the end date of the last related study[1], thus starting from 2019 to 2021; and lastly, considering the UK is now recognized as one of the leading countries in AI development, this study will follow the example of Brennen et al (2018) and be UK-based in its scope.

1.3 Public understanding of AI

Public perception of a subject as vague and varied as AI can be difficult to pinpoint, in addition to being likely to change quickly with every new innovation, scandal or statement by a famous individual. It is only recently that people have started to distance AI from the creations of Hollywood (SCAI 2018: 23). It is pertinent to understand how the public perceives AI, which can then be linked to what aspects should the public be educated on, for example, by the news media; and if the news media has succeeded in this task.

As previously mentioned, AI is a broad term but not only is the term broad but also the public's understanding of it. Although media coverage on the topic of AI is frequent, people from the UK are confused by what it actually is and how it works - leading to a distorted view of the topic (Gallup 2019; Holder et al 2018; SCAI 2018; Kantar Public 2019; CDEI 2022). In 2019, 63% of people reportedly knew something about AI (PAS 2020: 6), whereas in 2021, CDEI (2022: 9) found that 90% have heard of the term "AI", potentially indicating an increase in knowledge or at least in exposure to the topic. Yet, only 13% of the respondents felt they could explain the term 'artificial intelligence' (CDEI 2022: 48). Holder et al (2018: 7) also conducted a survey and found that the a large proportion of the British public (39.5%) consider

[1] Bunz and Braghieri (2022) analysed news media coverage in 2021, but it was only for the duration of one month and regarding the healthcare sector

their knowledge of AI ‘limited’, whereas only 8.4% claim to know ‘a lot’ about it or consider themselves experts; 16.4% have heard the term but are unsure of its meaning; and 9.1% claimed to have never heard of it. Another survey reported that those who ‘knew a lot’ about AI was 12% in 2019 (Kantar Public 2019: 1), thus having grown. Holder et al (2018: 8) expanded on their findings and tested respondents' knowledge. The findings link with previous and future findings (SCAI 2018; Gallup 2019) that people tend to have a false perception of AI, even those who claim to understand it. For example, their survey concluded that 1 in 5 respondents think AI is currently capable of modifying itself, and 1 in 6 that predicting human actions was in its capacity. These responses were high amongst those who had said they ‘knew a lot about AI’, which indicates that even those claiming to be knowledgeable of AI, still do not understand the exact capabilities of the technology. Additionally, 62% of respondents claimed they had had no contact with or had not used an application which utilizes AI, whereas only 15% said they had used or had had contact with AI. This is explained by Zhang and Dafoe (2019) who state that most do not associate, for example, Google Translate, tagging people on Facebook, Google Search, and Netflix recommendations with AI. Thus, public understanding on this topic is mostly surface-level. Furthermore, Nader et al (2022: 6) enquired after what technologies Americans specifically use and found that most often, either regularly or not, they use digital assistants, predictive texts, wireless networks, and digital recommendation systems. On the other hand, the technologies that most admitted to never using included self-driving cars, virtual reality gaming, smart home devices, and big data. Of these technologies, respondents did not associate AI with the following: wireless networks, big data and digital recommendation systems. Another survey found that British people are uncomfortable with AI being used to power Internet search engines which is a common practice (CDEI 2022: 48).

Furthermore, when examining the different terms associated with AI (e.g., machine learning, robotics, automation), Zhang and Dafoe (2019) demonstrate that most Americans tend to have some knowledge of the terms but they are not aware how widespread AI is. Whereas only 9% of the British public said they had heard the term 'machine learning', it turned out that they are in fact familiar with machine learning's different applications, as the majority recognized at least one of the eight examples given (Ipsos MORI 2018: 22).

Data, which AI technologies use, is another topic that warrants exploration. CDEI (2022: 24) reports that British people do not know how their personal data is being collected and used, in addition to uncertainties whether organisations are transparent in this regard. Data security was indeed seen as the biggest risk for data use in society (could be hacked, stolen or even sold). Interestingly, risks of bias and job loss were at the bottom of risks perceived with new technologies (CDEI 2022: 27). According to another survey, willingness in the UK to share personal data is highest in relation to improving medical research and care, which, in fact, is higher in the UK than in the rest of the European Union by 7% (Eurobarometer 2020: 2). In fact, AI in healthcare is one of the aspects people feel most positive about and could see the greatest potential for benefits, for example, by faster and more accurate diagnosis of diseases (Ipsos MORI 2017; Kantar Public 2019; PAS 2020; Nader et al 2022; CDEI 2022). However, it was felt that human involvement was pertinent to ensure that personal contact continues where it is needed (Ipsos MORI 2017: 32)

As to AI's impact, Pew Research Center (Smith et al 2017) revealed in 2017 that Americans tend to observe AI as a topic of apprehension, not enthusiasm. On the other hand, Gallup (2018) points out that, when it comes to AI in general, public discussion in the US and the UK has remained optimistic rather than negative, but some topics (e.g., job and economic

impact) are generally regarded as potentially having a growing negative correlation. To support this, Deloitte (2017: 11) found that most people believe AI will empower work instead of replacing humans. This was similar to the findings of Holder et al (2018) regarding specifically the perceptions of the British public. When asked when their current skills will become obsolete, the US is more optimistic than the UK. Of those who think their skills will become obsolete, a large proportion in the US (22.4%) answered that it will happen in 10 years or more, whereas people from the UK thought it would happen in 1 to 4 years (20.9%). People in both countries though, think that AI will eliminate more jobs than it will create (64.3%) (Gallup 2018). Experts, on the other hand, say that AI and automation will generate more jobs than they will displace (World Economic Forum 2020). However, people rarely fear to lose their *own* job and is therefore perceived as a problem which will affect others (Gallup 2018; Archer et al 2018: 9).

In regard to trust, the National Health Service (NHS) is highly trusted (89%) (CDEI 2022: 64), whereas social media companies and the government are the least trusted entities to act in people's best interest (36% and 39% trust, respectively). Notably, big technology companies (24%) are more trusted (60%) than social media companies. The academics (academic researchers at universities), police and banks are trusted by most (76%, 72% and 71%, respectively) (CDEI 2022: 64). Respondents with the lowest digital familiarity expressed high trust in the police and banks, but were less likely to trust academics (CDEI 2022: 63).

Some surveys even enquired about specific emotions and words respondents would associate with AI (Kantar Public 2019; CDEI 2022; Nader et al 2022). Kantar Public (2019: 2) found that most respondents (64%) were excited to see what AI can do, but still felt that more research is required (82%) on how AI will be used in everyday life. More recently (Figure 2), AI was predominantly associated with being 'scary' and 'futuristic', as well as feeling

'concerned' and 'nervous' (CDEI 2022: 50). Surveys were similar in their findings in regard to respondents' strong association of AI with 'robots' and 'computers' (CDEI 2022: 50; Nader et al 2022: 6); with advanced but existing technology in general, such as speech recognition or virtual reality (Nader et al 2022: 6). It is noteworthy that a shift in perception can be perceived as in 2019 AI was still often related to science fiction (PAS 2020), whereas in 2022 only 11% of people perceived AI as a futuristic technology (Nader et al 2022: 6).

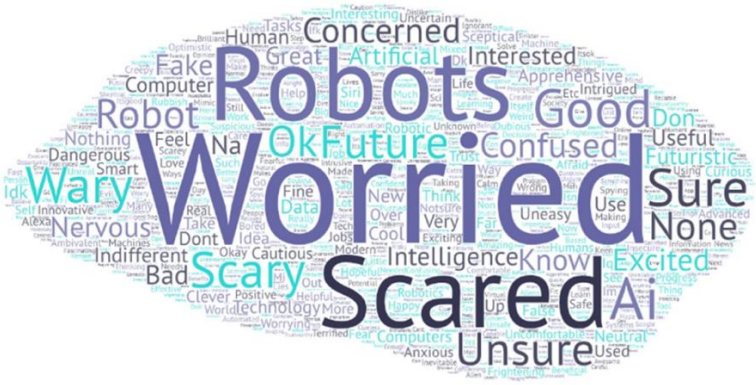

Figure 2. Feelings about 'Artificial Intelligence'.
Source: CDEI 2022

Research regarding public understanding of AI mostly coincides but disagreements can be found in different age groups' attitudes. For example, Gallup (2019: 13) observes that in the UK, the younger the member of the public, the greater the level of worry regarding job loss (due to automation). However, other research (Holder et al 2018; CDEI 2022) has found that younger people in the UK are more comfortable with AI and the automation of jobs.

Public understanding of AI and its associated terms is not extensive, even by those in the general public who claim to have some knowledge of it. Generally, people view AI optimistically but approach the topic with caution when it comes to more specific aspects. On the topic of job automation, public perception differs from those of experts who stress

favourable outlooks, further asserting that people have a false understanding of AI. Where these research findings lack, however, is that they have not been linked with how people consume information about artificial intelligence. More importantly, as the public mostly receives information from and is influenced by the news media, it is important that the public is educated on the aspects that they are least familiar with by the news media.

An overview of the terms associated with AI, what the public does not understand, what has negative and what positive connotations regarding AI can be found in Appendix 1.

2. CORPUS-BASED STUDY OF THE COVERAGE OF ARTIFICIAL INTELLIGENCE IN *THE GUARDIAN* AND *MAILONLINE* DURING 2019-2021

This chapter analyses the coverage of artificial intelligence in the British online news media on the example of *The Guardian* and *MailOnline* and compares it to the public perception of AI in the UK as determined by previous studies. The chapter describes the methods and procedure of the study discusses its findings as to how AI is presented, what differences in the different news media occurred during 2019-2021 and whether public perception of AI matches that of the news media coverage.

2.1 Method

The aim of the study was to examine the coverage of AI in two British news media sources and to compare these to the general public's perception of AI. To this purpose, an electronic text corpus ('AI corpus', hereafter referred to as AIC) was compiled from two online UK news sources, by searching for 'artificial intelligence' during the period of 01 January 2019 until 31 December 2021. The set of pragmatic decisions that presided over the corpus building phase will be discussed in detail below.

The outlets were strategically selected on the basis of the following criteria: that they were national and UK-oriented in their scope; that they publish articles based on news reporting on a daily basis; that they have a large number of digital readers; that they represent different types of news media (quality paper and middle-market tabloid); and lastly, that they represent different political leanings (liberal and conservative), providing the means for comparison of two different types of news media. Two of the most widely read digital titles in the UK are *The Guardian* and *MailOnline*, which are also of different political views and formats with the

former being a more liberal quality paper and the latter a more conservative middle-market tabloid. The features of the two chosen outlets are outlined in Table 1.

Table 1. Overview of chosen news media sources.

Newspaper	Established	Digital Delivery (Montly)	Type	
The Guardian	1821	18,422,000	Liberal	Quality paper
MailOnline	1896	22,430,000	Conservative	Middle-market tabloid

Source: Newsworks n.d.; Britannica 2019; Britannica 2021.

Taking the limitations in previous research into consideration and for the analysis to be as up-to-date as possible, this study will include articles from the end date of the last related study conducted by Brennen et al (2018) thus starting from 2019 and ending in 2021. Additionally, the time period allows for comparison to be drawn between the three years.

The articles were collected from the LexisNexis archive (following the example of Duberry and Hamidi 2021, and Sun et al 2020) through targeted search of the phrase 'artificial intelligence'. The author decided to focus only on 'artificial intelligence' and not its related phrases, such as 'machine learning' and 'robotics' (as opposed to Sun et al 2020), because as mentioned in 1.2, people do not necessarily associate the terms with 'artificial intelligence'.

The primary search produced 276 articles for *The Guardian* and 417 articles for *MailOnline*. The exclusion criteria included: if (a) 'artificial intelligence' was not mentioned in the headline or lead section, and (b) article had less than 100 words, as such articles do not provide sufficient information on the dominant AI keywords promoted by the British media outlets (Sun et al 2020). After reading 50 sampled pieces, the author found that some of the articles only mentioned AI in passing, without presenting any substantial discussion of the topic. Although past studies (Sun et al 2020; Burcher et al 2016) have utilized the approach of having the keyword appear at least twice in the full text in addition to at least once in the title or lead

paragraph, in this case this prompted articles which may have mentioned AI but were still irrelevant to the topic[2]. LexisNexis enables a relevancy score "based on the term's frequency, location in document, and how much discussion of the topic occurs" (LexisNexis n.d.: 3). Thus, to focus on the news substantially relevant to AI, the following rule was applied: the relevance score had to be 80% or more (major or strong passing reference) to be included. The author then screened articles by manually removing the following: book, stage, videogame, film and TV reviews; obituaries; letters and other genres not classified as news (Sun et al 2020). This yielded 171 news articles for *The Guardian* and 315 for *MailOnline*. The full-text articles were downloaded from the newest to the oldest to form an overall news corpus (AIC) of 486 published items. Altogether the corpus amounts to 417,523 words, with a 5.9% type-token ratio (TTR = unique words / tokens x 100) as a raw measure of lexical diversity (Paganoni 2019). AIC consists of two sub-corpora which are *The Guardian* (TG) sub-corpus, consisting of about 154,040 words, and the *MailOnline* (MO) sub-corpus, consisting of about 263,482 words.

The Corpus Linguistics software used for the analysis is the March 2022 version of the Sketch Engine software. Sketch Engine's features for analysing different aspects of corpora include word sketch, word sketch difference, wordlist, concordance, thesaurus, n-grams and keywords (Kilgariff et al 2014). Keyword and term extraction showcases units which are typical of a focus corpus or which "define its content or topic" (Sketch Engine n.d.-a). Keywords and term extraction n.d.). Keywords are individual words and terms are "multi-word expressions" which appear more frequently in the focus corpus than in the reference corpus, and are displayed as lemmas. Mainly nouns and adjectives will be included in the result since the frequencies of

[2] For example, this 2021 *MailOnline* article titled: "'Armed and dangerous' MIT grad, 29, is sought in shooting death of newly engaged Yale student - as it's revealed cops quizzed him over killing before he went on the lam"; 'artificial intelligence' is mentioned five times but only in relation to the person's occupation.

other parts of speech tend to be similar in all texts. This corpus-assisted study employs the keyword, concordance and word sketch features. This approach enables much larger amounts of material to be analysed than in conventional content or qualitative thematic analysis (Seale 2012: 476). While a simple word list provides frequency, a keyword list compares the frequencies of one corpus against another in order to determine which words occur statistically more often in one or the other and thus, the keywords produced indicate that the given words have high saliency and that they are worth exploring (Baker 2006: 124-125). The limitation of keyword lists is that they lack context. Fortunately, it is not difficult in keyword analysis to discover the context in which words are being used in order to further assess predominant meanings. Concordance enables to explore the "company that words keep" (Firth 1957).

The corpus software Sketch Engine was subsequently implemented to extract keywords and multi-word terms by keyness, a score computed by a statistical formula that shows what words are typical of the focus corpus because of their higher (or unusual) frequency against a reference corpus (Scott 1997: 236), here the English Web Corpus (hereafter enTenTen20). As the name suggests, the enTenTen20 is an English corpus made up of texts collected in 2020 from the Internet and, as of April 2022, consists of 36.5 billion words (Sketch Engine n.d.-b). The corpus has been built using a web crawler, which can effectively avoid data not suitable for text corpora (e.g., forms, advertisement) (Suchomel and Pomikálek 2020). The N variable, "a variable which allows the user to focus on higher or lower frequency words" (Kilgariff 2009: 1), was set for mid-frequency words (i.e. from rank 500 upwards), thus excluding rare words. The procedure of extracting key keywords and multi-word terms was performed to guide the analysis towards the most prominent concepts in the study corpus.

By comparing the AIC to the enTenTen20 reference corpus, Sketch Engine's keyword tool produced 100 keywords and multi-word terms, ordered by keyness, was created in Sketch Engine to be used as a starting point in the analysis of AI coverage. Arguably, the full keyword list of nouns and adjectives (which includes verb forms in the case of verbal nouns) sums up the basic vocabulary of AI coverage and highlights recurrent subtopics. The same amount of multiword terms (nominal lexical collocations in this case) were extracted as well which have been proven to be very useful to better circumscribe the semantics of keywords, whose social and cultural implications could not just be captured by automatic extraction (Paganoni 2019: 29), showing meaning relationships between them and thus preparing the ground for the analysis.

In order to find out how *The Guardian* and *MailOnline* differ in their coverage, a list of 100 keywords was extracted from the TG and MO sub-corpora as the focus corpora against the enTenTen20 reference corpus. Furthermore, the author looked at concordances for some keywords to better understand their context. Lastly, two word sketches were created with the term 'artificial intelligence' to understand which collocates were the most strongly related to the specific term.

In order to discover what topics were most prevalent in each year (2019, 2020, 2021), three additional sub-corpora were created (AI 2019, AI 2020, AI 2021) and the top 50 multiword terms were extracted from each sub-corpus. Here, the reference corpus was chosen to be the rest of the AIC (i.e., AI 2019 against AI 2020 and AI 2021) because the enTenTen20 does not include the year 2021 and one of the limitations of the enTenTen20 corpus is that it contains minimal metadata, i.e., date of production (Jakubíček, n.d.: 2). The keywords were then categorised into five thematic groups: technology, healthcare, implications, products and arms.

The fourth and final research question regarding public perception of AI and whether this is connected to news media coverage is presented as a discussion using previous studies and a figure created on the basis of the table of terms connecting AI and public perception.

2.2 Findings and Discussion

The full keywords and multi-word terms list of 100 lemmas for the AIC in general, TG and MO, as well as the AI 2019, AI 2020 and AI 2021 sub-corpora can be found in Appendix 2. Table 2 includes the first 30 multi-word terms for AIC with frequencies and keyness scores, as well as, keywords for *The Guardian* and *MailOnline*.

Table 2. 30 keywords for AIC compared with enTenTen20, with parameter N=500, using Sketch Engine.

No.	*AI multi-word terms*	*Frequency*	*Score*	*The Guardian*	*MailOnline*
1	artificial intelligence	1057	5,294	human	**Facebook**
2	social media	150	1,522	machine	human
3	machine learning	128	1,516	**Google**	researcher
4	facial recognition	98	1,401	tech	image
5	neural network	85	1,342	**UK**	Google
6	ai system	78	1,321	company	machine
7	deep learning	68	1,277	**bias**	**patient**
8	using artificial intelligence	67	1,276	researcher	predict
9	chief executive	57	1,234	digital	use
10	new technology	54	1,222	cancer	scientist
11	use of ai	58	1,221	scientist	video
12	science fiction	56	1,205	tool	detect
13	breast cancer	47	1,193	science	app
14	tech company	43	1,159	develop	brain
15	fake video	42	1,151	**deepfake**	firm
16	human brain	36	1,145	Facebook	science
17	killer robot	34	1,14	expert	user
18	tech giant	34	1,136	use	**drone**
19	facial expression	32	1,132	research	train
20	tech firm	32	1,13	future	camera
21	recognition technology	32	1,128	train	develop
22	mental health	31	1,127	learn	company
23	computer program	31	1,127	potential	identify
24	clinical trial	31	1,127	app	**cancer**
25	dark matter	42	1,122	firm	expert
26	computer science	31	1,122	faculty	scan
27	prime minister	35	1,122	predict	UK
28	learning algorithm	30	1,119	**US**	tech
29	self-driving car	32	1,114	drug	create
30	home office	36	1,112	**automation**	able

With *artificial intelligence* at the top as expected, the selection of content words can be read as a snapshot of the technologies, actors and crucial issues in AI. In between are a number of terms related to the technological domain and references to Silicon Valley *chief executives* (57 hits) and *tech giants* (34 hits), and other *tech companies* (43 hits) and *firms* (32 hits) occupying the middle ground. The actors with a lower keyness score are government-related (*prime minister* is identified 35 times and the *home office* 36 times). Interestingly, the first academia-related actor *computer scientist* comes in at rank 37 with 24 occurrences in the whole corpus. That means that *chief executives* from the industry sector (such as Deepmind's, Benevolent AI's and Youtube's) are two times more prevalent in AI news than academics. This could be tied to the fact that news introducing novel products or (advances in) industry initiatives are far more frequent than those reporting on academic studies or reports. This corresponds with the findings of Brennen et al (2018). On the other hand, they found that government sources were behind academia which is not the case in this study. Concordances reveal that *prime minister* is mostly modified by *former* (David Cameron) and his dealings with becoming chair of a US firm; as well as *home office* having collocates of *according to* (mostly in MO) - using the Home Office as a credible source -, and *racial bias* and *legal challenge* (mostly in TG) regarding a visa application system. Thus, even though mentions have grown since 2018, the context is mostly neutral or negative and not reflective of UK government's pursuit of AI supremacy.

Moreover, the results highlight the continuous digital duopoly of Facebook (443 occurrences) and Google (463), against the 107 occurrences of Microsoft. In actual fact, the topicality of the three platforms seems to correspond to their respective positioning in the digital market. According to the 2021 report released by the Reuters Institute for the Study of

Journalism, Google and Facebook are the Big Two of online advertising (Newman et al 2021: 13). The most referenced individual continues to be Elon Musk (Brennen et al 2018) with 110 hits; of these more than 75% are in MO which could also be related to the higher number of articles in the MO sub-corpus.

Undoubtedly, the AI topic is reported first and foremost as a technological system (*machine learning, neural network, AI system, deep learning, computer program, computer science, learning algorithm*), followed by technological innovation (*facial recognition, new technology, self-driving car*) and societal change. Notably, the preferred meaning of *science* appears to be *science fiction* (frequency 43) and not *computer science* (frequency 32). Moreover, the noun *video* is modified by *fake*, and *robot* by *killer*. Killer robots refer to 'Lethal Autonomous Weapon Systems' (LAWS) which are essentially 'combat robots'. Killer robots are mostly a thing of the future still but there already numerous movements trying to stop the development of these systems (WILPF 2020: 2). Notably, the higher appearance of the term in the MO sub-corpus could indicate that MO might still occasionally lean towards being sensationalist, even though it is a middle-market newspaper. The high frequency of keyword *use* (both as a verb and a noun: *using artificial intelligence*, *use of AI*) is a significant clue that alludes to the human intentionalities on which AI ultimately relies, despite more radical claims that see it as being independent of human reason.

Finally, words related to the healthcare sector (*breast cancer, human brain, mental health, clinical trial*) are present at the top of the list. Analytics firm GlobalData (2021) found that AI is expected to have the greatest impact on the sector. An example of this is a 2019 *MailOnline* (9 Aug 2019) news article which explains that AI "…could diagnose breast cancer better than doctors after being trained to read MRI scans.". One strong theme is indeed AI

systems that can diagnose *patients*, or *detect* diseases. The use of AI in healthcare has been sped up due to the COVID-19 outbreak in 2020. In the same year, 90% of large pharmaceutical firms initiated AI projects (Waltham 2021). Britain's two biggest biopharmaceutical companies, AstraZeneca and GSK, entered into a five-year partnership with the University of Cambridge to fund the Cambridge Centre for AI in Medicine (CCAIM) with the aim to *develop* AI technologies which will advance *clinical trials*, personalised medicine and drug discovery (O'Neill 2020). An example, this time from *The Guardian* (31 Mar 2020), demonstrates how AI is being used to screen 15,000 drugs in search for a coronavirus cure.

When comparing the two sub-corpora (Appendix 2) in order to answer the research question "How do *MailOnline* and *The Guardian* vary in their coverage of AI?", it appeared that some keywords (e.g., *ai*, *intelligence*, *technology*, *computer*, *robot*) are common to both news media and were thus removed from comparison (Table 2).

In regard to stakeholders, *MailOnline* mentioned Facebook far more often than Google, whereas in *The Guardian* it was the other way around. This could allude to the fact of *MailOnline* as a middle-market tabloid and that Facebook The nations most frequently mentioned in TG are the UK, followed by the US; similarly in MO, the UK is mentioned the most, but is followed by China. This is explained by the fact that these newspapers are UK-based in scope and thus report on the happenings in this country. American newspapers, on the other hand, mostly mentioned China, America and Japan (Sun et al 2020), which is understandable considering that these were (and still are) the leaders in AI during the time period chosen for the study (1977-2018). Looking at the concordances of the countries, *MailOnline*'s references to China are mostly in relation to the AI arms race, including its constant developments in the military industry, e.g.,

China is developing underwater AI robots that can hide in the sea and attack enemy vessels with torpedoes WITHOUT human guidance. (8 July 2021)

Another context is China using the technology to 'spy' on their people. *The Guardian* mentions the US mostly in connection with its tech companies.

Furthermore, when discussing the adjectivity of *humans*, *The Guardian* mostly pairs it with *human right, human being,* and *human player* (i.e. poker, videogames). MO has a higher tendency to use *human* in context with the *human brain*, *human pilot* (in regard to autonomous fighter *drones*). MO's tendency to highlight the militaristic uses of AI is further illustrated by the keyword *machine* modifying the noun *gun*, in addition to *learning* (the latter is the only noun modified by *machine* in the TG sub-corpus's first 100 multiword terms, and the former only appears as a noun compound once in entirety of the TG sub-corpus).

In 1.2., it was revealed that left-leaning outlets usually focus on issues of ethics of AI (discrimination, data bias, privacy). This is supported by the findings of this study in that *The Guardian,* a more liberal outlet, has a high keyness score for keywords such as *bias, human right, gender* and *diversity* (especially the *lack of diversity*). Moreover, the articles highlight ethical concerns surrounding topics such as *deepfakes* and *autonomous weapons*. It can be said then that The Guardian places greater emphasis on the various limits and dangers of AI, not just ethical concerns. Even the use of the word *potential* in *The Guardian* is related not to potential benefits but to *potential harm* instead. *MailOnline,* a more conservative outlet, mainly discusses two industry-related topics: the arms industry and healthcare. The different newspaper types play a role here, as *The Guardian,* a quality paper goes more in-depth on the topic of AI, whereas *MailOnline*, a middle-market tabloid focuses coverage of significant events in different industries, as well as foreign news (China-US arms race). Brennen (et al 2018) found that

although the discussion of automation and job loss is common across coverage, left-leaning outlets frequently discuss jobs lost through automation, whereas right-leaning outlets claim AI will lead to the creation of new jobs. In the time range 2019-2021, it is apparent that these stances/ have prevailed with *The Guardian* discussing jobs in the context of verbs such as *take*, *lose* and *displace* more than *generate* and MailOnline in the context of *create* more than *lose* or *replace*.

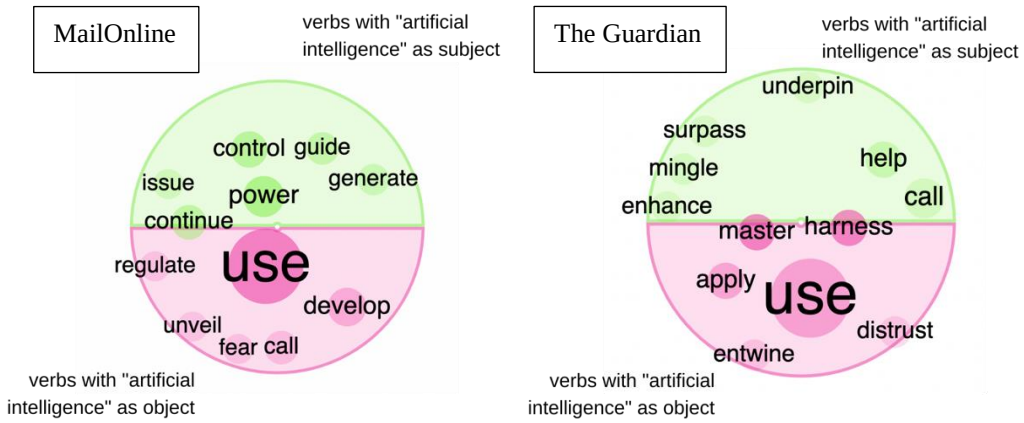

Figure 3. Word sketch of 'artificial intelligence' in the *MailOnline* and *Guardian* sub-corpora.

When examining the term 'artificial intelligence's' grammatical and collocational behaviour (Fig. 1), it should be noted that the typicality score (LogDice) indicates how strong the collocation is. The higher the score, the stronger the collocation, meaning the words in the collocation do not frequently combine with many other words (Kilgariff et al 2014). The word sketch identified that the following verbs with 'AI' as the object have the strongest collocations in the *MailOnline* sub-corpus: many still *fear* (score = 8.1) and are thus calling for a need to *regulate* (8.1) AI. Similarly in *The* Guardian, the verb *distrust* is connected to AI, as well as how different sectors should *harness* (9.8) AI to improve outcomes and that other countries/companies are already *mastering* (9.8) its technologies ahead of the UK. Verbs with 'AI' as a subject in MO include: *powered* (11.3), *controlled* (10.5) and *guided* (9.6) by AI,

demonstrating AI's prevalence and potential uses; that it *continues* (10.0) to advance and be a priority, although some still *issue* (8.6) warnings. AI has the ability to *generate* (8.9) (photos, cancer treatment plans) and *create* (8.4) (music, images). In TG: AI can *help* (9.4), e.g., spot fake videos. Similarly to MO, verbs that illustrate how AI is already among us were also highly scored: AI and humans *mingle* (9.1), and that AI *underpins* (9.0) many aspects of everyday life. There are also discussions whether AI will truly *enhance* (9.0) rather than replace us.

Brennen (et al 2018) found that right-leaning outlets emphasise that there should be no regulating AI as this would limit companies and lead to their slower development compared to, for example, China. Coverage regarding AI regulation has thus changed in conservative outlets from against regulation:

> *the government must resist the temptation to slow down its* [AI's] *advance through inappropriate red tape and burdensome taxation* (Telegraph, 23 Jan. 2018)

to in favour of it:

> *We have to prevent the avoidable tragedy that is coming if we do not regulate our killer robots* (MailOnline, 17 Jan. 2020).

Following along the topic of change in yearly reporting, the years 2019-2021 are looked at. Presented in Table 3 are the five topics retained for final interpretation. Additionally, five representative keyword examples are presented depicting the meaning of each topic.

Table 3. Keyword topics in AI 2019, AI 2020 and AI 2021 sub-corpora.

Topics	**technology**	**healthcare**	**implications**	**product**	**arms**
Keyword examples	facial recognition, virtual reality, computer science, intelligence algorithm, big tech	breast cancer, MRI scan, protein structure, clinical trial, mental health	AI ethic, new job, porn scam, trans people, black people, human right	pie recipe, electronic toothbrush, sex doll, pressure sensor, chat service	autonomous weapon, human pilot, air force, drone swarm, robot submarine
Keywords in AI 2019, AI 2020 and AI 2021	2019 (6)	2019 (6)	2019 (14)	2019 (9)	2019 (3)
	2020 (2) ↓	2020 (22) ↑	2020 (4) ↓	2020 (8)	2020 (4)
	2021 (6) ↑	2021 (12)	2021 (6)	2021 (19) ↑	2021 (2)

The topic 'product' has two sub-categories: AI goods and programs (e.g., pressure sensor), and results/achievements (Henry VIII in the context of "Scientists use artificial intelligence to work out how much of Shakespeare's collaborative history play he actually wrote" (*MailOnline*, 25 Jan 2019)). Throughout 2019-2021, keywords representative of the topic the 'arms' industry remained somewhat similar in frequency/relevance with the number rising by one keyword in 2020 and dropping by two the following year.

The topic by which 2019 differs from 2020 and 2021 the most is implications with 14 representative keywords out of 50. Implications keywords include *ai ethic*, *new job*, *porn scam*, *black patient*, *public service* and *vulnerable people*, thus showcasing that AI ethics was a major discussion point in 2019. Of the entities mentioned, 80% are from the US and 20% are related to the arms industry. Notably, none are from the UK.

Healthcare (22 keywords) and the arms industry (4 keywords) are the two keyword topics that appeared most frequently in the 2020 news articles. The increase in keywords related to healthcare was to be expected, again, because of Covid-19. However, advances were not made just in regard to the pandemic, but also cancer detection (*breast cancer*, *cancer screening*, *cancer patient*) and the *protein folding problem*. In regard to the entities mentioned, about 60%

are from the UK and only one is distinctly American. Thus, a complete shift occurred as the happenings in the UK pulled almost complete focus.

In the year 2021, healthcare is not as relevant as in 2020 but is still being reported on in regard to, for example, *mental health*, *brain scans*, and *brain wave*. The highest number of keywords related to products are also from that year (19 keywords). Furthermore, while in previous years, keywords illustrating AI goods and programs appear the most (66.7% in 2019, 87.5% in 2020), in 2021, results/achievements by AI are higher with 58%. Some examples of these include: AI decodes *sperm whale* clicking sounds, AI able to reveal climate change *tipping point*, and AI ends debate whether *Jaffa Cake* is a cake or biscuit.

The Select Committee on Artificial Intelligence (SCAI) stated in 2018 that people have just started distancing AI from the creations of Hollywood. This is supported by this study in that the term *science fiction* appears predominantly in 2019 in reference to it potentially being a thing of the past: "AI has evolved rapidly from the realm of science fiction to something that permeates our everyday lives." (*The Guardian* 2019). Lastly, this study will analyse where else public perception (Fig. 3) and news media presentation of AI coincide and where they do not.

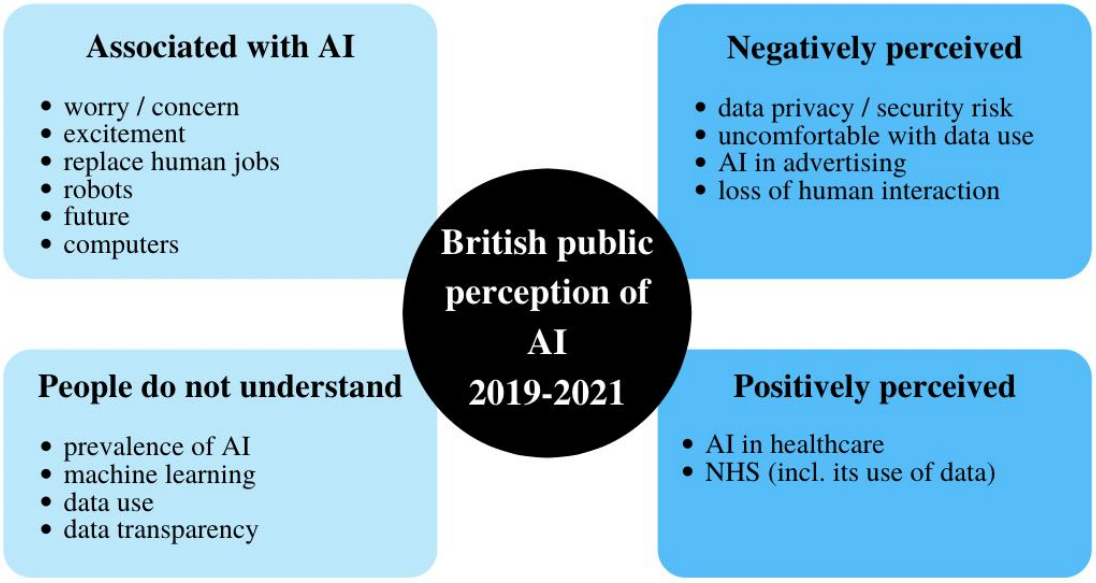

Figure 3. British public perception of AI during 2019-2021 according to research surveys (Gallup 2019; Kantar Public 2019; PAS 2020; Nader et al 2022; CDEI 2022)

As explored in 1.3 people 'worry' or have 'concerns' regarding 'scary' AI. The noun *concern* even appears in the AIC, though at the bottom of the keyword list with a keyness ranking of 98 and with 206 occurrences. Specifically, as mentioned above, MO includes the word *fear* and TG the word *distrust* regarding AI. However, even though 'concern' regarding AI has mostly prevailed throughout the years, the British public have started to feel *excitement* as of late regarding the potential benefits of AI, which cannot be related to any AIC top 100 keywords or multi-words. Another quality associated with AI is *replacing human jobs.* Although previous studies (Brennen et al 2018) indicate that job loss is connected to automation, in this corpus 'job loss' is most strongly connected to *robots* (verbs with 'robot' as subject: *replace*, *take over*, both appear four times). Similarly, nouns modified by 'robot' include nouns related to war or force: *takeover*, *revolution*, and *overlord*. The surveys revealed that people indeed most associate AI with *robots* as well as *computers*, which are two AI technologies with high keyness scores in the AIC (7th with 447 occurrences, 12th with 472 occurrences,

respectively). Collocates surrounding *robots* in the AIC reveal that in the category 'robot' and/or…, *artificial intelligence* has the highest frequency with five hits. This reveals a direct correlation between public association and news media coverage of certain topics. Additionally, people perceive AI as something which will affect the *future* (48th with 321 occurrences) rather than the present; but that the future is coming sooner rather than later. In AIC as well, *future* is used as a noun 197 times and is modified by adjectives implicating nearness (*near* (4 times), *not-too-distant* (once)), instead of distance (*dystopic* and *distant* (both once)).

'Machine learning', on the other hand, is a technology which most people do not understand or have not even heard of. In the AIC, *machine learning* is the top third multi-word term, ahead of *robots* and *computers*, showcasing its high relevancy to AI news articles. However, it has only 128 hits in the AIC, thus it appears about 70% or four times less than *robot* or *computer*, again demonstrating that what the news media portrays more frequently, the public understands or has heard of. There is also the probability that people remember topics that are more specific, like *robots* and *computers*, and emotion-evoking (*killer robots*) than more general and hard to understand. 'Data use' and 'transparency' are two other topics which people are confused by; in the top 100 multi-word terms list *data* only modifies *science* (22 occurrences). When investigating concrete collocations of *data* in the AIC, it is revealed that *data privacy* and *data protection* appear once throughout the corpus, and of the verbs with 'data' as object: data is *collected* (3 times), *used to* (2 times) and *gathered* (1 time) which are verbs that with more frequent use, could have alluded to *data* collection and its use. Data privacy is one aspect that people perceive as negative and the previously mentioned noun-form of *concern* is indeed most frequently modified by *privacy* with 9 occurrences.

Although 'the need for regulation' has apparently dropped in 2019-2021 public's association with AI, maintaining the 'human' (loss of human interaction, loss of human skills, human experts involvement required) aspect of AI has remained; this is also present in the news media with emphasis put on *human expert* (AI 2019) and *human being*. Moreover, the public is favourably inclined towards the use of AI in healthcare and by the NHS. As seen from the findings of this study, the healthcare industry is indeed covered extensively and the NHS appears 64th with 121 mentions, mostly in relation to *huge potential* and *exploring* how AI can help.

Interestingly, *autonomous weapons, risks of bias* and *cyber-attacks* are three topics that purportedly only Americans associate with AI. Only *cyber-attacks* does not appear in the AIC top 100 but the other two do (*autonomous weapons* with 24 occurrences; *racial bias* with 17 occurrences). There is no apparent reasoning for this other than the fact that the frequencies of these terms are simply lower than other terminology which is more prevalent. Another concept that only came out of American surveys is people's understanding of *smart home devices* as related to AI. In the AIC, there is no mention of any smart home device (such as virtual assistant or robot vacuum cleaner), implicating that the news media usually portrays AI in the context of harder-to-grasp technology, which leaves people confused as to the prevalence of AI and its uses in daily life.

CONCLUSION

AI technologies are making remarkable progress and are having real-world impact on people, institutions and culture. Although the United States and China continue to fight over the AI frontier, one of the leading European countries in AI readiness is, in fact, the United Kingdom. Currently, the UK is already recognised for its technology for healthcare.

Previous research about AI in the news media shows that news is mostly technology-related, even in specific industries, such as healthcare. Differences occur in newspapers with different political views, with left-leaning outlets focusing on issues of ethics of AI, and right-leaning on issues of economics and geopolitics. Overall, discourse remains pragmatic and even optimistic.

On the other hand, people show a general understanding of what AI is, but no real comprehension of how AI is used or how it works. The present study analysed *The Guardian* and *MailOnline* from January 2019 to December 2021 to find out how AI is presented in the British news media, in addition to whether the representation corresponds with public perception of the topic. Moreover, each year in the study was examined to find how they differed from one another, i.e., what topics were relevant to them the most and also how did the different online newspapers vary in their coverage of AI. Articles were collected from the LexisNexis database with the combined study corpora having a word count of 417,523. The analysis was conducted with the Sketch Engine corpus analysis tool.

The findings of the specific research questions presented in the Introduction are as follows: firstly, the keywords most relevant in British news media about AI are precise technical terms representing systems what/which the general public mostly does not understand. Overall, industry-related entities remain the most cited but there has been a growth in the mentioning of

the UK government during 2019-2021. The context, however, is not in trajectory with their/its own development plans but rather related to facts released by them or, more frequently, of their controversies. Healthcare is the most prominent topic during the analysed time period, partially in relation to Covid-19 but also just general advances. Secondly, *MailOnline* and *The Guardian* differ in their coverage of AI in both format and political leaning with *The Guardian* emphasising the potential limitations and pitfalls of AI, thus going more in-depth, and *MailOnline* focusing on the coverage of significant events in the healthcare and arms industry. Yet, keywords occurring in both newspapers mostly highlighted how AI is already present in our everyday lives. Thirdly, AI keywords have changed between 2019-2021 from emphasis on AI ethics in 2019 to healthcare and the UK in 2020 to innovative achievements by AI in 2021 highlighting the ever-evolving nature of AI. Lastly, this study highlighted that public perception has a direct correlation with news media coverage of certain topics, since the examined news media did indeed more frequently cover topics which the public associated with AI, as opposed to those which the public was confused about.

REFERENCES

Acheson, Ray. 2020. A WILPF Guide to: Killer Robots. Available at https://www.wilpf.org/wp-content/uploads/2020/04/WILPF_Killer-Robots-Guide_EN-Web.pdf, accessed May 2022.

Archer, Henry, Rebecca Writer-Davies and Mark McGeoghegan. 2018. AI, Automation, and Corporate Reputation. Ipsos MORI. Available at https://www.ipsos.com/sites/default/files/ct/publication/documents/2018-11/ai_automation_cr_ipsos.pdf, accessed April 2022.

Baidu. 2021. These five AI developments will shape 2021 and beyond. *MIT Technology Review*. Available at https://www.technologyreview.com/2021/01/14/1016122/these-five-ai-developments-will-shape-2021-and-beyond/, accessed April 2022.

Baker, Paul. 2006. Using Corpora in Discourse Analysis. London and New York: Continuum.

Bar Standards Board. 2019. Risk Outlook. Available at https://www.barstandardsboard.org.uk/uploads/assets/6bb68426-03bf-4e97-86c6cb4780c76743/bsbriskoutlook2019.pdf, accessed April 2020.

BBC. n.d. Newspapers. In *Media Studies*. Available at https://www.bbc.co.uk/bitesize/guides/zps4qty/revision/1, accessed May 2022.

Brennen, J. Scott, Philip N. Howard and Rasmus K. Nielsen. 2020. Balancing Product Reviews, Traffic Targets, and Industry Criticism: UK Technology Journalism in Practice. *Journalism Practice*, ahead of print, 1-18. Available at https://www.tandfonline.com/doi/full/10.1080/17512786.2020.1783567?scroll=top&needAccess=true, accessed November 2020.

Brennen, J. Scott, Philip N. Howard and Rasmus Kleis Nielsen. 2018. An Industry-Led Debate: How UK Media Cover Artificial Intelligence. Available at https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-12/Brennen_UK_Media_Coverage_of_AI_FINAL.pdf, accessed November 2020.

Britannica, The Editors of Encyclopaedia. 2019. Daily Mail. *Encyclopedia Britannica*. Available at https://www.britannica.com/topic/Daily-Mail, accessed May 2022.

Britannica, The Editors of Encyclopaedia. 2021. (2021. The Guardian. *Encyclopedia Britannica*. Available at https://www.britannica.com/topic/The-Guardian-British-newspaper, accessed May 2022.

Bunz, Mercedes and Marco Braghieri. 2022. The AI doctor will see you now: assessing the framing of AI in news coverage. *AI & Society*, 37: 9-22. Available at https://link.springer.com/article/10.1007/s00146-021-01145-9#Sec3, accessed April 2022.

Burscher, Bjorn, Rens Vliegenthart, and Claes H. de Vreese. 2016. Frames Beyond Words: Applying Cluster and Sentiment Analysis to News Coverage of the Nuclear Power Issue. *Social Science Computer Review*, 34: 5, 530–45. https://doi.org/10.1177/0894439315596385, accessed April 2022.

Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, Hyrum Anderson, Heather Roff, Gregory C. Allen, Jacob Steinhardt, Carrick Flynn, Seán Ó hÉigeartaigh, Simon Beard, Haydn Belfield, Sebastian Farquhar, Clare Lyle, Rebecca Crootof, Owain Evans, Michael Page, Joanna Bryson, Roman Yampolskiy and Dario Amodei. 2018. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.

Available at https://www.eff.org/files/2018/02/20/malicious_ai_report_final.pdf, accessed April 2020.

Bruner, Jon. 2016. Why 2016 is shaping up to be the Year of the Bot. *O'Reilly*, June 15. Available at https://www.oreilly.com/radar/why-2016-is-shaping-up-to-be-the-year-of-the-bot/, accessed April 2020.

Insider. 2020. Artificial Intelligence News. Available at https://www.businessinsider.com/artificial-intelligence#featured, accessed April 2022.

Central Digital and Data Office and Office for Artificial Intelligence. 2019. A guide to using artificial intelligence in the public sector. Available at https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector, accessed April 2020.

Centre for Data Ethics and Innovation (CDEI). 2020. AI Barometer 2020. Available at https://www.gov.uk/government/publications/cdei-ai-barometer, accessed April 2022.

Centre for Data Ethics and Innovation (CDEI). 2021. AI Barometer 2021. Available at https://www.gov.uk/government/publications/ai-barometer-2021, accessed April 2022.

Centre for Data Ethics and Innovation (CDEI). 2022. Public Attitudes to Data and AI Tracker Survey. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1064525/Public_attitudes_to_data_and_AI_-_Tracker_survey.pdf, accessed May 2022.

Chandler, Daniel and Rod Munday. 2011. *A Dictionary of Media and Communication* (1 ed.). Oxford: Oxford University Press.

Chuan, Ching-Hua, Wan-Hsiu Sunny Tsai and Su Yeon Cho. 2019. Framing Artificial Intelligence in American Newspapers. AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery: New York, 339-344. Available at https://dl.acm.org/doi/abs/10.1145/3306618.3314285?casa_token=216FhO2oGGkAAAAA:yzmO52p1wpjzqyCpz8DMKeKp4FHeguYcp5EgVMktPwhXoFgiuJH6aQEcyjPpFT4yANs4iTirwlrn, accessed November 2020.

Committee on Standards in Public Life. 2020. Artificial Intelligence and Public Standards: A review by the Committee on Standards in Public Life. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF, accessed April 2020.

Committee on Standards in Public Life. 2020. Artificial Intelligence and Public Standards: public polling. Available at https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-public-polling, accessed April 2020.

Crawford, Kate, Meredith Whittaker, Madeline Clare Elish, Solon Barocas, Aaron Plasek and Kadija Ferryman. 2016. The AI Now Report. Available at http://acikistihbarat.com/Dosyalar/AINowSummaryReport-artificial-intelligence-effects-in-near-future.pdf, accessed April 2022.

Dafoe, Allan. 2018. AI Governance: A Research Agenda. *Centre for the Governance of AI, Future of Humanity Institute, University of Oxford*.

Deloitte. 2017. AI and you: Perceptions of Artificial Intelligence from the EMEA financial services industry. Available at

https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/technology/deloitte-cn-tech-ai-and-you-en-170801.pdf, accessed April 2020.

Deng, Li. 2018. Artificial Intelligence in the Rising Wave of Deep Learning: The Historical Path and Future Outlook [Perspectives]. *IEEE Signal Processing Magazine*, 35: 1, 180-177.

Dobrescu, Emilian M., Edith M. Dobrescu. 2017. The Future of the Artificial Intelligence in Economics and Management. *Review of General Management*, 26: 2, 81-89.

Duberry, Jerome and Sabrya Hamidi. 2021. Contrasted media frames of AI during the COVID-19 pandemic: a content analysis of US and European newspapers. *Online Information Review*, 45: 4. Available at https://www-emerald-com.ezproxy.utlib.ut.ee/insight/content/doi/10.1108/OIR-09-2020-0393/full/html, accessed April 2022.

Dudo, Anthony, Sharon Dunwoody and Dietram A. Scheufele. 2011. The Emergence of Nano News: Tracking Thematic Trends and Changes in U.S. Newspaper Coverage Of Nanotechnology. *Journalism & Mass Communication Quarterly*, 88: 1, 55-75. Available at https://journals.sagepub.com/doi/10.1177/107769901108800104, accessed April 2022.

Eurobarometer. 2020. Attitudes towards the Impact of Digitalisation on Daily Lives. Factsheets: United Kingdom. Available at https://europa.eu/eurobarometer/surveys/detail/2228, accessed May 2022.

Everitt, Tom, Gary Lea and Marcus Hutter. 2018. AGI Safety Literature Review. In *International Joint Conference on Artificial Intelligence (IJCAI)*. arXiv: 1805.01109.

Fast, Ethan and Eric Horvitz. 2016. Long-Term Trends in the Public Perception of Artificial Intelligence. ArXiv:1609.04904. Available at https://arxiv.org/abs/1609.04904, accessed November 2020.

Firth, John Rupert. 1957. *Papers in linguistics, 1934-1951*. London: Oxford University Press,

Gallup and Northeastern University. 2019. Facing the Future. U.S., and U.K. and Candian citizens call for a unified skills strategy for the AI age. Available at https://www.northeastern.edu/gallup/pdf/Northeastern_Gallup_AI_2019.pdf, accessed January 2021.

Gerner, Marina. 2020. Is the UK still an AI leader? *Racounteur*. December 10. Available at https://www.raconteur.net/global-business/uk/uk-ai-leader/, accessed April 2020.

GlobalData. 2021. Artificial intelligence will be most disruptive technology across pharmaceutical industry in 2021 and beyond. Available at https://www.globaldata.com/artificial-intelligence-will-disruptive-technology-across-pharmaceutical-industry-2021-beyond/, accessed May 2022.

Government Office for Science. 2021. Large-scale computing: the case for greater UK coordination. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1018875/UK_Computing_report_-_Final_20.09.21.pdf, accessed April 2022.

Gov.uk. 2017. The Grand Challenges. [Policy Paper]. (updated 26 January 2021). Available at https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/industrial-strategy-the-grand-challenges, accessed April 2021.

Gov.uk. 2018. AI Sector Deal. [Policy Paper]. (updated 21 May 2019). Available at https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal, accessed April 2021.

Gov.uk. 2019. A guide to using artificial intelligence in the public sector. [Guidance]. Available at https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector#defining-artificial-intelligence, accessed February 2022.

Gov.uk. 2021a. New strategy to unleash the transformational power of Artificial Intelligence. [News story]. Available at https://www.gov.uk/government/news/new-strategy-to-unleash-the-transformational-power-of-artificial-intelligence, accessed April 2022.

Gov.uk. 2021b. The potential impact of AI on UK employment and the demand for skills. [Research and analysis]. Available at https://www.gov.uk/government/publications/the-potential-impact-of-ai-on-uk-employment-and-the-demand-for-skills, accessed April 2022.

Holder, Chris, Vikram Khurana and Mark Watts. 2018. Artificial Intelligence: Public Perception, Attitude and Trust. *Bristows*. Available at https://www.bristows.com/app/uploads/2019/06/Artificial-Intelligence-Public-Perception-Attitude-and-Trust.pdf, accessed January 2021.

House of Lords Select Committee on Artificial Intelligence (SCAI). 2018. AI in the UK: ready, willing and able? Available at https://allcatsrgrey.org.uk/wp/download/informatics/100.pdf, accessed November 2020.

Hutson, Matthew. 2017. AI Glossary: Artificial intelligence, in so many words. *Science*, 357: 6346, 19. Available at https://science.sciencemag.org/content/357/6346/19, accessed April 2021.

Informatics Europe & EUACM. 2018. When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making. https://www.acm.org/binaries/content/assets/public-policy/ie-euacm-adm-report-2018.pdf, accessed April 2020.

IBM. n.d. Deep Blue. *Icons of Progress*. Available at https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/, accessed March 2022.

IBM Cloud Education. 2020. Artificial Intelligence (AI). Available at https://www.ibm.com/cloud/learn/what-is-artificial-intelligence, accessed March 2022.

Ipsos MORI. 2018. Public Views of Machine Learning: Findings from Public Research and Engagement Conducted on Behalf of the Royal Society. *The Royal Society*. Available at https://perma.cc/79FE-TEHH, accessed January 2021.

Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý and Vít Suchomel. n.d. The TenTen Corpus Family. Available at https://www.sketchengine.eu/wp-content/uploads/The_TenTen_Corpus_2013.pdf, accessed May 2022.

Kantar Public. 2019. AI poll: summary of findings. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/802548/BEIS_AI_PR_Survey_40309009_Topline_summary_V1__1_.pdf, accessed May 2022.

Kilgariff, Adam. 2009. Simple maths for keywords. In *Proceedings of Corpus Linguistics Conference CL2009*, Mahlberg, M., González-Díaz, V. & Smith, C. (eds.), University of Liverpool. Available at https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf, accessed May 2022.

Kilgariff, Adam, Vit Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchome. 2014. The Sketch Engine: ten years on. *Lexicography ASIALEX* 1, 7-36.

LexisNexis. n.d. SmartIndexing Technology. Available at https://www.lexisnexis.com/bis-user-information/docs/LexisNexisSmartIndexingTechnology.pdf, accessed October 2021.

Littman, Michael, Ifeoma Ajunwa, Guy Berger, Craig Boutilier, Morgan Currie, Finale Doshi-Velez, Gillian Hadfield, Michael C. Horowitz, Charles Isbell, Hiroaki Kitano, Karen Levy, Terah Lyons, Melanie Mitchell, Julie Shah, Steven Sloman, Shannon Vallor and Toby Walsh. 2021. Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. Stanford: Stanford University. Available at http://ai100.stanford.edu/2021-report, accessed April 2022.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester and Claude E. Shannon. 2006. A Proposal for the Dartmouth Project on Artificial Intelligence [August 31, 1955]. *AI Magazine*, 27: 4, 12-14.

McKinsey Global Institute. 2019. Artificial intelligence in the United Kingdom: Prospects and challenges. Available at https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Artificial%20intelligence%20in%20the%20United%20Kingdom%20Prospec

ts%20and%20challenges/Artificial-intelligence-in-the-United-Kingdom-VF2.ashx, accessed November 2020.

Nader, Karim, Paul Toprac, Suzanne Scott and Samuel Baker. 2022. Public understanding of artificial intelligence through entertainment media. *AI & SOCIETY*. Available at https://link.springer.com/article/10.1007/s00146-022-01427-w#citeas, accessed May 2022.

Nadimpalli, Meenakshi. 2017. Artificial Intelligence Risks and Benefits. *International Journal of Innovative Research in Science, Engineering and Technology*, 6: 6.

Naughton, John. 2018. Written evidence (AIC0144). Available at http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69649.html, accessed November 2020.

Newman, Nic, Richard Fletcher, Anne Schulz, Simge Andı, Craig T. Robertson and Rasmus Kleis Nielsen. 2021. Reuters Institute Digital News Report 2021. 10th ed. *Reuters Institute for the Study of Journalism*. Available at https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf, accessed May 2022.

Newman, Nic. 2022. Journalism, Media, and Technology Trends and Predictions 2022. *Reuters Institute for the Study of Journalism*. Available at https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-01/Newman%20-%20Trends%20and%20Predictions%202022%20FINAL.pdf, accessed May 2022.

Nilsson, Nils John. 2010. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge: Cambridge University Press.

Ofcom. 2021. News Consumption in the UK: 2021. Available at https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/news-media/news-consumption, accessed March 2022.

Office for Artificial Intelligence. 2021. National AI Strategy. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf, accessed April 2022.

O'Neill, Sean. 2020. Cambridge Centre for AI in Medicine announces its official launch. *CCAIM*. Available at https://ccaim.cam.ac.uk/cambridge-centre-for-ai-in-medicine-announces-its-official-launch/, accessed May 2022.

Ouchchy, Leila, Allen Coin and Veljko Dubljević. 2020. AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. *AI & Society*, 35, 927-946. Available at https://link.springer.com/article/10.1007/s00146-020-00965-5#Abs1, accessed November 2020.

Paganoni, Maria Cristina. 2019. Big Data in the News Media. In *Framing Big Data*. Palgrave Pivot, Cham, 21-58.

Public Attitudes to Science 2019 (PAS). 2020. UK public attitudes to science. *BEIS Research Paper Number* 2020/012.

Perrault Raymond, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz, Terah Lyons, James Manyika, Saurabh Mishra and Juan Carlos Niebles. 2019. The AI Index 2019 Annual Report, AI Index Steering Committee, Human-Centered AI Institute, Stanford: Stanford University. Available at https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf, accessed April 2022.

Scott, Mike. 1997. PC Analysis of Key Words - And Key Key Words. *System*, 25: 2, 233–245. Available at doi:10.1016/s0346-251x(97)00011-0, accessed May 2022.

Siapera, Eugenia. 2018. *Understanding new media*. London: SAGE.

Sketch Engine. n.d.-a. Keywords and term extraction. Available at https://www.sketchengine.eu/guide/keywords-and-term-extraction/, accessed May 2022.

Sketch Engine. n.d.-b. enTenTen: Corpus of the English Web. Available at https://www.sketchengine.eu/ententen-english-corpus/#toggle-id-2, accessed May 2022.

Smith, Aaron and Monica Anderson. 2017. *Pew Research Center*. American's attitudes toward a future in which robots and computers can do many human jobs.

Spiro, Hannah and Holly Clarke. 2022. Monitoring how public attitudes towards data and AI vary over time. Available at https://cdei.blog.gov.uk/2022/03/30/monitoring-how-public-attitudes-towards-data-and-ai-vary-over-time/, accessed April 2022.

Stone, Peter, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee (Anno) Saxenian, Julie Shah, Milind Tambe and Astro Teller. 2016. Artificial Intelligence and Life in 2030. Stanford: Stanford University. Available at https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/ai100report10032016fnl_singles.pdf, accessed April 2022.

Suchomel, Vit and Jan Pomikálek. 2020. Efficient Web Crawling for Large Text Corpora. Available at https://www.sketchengine.eu/wp-content/uploads/Efficient_Web_2012.pdf, accessed May 2022.

Sun, Shaojing, Yujia Zhai, Bin Shen, and Yibei Chen. 2020. Newspaper coverage of artificial intelligence: A perspective of emerging technologies. *Telematics and Informatics*, 53, Available at https://doi.org/10.1016/j.tele.2020.101433, accessed April 2022.

Waltham, Mass. 2021. Ninety Percent of Large Pharma Companies Initiated Artificial Intelligence/Machine Learning Projects In 2020. *Trinity Life Sciences*. Available at https://www.businesswire.com/news/home/20210119005100/en/Ninety-Percent-of-Large-Pharma-Companies-Initiated-Artificial-IntelligenceMachine-Learning-Projects-In-2020, accessed May 2022.

World Economic Forum. 2020. The Future of Jobs. Report. Available at http://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf, accessed January 2021.

Zhang, Baobao and Allan Dafoe. 2019. Artificial Intelligence: American Attitudes and Trends. Oxford: Center for the Governance of AI, Future of Humanity Institute, University of Oxford. Available at https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/executive-summary.html, accessed December 2020.

Zhang, Daniel, Nestor Maslej, Erik Brynjolfsson, John Etchemendy, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Michael Sellitto, Ellie Sakhaee, Yoav Shoham, Jack Clark and Raymond Perrault. 2022. The AI Index 2022 Annual Report. *AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford*

University. Available at https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf, accessed February 2022.

APPENDICES

Appendix 1. Public perception of AI according to research; results which were only present in American surveys are marked in blue, and results which had a high number of results between 2019-2021 are marked in green.

	2018			2019			2020	2021	**No. of UK related**	**No. 19-21**
	Holder et al 2018	Ipsos MORI 2018	Zhang & Dafoe 2019	Gallup 2019	Kantar Public 2019	BEIS 2020	Nader et al 2022	CDEI 2022		
Associated with AI										
worry / concern	x	x	x	x		x	x	x	**5**	**3**
replace human jobs	x?	x		x		x	x		**4**	**2**
robots		x	x			x	x	x	**3**	**2**
tougher regulation needed	x	x	x			x			**3**	**1**
future	x		x			x		x	**3**	**2**
excitement					x	x	x	x	**3**	**3**
able to learn (new things)		x				x	x		**2**	**1**
computers						x	x	x	**2**	**2**
self-driving cars			x			x	x		**1**	**1**
facial recognition						x	x		**1**	**1**
voice / speech recognition						x	x		**1**	**1**
predict	x						x		**1**	**0**
solve problems							x	x	**1**	**1**
scary								x	**1**	**1**
science fiction						x			**1**	**1**
smart home devices			x				x		**0**	**0**
risks of bias			x				x		**0**	**0**
virtual reality							x		**0**	**0**
think (logically)							x		**0**	**0**
autonomous weapons			x						**0**	**0**
Not associated with AI										
AI and art	x	x							**2**	**0**
digital recommendation systems			x				x		**0**	**0**
wireless networks							x		**0**	**0**
big data							x		**0**	**0**
machine translation			x						**0**	**0**
search engines			x						**0**	**0**
People do not understand										
prevalence of AI	x	x	x		x	x			**4**	**2**
data use	x	x						x	**3**	**1**
machine learning		x	x	x					**2**	**1**
data transparency								x	**1**	**1**

	2018			2019			2020	2021	No. of UK related	No. 19-21
	Holder et al 2018	Ipsos MORI 2018	Zhang & Dafoe 2019	Gallup 2019	Kantar Public 2019	BEIS 2020	Nader et al 2022	CDEI 2022		
Negatively perceived										
data privacy / security risk	x	x	x			x		x	**4**	**2**
job impact	x	x	x	x					**3**	**1**
uncomfortable with data use	x					x		x	**3**	**2**
AI in retail / advertising		x?				x		x	**3**	**2**
loss of human interaction		x				x			**2**	**1**
de-skilling		x				x			**2**	**1**
(misuse) accountability								x	**1**	**1**
surveillance			x				x		**0**	**0**
cyber-attack			x				x		**0**	**0**
Positively perceived										
AI in healthcare		x			x	x	x	x	**4**	**3**
NHS						x		x	**2**	**2**

Appendix 2. Top 100 keywords and multi-word terms in the AIC, TG and MO, and top 50 in 2019-2021 sub-corpora.

No.	***Keyword***	*Frequency*	*Score*	***Terms***	*Frequency*	*Score*	*The Guardian*	*Frequency*	*MailOnline*	*Frequency*	***AI 2019***	***AI 2020***	***AI 20201***
1	AI	2437	10,379	artificial intelligence	1057	5,294	ai	1013	ai	1424	science fiction	breast cancer	social medium
2	artificial	1191	5,655	social media	150	1,522	intelligence	383	artificial	837	facial recognition	milky way	dark matter
3	intelligence	1249	5,539	machine learning	128	1,516	artificial	354	intelligence	866	brain tumour	kidney disease	jaffa cake
4	technology	1168	3,778	facial recognition	98	1,401	technology	475	technology	693	ai ethic	roc nation	sea ice
5	human	1041	3,329	neural network	85	1,342	human	477	facebook	361	virtual reality	mri scan	hate speech
6	algorithm	566	3,148	ai system	78	1,321	algorithm	227	algorithm	339	ethics board	lyme disease	stem cell
7	robot	447	2,709	deep learning	68	1,277	machine	260	human	564	new job	prime minister	police department
8	machine	597	2,702	using artificial intelligence	67	1,276	google	179	robot	314	computer science	human right	tipping point
9	Facebook	443	2,58	chief executive	57	1,234	tech	151	researcher	335	henry viii	protein structure	live chat
10	Google	463	2,547	new technology	54	1,222	uk	194	image	439	imagenet roulette	lung cancer	big tech
11	researcher	460	2,521	use of ai	58	1,221	computer	201	google	284	fake news	security research	skin condition
12	computer	472	2,233	science fiction	56	1,205	robot	133	machine	337	self-driving car	human pilot	drone swarm
13	system	1027	2,201	breast cancer	47	1,193	company	327	patient	349	pie recipe	blood flow	hot streak
14	image	544	2,188	tech company	43	1,159	bias	112	system	666	heritage foundation	brain activity	stock market
15	UK	408	2,101	fake video	42	1,151	system	361	predict	218	elon musk	blood sugar	brain wave
16	tech	306	2,1	human brain	36	1,145	researcher	125	use	1541	covenant eye	using ai	night watch
17	scientist	329	2,079	killer robot	34	1,14	digital	141	scientist	229	mr zuckerberg	real world	historical figure

18	company	807	2,056	tech giant	34	1,136	cancer	116	video	329	autonomous weapon	home office	psychedelic drug
19	use	2225	2,044	facial expression	32	1,132	scientist	100	detect	208	arms race	false positive	virtual influencer
20	predict	284	2,013	tech firm	32	1,13	tool	134	app	236	automating poverty	red light	high blood pressure
21	patient	445	1,979	recognition technology	32	1,128	science	139	computer	271	cervical cancer	vocal synthesis	sperm whale
22	app	320	1,971	mental health	31	1,127	develop	176	brain	216	atypical hyperplasia	soft skill	vital sign
23	science	420	1,963	computer program	31	1,127	deepfake	68	firm	223	dr roper	hormone therapy	loved one
24	detect	268	1,93	clinical trial	31	1,127	facebook	82	science	281	porn scam	space command	jeff bezos
25	develop	521	1,924	dark matter	42	1,122	expert	98	user	321	local road	ai abacus	ai technology
26	video	435	1,92	computer science	31	1,122	use	684	drone	164	trans people	pressure sensor	untreated sewage
27	cancer	313	1,898	prime minister	35	1,122	research	191	train	232	breast lump	university college london	emotion recognition
28	firm	310	1,892	learning algorithm	30	1,119	future	131	camera	210	public service	downing street	pickup line
29	train	329	1,885	self-driving car	32	1,114	train	97	develop	345	electric toothbrush	bionic hand	oesophageal cancer
30	brain	276	1,859	home office	36	1,112	learn	156	company	480	intelligence official	air force	Deep Nostalgia
31	expert	293	1,838	virtual reality	27	1,11	potential	101	identify	245	care home	artificial intelligence company	head movement
32	identify	333	1,793	ai model	27	1,11	app	84	cancer	197	assistant professor	cancer screening	critical technology
33	user	419	1,79	big tech	28	1,109	firm	87	expert	195	artificial intelligence	clinical trial	black people
34	tool	329	1,759	law enforcement	28	1,108	faculty	74	scan	160	secretary of state	chief adviser	conspiracy theory
35	drone	193	1,747	hate speech	26	1,107	predict	66	uk	214	criminal justice	sugar level	chat service
36	deepfake	175	1,72	fake news	26	1,106	us	280	tech	155	practical application	protein folding problem	facebook whistleblower
37	create	546	1,69	computer scientist	34	1,104	drug	89	create	373	human face	bank account	family violence
38	camera	240	1,686	autonomous weapon	25	1,099	automation	58	able	269	baby monitor	oral cancer	rem sleep
39	learn	430	1,68	early stage	24	1,097	recognition	65	china	184	bot swarm	jaguar land rover	health data
40	facial	172	1,672	brain tumour	28	1,096	facial	55	test	254	kay coles james	diabetes patient	brain scan
41	scan	196	1,671	ai ethic	23	1,095	nhs	56	help	482	silicon valley	false negative	mental health
42	bias	178	1,654	jaffa cake	26	1,094	software	96	facial	117	human expert	rights group	free speech
43	software	280	1,637	ai tool	23	1,093	police	98	fake	119	surveillance system	chronic kidney disease	new zealand
44	fake	168	1,627	data science	22	1,091	ethic	54	dr	132	real madrid	sex doll	health condition
45	help	695	1,62	elon musk	22	1,089	voice	84	deepfake	107	black patient	nuclear deal	human operator
46	China	251	1,611	facial recognition technology	21	1,086	work	437	tool	195	killer robot	intelligence algorithm	law enforcement

47	digital	247	1,61	tipping point	21	1,084	gpt-3	48	learn	274	learning system
48	future	321	1,578	sea ice	21	1,083	detect	60	software	184	Project Debater
49	research	461	1,567	lung cancer	22	1,082	create	173	disease	168	vulnerable people
50	risk	284	1,563	new artificial intelligence	20	1,082	decision	95	professor	150	professor kirchengast
51	able	347	1,563	new ai	20	1,082	video	106	risk	190	
52	test	332	1,551	human being	27	1,08	data	76	musk	95	
53	recognition	162	1,543	natural language	20	1,079	identify	88	study	302	
54	disease	230	1,54	silicon valley	20	1,078	chief	72	accuracy	104	
55	Dr / Doctor	164	1,537	learning system	19	1,077	fiction	52	model	228	
56	analyse	139	1,532	high risk	21	1,077	fake	49	sensor	112	
57	model	326	1,507	milky way	19	1,074	surveillance	48	doctor	137	
58	professor	196	1,496	protein structure	18	1,073	risk	94	monitor	126	
59	potential	223	1,492	blog post	21	1,073	analyse	46	person	225	
60	DeepMind	119	1,489	data scientist	18	1,072	image	105	analyse	93	
61	voice	209	1,488	donald trump	20	1,071	government	151	neural	87	
62	university	407	1,46	drone swarm	17	1,07	concern	89	deepmind	81	
63	accuracy	127	1,456	using deep learning	17	1,07	brain	60	future	190	
64	NHS	121	1,453	ai algorithm	17	1,07	deep	72	claim	185	
65	monitor	161	1,449	artificial intelligence technology	17	1,07	job	108	warn	97	
66	improve	250	1,446	facial recognition software	17	1,07	welfare	47	face	206	
67	Musk	110	1,444	racial bias	17	1,069	patient	96	recognition	97	
68	Amazon	128	1,439	recognition software	17	1,069	problem	133	improve	169	
69	face	298	1,439	study author	17	1,069	university	145	reveal	122	
70	police	221	1,435	learning model	17	1,069	deepmind	38	research	270	
71	study	398	1,43	police department	18	1,067	write	149	robotic	79	
72	reveal	174	1,429	video game	23	1,067	help	213	diagnose	84	
73	US	651	1,428	human right	30	1,066	gender	48	protein	98	
74	surveillance	115	1,425	deepfake video	16	1,066	gebru	37	university	262	
75	warn	127	1,418	ethics board	16	1,066	doctor	60	treatment	147	
76	treatment	214	1,414	ai research	16	1,066	music	99	amazon	86	
77	neural	106	1,413	customer service	23	1,065	healthcare	46	health	251	
78	diagnose	113	1,409	driverless car	16	1,065	social	105	smart	97	
79	claim	253	1,407	mri scan	16	1,065	diversity	44	bot	74	
80	protein	134	1,402	machine learning algorithm	16	1,065	automate	37	spot	112	
81	robotic	103	1,401	surveillance system	16	1,064	staff	79	device	156	
82	person	286	1,4	black people	17	1,064	world	202	voice	125	
83	sensor	127	1,399	air force	19	1,063	china	67	mr	100	
84	social	282	1,386	kidney disease	16	1,062	amazon	42	covid-19	81	
85	deep	173	1,379	ai program	15	1,062	prof	35	driver	109	
86	bot	96	1,378	intelligence algorithm	15	1,062	programme	55	target	128	

cancer patient	real person	
new technology	learning model	
deep blue	customer service	
robot submarine	neurological disease	

87	task	161	1,378	blood vessel	17	1,062	user	98	footage	74
88	team	395	1,378	intelligence program	15	1,062	openai	33	team	266
89	Covid-19	111	1,377	dating app	15	1,061	generate	56	dementia	70
90	media	217	1,373	digital technology	16	1,061	employee	65	network	159
91	health	356	1,372	training datum	15	1,061	chess	35	autonomous	70
92	spot	153	1,371	social media post	15	1,061	ethical	37	photo	132
93	generate	152	1,367	university college london	15	1,06	improve	81	giant	81
94	target	179	1,365	live chat	15	1,06	microsoft	42	military	120
95	automation	98	1,363	stem cell	18	1,059	replace	57	task	109
96	speech	129	1,362	body language	15	1,058	model	98	launch	116
97	smart	122	1,348	traffic light	15	80,501	speech	46	potential	122
98	concern	206	1,348	human pilot	14	0	public	129	tweet	68
99	autonomous	90	1,342	new system	14	906	recognise	38	medium	141
100	decision	201	1,342	intelligence company	14	1,991	medium	76	instagram	65

RESÜMEE

TARTU ÜLIKOOL
ANGLISTIKA OSAKOND

Maria Johanna Milder

News media coverage of artificial intelligence: a corpus-based study of two British online media sources

Tehisintellekti meediakajastus: kahe briti võrguväljaande korpusanalüüs

Bakalaureusetöö
2022
Lehekülgede arv: 59

Annotatsioon:
Tehisintellekt on tehnoloogia, mis järjest areneb ning mis seeläbi on ka meedias ülemaailmselt tähelepanu pälvinud. Üks tehisintellekti juhtivaid riike Euroopas on Ühendkuningriigid, kuid kelle elanikud ei mõista kui palju nad juba praegu tehisintellektiga kasvõi igapäevaselt kokku puutuvad. Bakalaureusetöö eesmärk on uurida tehisintellekti kajastumist kahes briti võrguväljaandes, The Guardian'is ja MailOnline'is. Uurimistöö korpus koosneb 486 artiklist, mis koguti LexisNexis andmebaasist. Artiklid sisestati Sketch Engine tarkvarasse, mille abil viidi läbi korpuse analüüs, rõhuga võtmesõna ja konkordantsi tööriistadel. Sissejuhatuses tutvustatakse Suubritannia positsiooni tehisintellekti maastikul, meedia rolli arvamuse kujundamisel ning kahte erinevat briti võrguväljaannet. Esimene peatükk annab täpsema ülevaate tehisintellektist ja selle kasutustest, kuidas meedia on kajastanud tehisintellekti ning milline on avalik arusaam tehisintellektist. Teine peatükk kirjeldab korpuse kompileerimist ning tutvustab analüüsimeetodit. Seejärel viiakse läbi korpuse analüüs ning tuuakse välja tulemused, millele järgneb diskussioon. Kokkuvõttes kommenteeritakse tulemusi ning antakse soovitus võimalikuks edaspidiseks uuringuks. Korpuse analüüsist selgus, et briti meedias esinevad tehisintellektiga seoses enim tehnilised võtmesõnad, mis esindavad tehnoloogiad, millest üldsus enamasti aru ei saa. Tervishoid on analüüsitud ajavahemikul kõige silmapaistvam teema, osaliselt seoses Covid-19ga, aga ka üldiste tehisintellekti arendustega seoses. Teiseks, MailOnline ja The Guardian erinevad tehisintellekti kajastamisel nii formaadi kui ka poliitilise suunitluse poolest: The Guardian rõhutab tehisintellekti võimalikke piiranguid ja ohte, minnes rohkem süvitsi ning MailOnline keskendub tervishoiu- ja relvatööstuse oluliste sündmuste kajastamisele. Mõlemas ajalehes esinevad märksõnad rõhutasid siiski enamasti seda, kuidas tehisintellekt juba eksisteerib inimeste igapäevaelus. Kolmandaks on tehisintellektiga seotud võtmesõnad muutunud analüüsitud ajaperioodil: 2019. aastal rõhutati tehisintellekti eetikat, 2020. aastal tervishoidu ja Ühendkuningriiki ning 2021. aastal tehisintellekti uuenduslikke saavutusi. Lisaks, tuli uuringust välja, et üldsuse arusaam on otseses seoses uudismeedia kajastusega teatud teemadel.

Märksõnad: tehisintellekt, Ühendkuningriik, meedia, korpusanalüüs

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Maria Johanna Milder,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

News media coverage of artificial intelligence: a corpus-based study of two British online media sources,

mille juhendaja on Ülle Türk,

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Maria Johanna Milder

Tartus, 24.05.2022

Autorsuse kinnitus

Kinnitan, et olen koostanud käesoleva bakalaureusetöö ise ning toonud korrektselt välja teiste autorite panuse. Töö on koostatud lähtudes Tartu Ülikooli maailma keelte ja kultuuride kolledži anglistika osakonna bakalaureusetöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

Maria Johanna Milder

Tartus, 24.05.2022

Lõputöö on lubatud kaitsmisele.

[Juhendaja allkiri]
[Juhendaja nimi]

Tartus, [pp.kk.aaaa]