



King's Research Portal

DOI:

[10.1145/3600211.3604715](https://doi.org/10.1145/3600211.3604715)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Walker, J., Thuermer, G., Vicens, J., & Simperl, E. P. B. (2023). AI Art and Misinformation - Approaches and Strategies for Media Literacy and Fact Checking. In *AIES 2023 - Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 26-37). (AIES 2023 - Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society). ACM. <https://doi.org/10.1145/3600211.3604715>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



AI Art and Misinformation: Approaches and Strategies for Media Literacy and Fact Checking

Johanna Walker*
King's College London
London, UK
johanna.walker@kcl.ac.uk

Gefion Thuermer
King's College London
London, UK
gefion.theurmer@kcl.ac.uk

Julian Vicens
Eurecat Technology Institute of Catalonia
Barcelona, Catalonia, Spain
julian.vicens@eurecat.org

Elena Simperl
King's College London
London, UK
elena.simperl@kcl.ac.uk

ABSTRACT

Misinformation in its many forms is a substantial and growing problem for society today. Whether financially or ideologically motivated, purveyors of misinformation do not abide by legal, technical or moral rules. Therefore new, ludic, narrative, gamified and artistic approaches are needed. In this paper we analyse the approaches taken in countering misinformation by 18 AI and machine learning works of art, developed in the MediaFutures project. We examine how these align with existing AI approaches to countering misinformation, and how they address some of the key challenges. We show that AI artists engage with existing debunking and inoculating strategies, including highly technical aspects such as deepfakes, while also utilizing focused strategies of data literacy and collective intelligence. We also find that they are able to integrate hard-to-refute strategies such as narrative and emotion. These findings suggest that data as an art material and AI techniques as art tools are worth of further investigation as to their effectiveness for countering misinformation within society.

CCS CONCEPTS

• **Information systems** → *Social recommendation*; • **Computing methodologies** → **Machine learning**; • **Applied computing** → **Media arts**.

KEYWORDS

Misinformation, fake news, disinformation, AI art, media literacy, fact checking

ACM Reference Format:

Johanna Walker, Gefion Thuermer, Julian Vicens, and Elena Simperl. 2023. AI Art and Misinformation: Approaches and Strategies for Media Literacy and Fact Checking. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*, August 08–10, 2023, Montréal, QC, Canada. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3600211.3604715>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

AIES '23, August 08–10, 2023, Montréal, QC, Canada

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0231-0/23/08.

<https://doi.org/10.1145/3600211.3604715>

1 INTRODUCTION

Misinformation of various types has been around for centuries, but has grown as the technologies that enable its spread have grown. This matters as misinformation negatively affects the wellbeing of individuals, groups and society in a number of ways. It can undermine democracy [43], reduce climate change consensus [54], exacerbate crises [50] and even lead to death [13]. Further, the proliferation of ways in which misinformation can be encountered also matters, as repeated exposure to a piece of misinformation boosts its likelihood of being believed [35]. Misinformation is also more compelling when it is delivered in emotional language, or designed to be attention-grabbing [35].

Social media is a key forum for misinformation as it enables even those without formal knowledge of the workings of mass media to become content creators, challenging the domain of traditional media [25]. In the run up to the 2016 elections over one quarter of Americans visited a fake news site, visiting an average of 5.5 articles each [37]. Algorithms are a key component of social media, recommending content to users that will keep them engaged with the platform and so provide exposure for adverts [33]. This process is opaque to users and so has been called ‘invisible attention engineering’ [55]. Misinformation is frequently designed to work with the algorithms to find its way into users’ feeds [55].

Art, in the form of cave paintings, predates modern humanity. These paintings and stories shared orally were a way to share knowledge before the existence of written language. The narrative patterns enabled them to be more readily absorbed and remembered by their recipients, who needed the information contained within them to survive. The arts, therefore, are a key way of enabling people to develop the knowledge they need in order to engage effectively with the world around them.

Artists have been working with AI since the 1970s [25]. Algorithms and art are a current topic of much interest, both in popular culture and academia, fueled by the accessibility of algorithmically generated art using large language models such as Dall-E. However, there is also an emerging field of Critical AI art that seeks to engage and comment on society, and with it, attention given to the ethical use of AI tools in this critical art [18].

In this paper we look at art that exposes, refutes or counters various forms of misinformation, by using data science and AI approaches. We describe our use of the term misinformation and some of its key impacts and motivations. The key strategies used

to counter it, media literacy and fact checking, plus algorithmic methods for these, are explored. We show that AI art is used to highlight social tensions in AI itself and as a critique of technology, making it a potential strategy for dealing with misinformation. We identify the strategies used to challenge misinformation online by AI artists through interviews. We find that as well as using current technologies from the computing and journalism sectors, artists incorporate emotion, narrative and explicit ethical appeal in order to counter misinformation.

Our research questions therefore are: 1. What strategic approaches does AI art take to countering misinformation? 2. Which data, tools and techniques are utilized? 3. How does the artistic approach add value to algorithmic approaches to countering misinformation?

This paper makes two key contributions. Firstly, it contributes to the corpus on countering misinformation by analysing a specific and emerging AI approach, and examining how this may support more established approaches. Secondly, it contributes to the growing corpus on AI and the arts by describing the use of AI for artistic creation in a specific context.

2 BACKGROUND LITERATURE

2.1 The challenge of online misinformation

There is no agreed typology of the kinds of incorrect information that is disseminated online both knowingly and unknowingly, although there are some taxonomies created within specific terms [32], [40]. The term ‘misinformation’ is generally understood to mean incorrect information in the sense that it differs from the best available expert information or established fact [57]. ‘Disinformation’ is seen as more pernicious, and constitutes knowingly incorrect information created for ‘public harm or for profit’ [28]. Since 2016 these terms have been joined, and possibly superseded in popular imagination by the term ‘fake news’ [39] which itself includes a number of formulations, including propaganda, trolling, conspiracy theories and satire [53]. Throughout this paper we use employ the term ‘misinformation’ to represent all forms of incorrect news messaging.

Misinformation in the political sphere has led to “cognitive fallout” where people can continue to believe misinformation even after having been told it is untrue [35]. This “continued influence effect” of misinformation has been demonstrated repeatedly [30], [11]. One of the “fingerprints” of misinformation is its emotional appeal, with high emotionality ensuring virality and hindering analytical explanation [14]. As a consequence, misinformation spreads to up to 1,000 times more people than factual information [56], although some research has shown this is a function of the larger size of some fake news cascades [31].

The growth of large language models has increased the democratic access to content creation online [48], but this very facility presents a threat from machine generated fake news. Another form of machine generated fake news is ‘deepfake’ videos. These use generative adversarial networks to create believable yet false media [38]. The Korean news channel MBN created a deepfake of its anchor Kim Joo-ha for a bulletin [16], however, the majority of deepfakes are created for the purposes of misinformation. Well-known deepfakes exist of world leaders. such as US film director Jordan

Peele’s deepfake of Barack Obama discussing deepfakes [44], or more malignly, Vlodymyr Zylinsky suggesting Ukrainian soldiers lay down their weapons [17].

The key motivations for the production and dissemination of misinformation are financial and ideological [7]. Making money via advertising revenue generated by clicks and views on websites is attractive to the fake news media as well as the real news media. An example of this is the far-right conspiracy theory and fake news website InfoWars, which at one point in 2018 made USD800,000 a day. Financially motivated fake news sites often use narrative techniques (clickbait) to appeal to the unwary, such as ‘You won’t believe what Obama says in this video.’ [44], [7].

Ideological motivations are more complicated. Propaganda ensures when official bodies spread fake news for ideological reasons. Satire is fake news spread by legitimate bodies for the purpose of entertaining their audience. Trolling is fake news (usually) spread by individuals for their own entertainment or personal purposes. Largely, these are all driven by requirements for power and influence, albeit in different ways [7].

2.2 Data-driven approaches to countering misinformation

The two key approaches that have been taken to stemming the spread of misinformation online are fact-checking and the development of media literacy within populations (especially youthful ones).

2.2.1 Fact checking. Fact checking is “the systematic publishing assessments of the validity of claims made by public officials and institutions with an explicit attempt to identify whether a claim is factual” [59]. As such it is conducted after the dissemination of misinformation and is colloquially known as ‘debunking’ [35]. It is performed by both journalistic and non-journalistic bodies such as the Associated Press and FactCheck.org. Both of these organisations use human-led, investigative reporter techniques for fact checking. However, effectiveness of fact-checking as a tool for counteracting misinformation is somewhat undermined by people’s unwillingness to accept corrective fact-checking [10]. While “falsehoods” can be corrected, feelings are more challenging [36].

Computationally-oriented approaches are primarily based on knowledge graphs [19]. There are automated approaches that focus on fake news detection, such as Hoaxy [46], which visualise the spread of information on Twitter, and websites that identify how much content around a news story appears to be linked to bots such as Botometer [1]. ClaimBuster flags up claims that appear to be worthy of checking, based on a combination of natural language processing and machine learning [24]. However, this is a very emergent area, and “the potential for automated responses to online misinformation that work at scale and don’t require human supervision remains sharply limited today” [24]. With this aim, several automated fake news detection techniques have been developed, based on algorithmic techniques such as random forests (multiple decision trees), content features for classification and neural networks, although they have not necessarily been very successful [26]. One challenge is the variety of types of fake news which has no agreed taxonomy [32].

Further, collecting reliable datasets of fake and trustworthy news on which to train these techniques is not a trivial task and no benchmark dataset exists [36]. Datasets of importance include FEVER, consisting of 185,445 claims generated by altering sentences extracted from Wikipedia and LIAR, based on statements from Politifact.

Approaches to counter machine generated misinformation have been based on identifying the difference between machine and human generated writing. However, they have been less successful in distinguishing legitimate from false machine generated writing, as, unlike humans, machines do not alter their styles between true and false information [45]. Techniques for countering deepfakes include many versions of artifact detection, strict blockchain data provenance for multimedia, and counter attacks [8].

2.2.2 Media Literacy. Media literacy is the development by an individual of a set of skills around critical thinking, evaluation strategies, search skills and knowledge of the news and media industries [12],[20]. Limited research on games and gamification shows that these might also help media literacy. This suggests active engagement with different literacy skills, rather than passive instruction, is important [20]. Media literacy is seen as a ‘pre-emptive’ approach [35] and has in the past been characterised as an ‘psychological inoculation’ approach in that it seeks to arm the individual with the skills necessary to critically appraise and identify misinformation when coming into contact with it [36],[12].

A number of studies confirm that media literacy is correlated with the ability to identify misinformation, or engage critically with information online [20]. In one such study a practical test was administered to 63 adults aged 19-24, which showed that critical evaluation behaviours were positively correlated to the correct identification of fake news stories [34]. Unlike debunking, inoculation works equally despite prior attitudes [36].

Computational approaches to media literacy are based on identifying the algorithmic knowledge a media literate public would need to possess, and delivering that through computing education. A key component of this is raising awareness of aspects such as invisible attention engineering, bots/agents, content filtering and tracking [55]. This involves not only creating awareness of the technical capabilities, but their implications. In a post-digital world, in which the digital world and the ‘real world’ are no longer meaningfully separate, the technical and the political can no longer be separate [29].

The existence of ‘big’ datasets for training purposes is a key reason AI research is currently flourishing. Hence, the need for a data literacy within media literacy in order to understand the potential impacts of AI, given that the datasets themselves are frequently problematic [29]. Even more specifically, there have been calls for an algorithmic literacy [47].

In addition to these purely educational approaches, a computational approach has arisen that involves automatically notifying readers of the pragmatics of the content, in order to augment their media literacy in context. These could then identify ways that the reader is being manipulated without their awareness, such as by certain tropes being used as narrative frames [9].

Contemporary media literacy tends to organize around five themes: youth participation, teacher training and curricular resources, parental support, policy initiatives, and evidence base construction [12]. However, older adults are particularly susceptible to

fake news and consume substantially more than younger people. This may be in part due to not being ‘digital natives’ and having less facility with the technology on which they encounter fake news [42].

2.3 Critical AI art

According to the World Economic Forum, “Giving people access to data most often leaves them feeling overwhelmed and disconnected, not empowered and poised for action. This is where art can make a difference. Art does not show people what to do, yet engaging with a good work of art can connect you to your senses, body, and mind.” [21]. The participation of artists in technological or scientific projects has proven to question technologies, increase citizens’ awareness, explore hypothetical paths for progress, enhance and humanize technologies [21].

The first decade of this century saw AI artists exploring natural language processing, computer vision and pattern recognition. In the 2010s deep learning technologies enabled greater expansion [25]. Non-fungible tokens, facilitating a market for digital art, have further brought AI art into the mainstream [25]. High-profile projects such as *The Next Rembrandt*, which used deep learning to generate a ‘typical’ Rembrandt painting, tapped into corporate sponsorship as well as industry, public sector and academic partnerships [2].

A specific form of AI artwork is critical AI art – used to address social tensions arising from technology and to enable a sense of critical distance from the technology [51]. A number of artworks, such as *Capture* (2020) and *DataMasks* (2014), engage with what has been called “algorithmic anxiety” around the growing ubiquity of facial recognition technologies [58]. *DataMasks* created masks that, “are shadows of human beings as seen by the minds-eye of the machine-organism” [15]. Projects such as MIT’s *Crowd-Sourced Intelligence Agency* (2015) expose dataveillance practices by allowing visitors to assume the role of security analysts and monitor and analyse their friends’ tweets [23]. The well-known *ImageNet Roulette* (2019) uses classification techniques to illustrate what happens when technical systems are trained on problematic data [3].

As well as being fake news, deepfakes are used as art work to address the dangers of deepfakes themselves. The deepfake *In Event of Moon Disaster* (2020) purports to show a recording of President Nixon delivering a contingency speech in the event of the Apollo 11 moon mission ending disastrously. It was created by MIT to educate people about the existence of deepfakes and the challenges of identifying them [41]. Artist Bill Posters created a deepfake of Mark Zuckerberg, entitled ‘I wish I could...’ (2019), which was commissioned for exhibition to raise awareness of how people can be manipulated by social media. This video, showing Zuckerberg boasting that Facebook owns its users, was subsequently posted to Instagram (and was still there at the time of writing). Such works explicitly engage with and critically examine the ethics of AI.

In particular, critical AI art facilitates such aspects as linking “underlying technical systems to structural issues of power”, enables experiential learning, and crucially, allows interpretation rather than straightforward explanation [27]. There are a number of projects that aim to use art based on data technologies to interrogate existing structures or create new insights. The *British Antarctic Survey Data as Art* programme developed a series of artworks using Antarctic data, with the aim of engaging a broad audience with the question

of, ‘why is this data important for society?’ [4]. RAND Art + Data engaged artists to create visual stories with the aim of challenging the audience to think differently about policy analysis, including topics such as barriers to Covid vaccination and Russian propaganda [5]. The EU S+T+ARTS programme (science plus technology plus the arts) aims to support interdisciplinary teams of artists and technologists to creatively innovate in a host of fields [52]. The project DataStories aimed to investigate data in a “post-truth environment” using a variety of narrative approaches across a number of media including film [6]. There is sufficient volume of artists working in critical AI to provoke the need for a taxonomy of data as a material [22].



Figure 1: Data-Masks Installation in Karlsruhe, 2015 [Public domain], via the artist’s website.

3 METHODOLOGY

To investigate the use of fact-checking and media literacy strategies by critical AI artists, we conducted interviews with participants in the Horizon Europe project MediaFutures.

3.1 MediaFutures

MediaFutures is part of the European S+T+Arts programme. It offers grant funding and support for startups and artists, via open calls held in 2020, 2021 and 2022. In MediaFutures, artists are asked to use data as an art material to create works that question the impact of misinformation on individuals and society. The 1st cohort graduated in September 2021, the 2nd cohort graduated in April 2022 and a 3rd cohort started in November 2022 and will graduate in June 2023.

Currently, 10 art pilots have concluded, with a further 5 currently in the third cohort of the programme. A further 8 projects where artists collaborated with startups have also concluded, giving a participant pool of 18 projects. (There are currently 4 more such projects in the programme.) Of the concluded art projects, some are in their final version and have exhibited their work, while others are still pre-exhibition. Of these projects, some use AI to explore and challenge AI, while others use AI/machine learning to explore non-AI contexts. This includes both ‘pure’ art works and art works with a commercialisation element, developed by artists and startups

together. A description of all the art projects in the first two cohorts of the programme can be found in the Appendix.

3.2 Interviews and analysis

We interviewed the artists and startups behind 18 art projects in the MediaFutures programme. These interviews took place after the end of the first and second rounds of the MediaFutures residency and acceleration programme in 2021 and 2022. The interviews were structured, took place remotely and were accompanied by three questions on the data, tools, and techniques used which were sent and answered via email (see Appendix B). The interviews were then transcribed and thematic analysis was conducted. The projects are referred to by ID numbers in the results. Our interview pool was constituted by all the art and artist/startup projects that were selected for the second phase of the MediaFutures project (developing the project to the exhibition/pre-exhibition stage). The interview guide can be found in Appendix 2. We then applied inductive analysis to the interview transcripts, developing codes as we read.

4 RESULTS

4.1 Approaches to media literacy and fact checking

The projects in MediaFutures engaged in both debunking and inoculation strategies. The projects that focused on debunking ([774452], [580713], [504746] amongst others) were aware of the necessity for datasets in combatting fake news, and created new datasets by bringing together multiple different sources. “Specialists cannot combat misinformation if they don’t have data to analyse it... we wanted to do is to create a dataset of fake news, which are specific to [a] region, because [it] is highly underrepresented in terms of misinformation... So we talked with all these organisations, and we say, Let’s unite all these databases in order to create a big data set for the academic community” [580713]. Artistic engagement was seen as a unique way not only to spread information but also to gather hard to access data. “If you don’t understand street stories, you don’t understand what’s really going on, and you don’t get street stories in response to surveys or in response to experts going into interviews or running focus group” [776326]. Those datasets are used in many works for building AI models that allow classifying claims based on their feasibility, auto fact-checking, or fine-tuning pre-existing AI models for a specific context or topic. For instance, Computer-Assisted Recognition of Denial and Skepticism (CARDS) is employed in classifying different types of misinformation on climate change.

Many artists were familiar with and engaged with media literacy theory. “It’s very important to understand how to debunk them, how to find a way to raise the resilience and the capacity of people dealing with all this information... and also the creation or the critical thinking of people” [774452]. Artists frequently specified exactly how they believed media literacy creates protection against misinformation. “The main aspects, which we wish to deliver, as an impact to society is strong, critical thinking, super important, as we believe that educated persons can’t be influenced by propaganda so easily” [423794.] Others extended the link from critical thinking to behaviour change. “So [engaging with the artwork] as



Figure 2: Epic Sock Puppet Theater uses datasets of right and left wing ‘sock puppet’ social media accounts, the words of which are then spoken by animatronic sock puppets [Public domain], via MediaFutures.

the foundational of media literacy, of behavioural changes, I guess” [504746].

In terms of data, Twitter was a commonly used source of data in artworks that took a media literacy approach. [984662, 193374]. The GDELT Global Difference Graph (GDG) databases were used to analyze more than 250,000 headlines of Russian-language media publications for fake news identification [774452]. However, artists often created their own datasets for various reasons. One artist ran up against the problem of there simply being no appropriate existing dataset for Eastern Europe [580713]. The LIAR and LIAR+ datasets were used in this project not for their data, but to inform the structuring of the new dataset to address fake news. Another required a very small, specific dataset [859977]. Others sought to collect data that did not otherwise exist, for instance, micro-narratives [776326]. One artwork consisted of a browser plug in that used 11 data sources, including Wikidata and review site TrustPilot, to make visible the underlying ethics of certain websites [504746].

The artists were, overall, very clear on how data-driven art builds media literacy. For instance, “every time disinformation is picked up in the sense of sound or data, there’s always this critical approach in revealing mechanisms, but also in the systematic refutation of the disinformation. You see the digital analysis mechanisms and you hear refutations so there’s a strong critical thinking here” [831967]. One team believed that debunking approaches were ineffective, and disrupting people’s patterns of thinking would be more effective. “I think that there seems to be a dominant view, ... that you handle myths by making people more aware of facts, and that’s a really bad sign ... we think art can provide a line of flight by which people can escape from those dominant patterns” [776326]. One project moved away from either approach to consider less how misinformation is countered, towards, “how do we improve access to accurate information” [859977].

Another focus around data was increasing data literacy, enhancing the ability of the audience to understand aspects of the data, either what it contained or how it was created. This might be simply provoking interest in data, by creating an “immersive experience [that] can become a kind of stepping stone for these people to

get curious about” [369215]. Some goals were very ambitious, aiming for “people [to] understand better large amounts of data in the making. What is data, how it’s done, how you make sense of a database” [758112]. One artist found that collecting data from others, and analysing it, subverted the dominant trope, “as users, companies and governments are analysing us, and in this case we are analysing” [774452].

Educational aspects appear, therefore, to be very present in how the artists are thinking about their artworks and what they want participants, viewers or audiences to take away with them. “Instead of exit through the gift shop, we’re thinking it’s kind of like exit through the educational aspect of why this exists and kind of details about the process of how these were made, implications for how synthetic media is being used in the world today” [060672]. However, various strategies were employed here. Some were more ludic approaches, “the whole mission of our project is to educate people about what this information is methods of manipulation of their brains, but make it in a gamified way” [423794]. Others explicitly used the base of education to build more concrete behaviour change upon. “I call it an educational art tool. But we have ideas for how it’s not just be this kind of individual tool that you use to educate yourself. It’s how do we turn you from this individual into collective of individuals. The long objective for this is to use [the artwork] as a form of protest” [504746]. Adding these layers of complexity to simple educational tools also appeared in other ways. One of the artist/startup teams particularly noted that it was important to “educate from within” in terms of the infrastructures worked with, but also that their educational tool needed to “allow for these ambiguities and uncertainties” [758112] surrounding the particular data they worked with.

4.2 Tools and techniques

The artworks created varied widely in form. Some were installations which the audience could experience physically (e.g. 730 Hours of Violence, How many heartbeats to send an email), Critical Climate Change Machine) some were on the web (e.g. Invisible Voices, Social Sandwich) and some in the metaverse (e.g. Time Lapse Migration). Like ImageNet Roulette, many of the works took a hybrid form and can be exhibited in person and on the web. MediaFutures offered technical support to the artists, but the artists involved are often technically highly skilled. “Because I have an interest as an artist, I think that as an artist, we have to be involved in developing our digital tools... You know, that’s my art.” [758112] “the artists knew more on NFT than us [the startup partner] because they are really into digital art” [748452].

The use of AI and computational methods varied widely across the art works. Some projects were highly complex and used as many as 20 tools or libraries, among them AI models for classifying, predicting or generating content [101362]. A number of the more technically simple projects, such as generating sound or images processing data from social networks, like Twitter, utilized only 4 [984662]. There were also a number of artworks that used data as a source for building the narrative, rather than incorporating it directly in AI, or other computational, models (734815, 859977).

One artist working in the media literacy space described the process as “classical data science pipeline... classical data science

tools” [423794]. Many artists followed this traditional pipeline of data collection, preparation, exploration, visualisation, experimentation and classification. Some of them used visualisation as the end product [774452], and there was the use of supervised classification and regression models (e.g., K-nearest neighbours, DBSCAN or Neural Networks).

One network-based artwork, which had an accompanying commercial product, used knowledge graphs for extracting insights of data. Another work, in the same vein, used network science analysis techniques to achieve perform clustering of networks and in graphs and, ultimately, visualize them.

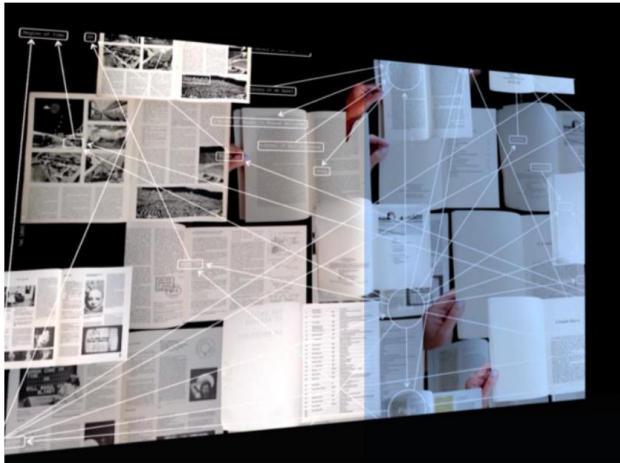


Figure 3: Illustration from Bibliograph, showing how objects in a library are related. [Public domain], via MediaFutures.

One of the most advanced uses of AI in the artwork was for generating synthetic content. Open-source tools like Tools such as the faceswapping framework SimSwap were used by artists addressing deepfakes who also used generative adversarial networks [353266]. Visitors to Oracle Network were introduced to deepfakes by animation of their own images, with First order motion models, which were also used for The Evil Magic Mirror and Soft Evidence. Not all artists developed their work from scratch. One work was based on an existing classifier that had been developed in research. Amongst artists generating, creating and exploring the impacts of deepfakes, existing datasets such as WIDER FACE were used. However, ethical concerns, primarily regarding whether consent for images of real people to be utilized for deepfakes could properly be described as informed, led some of the artists employing deep fake technologies to the creation of their own datasets using generative adversarial networks [353266, 060672].

Most of the artworks developed their work based on open-source tools following the recommendations of mentors in the residency and acceleration programme, however others created their tools from scratch. Consequently, MediaFutures encouraged the publication of datasets and code created by the artists and the algorithms created for an artwork on propaganda are available for others to use in working with propaganda narratives [774452] Again, concern about misuse meant that artists working with deepfake technology were reluctant to do this. [569260].

4.3 Emotion, narrative and ethics in misinformation countering strategies

“Art has a special way, in the way we make sense of things, and we make sense of data. And that art has that kind of level of abstraction that can bridge things” [776326]. The power of art was seen to be facilitative, with art going beyond existing data visualisation techniques and enabling exploration of “much more than representations of the data” [996510]. Others felt the relationship between data and art worked “efficiently” and, “the data approach can be enriched with the artistic methods” [774452]. It was believed that, “this interdisciplinary connection between data and art is becoming more and more important and more valid” [758112].

Narrative, or story-telling, was seen as a key aspect of many of the artworks. Simply consumed narratives easily go viral. “So one of the ways that people can find stories, and one of the ways they can discover meaning, is through art, through cartoons. Cartoons are very effective. They go viral” [776326] Another artist described their product as a “storytelling platform” [774452]. Narrative was seen by one team as effective in creating behaviour change. “That’s one of the things we do with narrative... So one of the ways you get people to change is ask them a question that they can’t answer without thinking or acting differently” [776326].

Emotion is often seen as a tool of misinformation, but the artists frequently engaged with this in their work. One artist used AI emotion analytics to gather information on people interacting with their artwork, and personalised the experience based on that. “There are three emotions, focused, energetic and rage. And for each of those, I create a specific soundtrack. So depending on the emotion when the user is using his mouse, it will say okay, it seems like you’re a bit energetic, so it’s going to play the energetic track and that way for everybody, the game is going to sound different” [423794]. Emotion was seen as a way to grasp the attention of users. “We need a very strong emotional reaction and engagement in order for people to start understanding how important it is to seek truth, you know, what is truth?” [266713].

Other artists focused on the importance of removing emotion to create neutral ground on which to discuss otherwise emotionally-heightened subjects. Two works explicitly facilitate this platform for dialogue, while others seek to find a way to integrate multiple sides of the story into their artwork. “I need to learn how to look at [a person who is very conditioned to believing misinformation] side of story as well” [266713]. Artists were seen as people who were particularly capable of communicating multiple points of view, “because artists in their worlds express a lot of narrative stories and different points of view” [734815]. Other pieces take what are normally highly-politicised subjects and make them, “purposefully neutral and apolitical. This allows people to engage on neutral ground for conversation about image manipulation and the risk of misinformation posed by such technologies” [060672].

Sound, visuals and touch were viewed as a way of bringing emotions into the misinformation area in a positive way, and combatting potential for attention drift. These emotions might be positive, “because it’s something engaging and funny... it would catch their mind as being a different way of explaining and visualising something that before was boring,” [201483]. An augmented reality project implemented ‘explosions’ into their game after discovering

users found it, “satisfying” [580713]. Related to this is art’s ability to be multi-sensory and transpose the abstract to the physical. “These abstract concepts, about fake news, ... this is an enormous amount of data and I think it’s very hard to imagine how it looks like, how it sounds like, so we’re trying to appeal to the different senses of the audience” [996510]. “Even though I’m using data and binaries, I wanted to make sure that people understand that [misinformation] has a huge material effect on everybody. I’m using a heartbeat, which, when you put your hand on it, you can literally feel the heartbeat” [266713]. This appeal was deemed necessary as an approach by some artists as a deliberate response to the perceived low refutation abilities of current strategies. “We want to create artistic interpretation of fake news, so that people are engaged that find them catchy, visually appealing, we strongly believe that having very strict and boring way of saying, this is fake news...is not engaging people” [774452]. A further art work built on this by segmenting their audience by psychological learning types and developing elements for visual, sound and kinaesthetic individuals [423794].

While all the artists considered the ethical aspects of their art-work as part of the MediaFutures process, some also surfaced ethical questions to their audience. One artist noted that engaging in a project that sought to educate people about harms required a high level of ethicality, especially around transparency, so people engage with the work can trust them. “Ethics has always been at the heart of it..it’s about trust. And if we’re not doing it right, then then who are we commenting on other people’s [ethics]?” [504746]. As noted above, dataset compilation raised many ethical concerns, in particular with regard to deepfakes. “Oftentimes, when you see a deepfake, this many images are scraped from Google without anyone’s consent. But we created every single one, so also raising awareness about consent and use of data.” [060672].



Figure 4: Still from deepfake video *Soft Evidence* [Public domain], via MediaFutures.

5 DISCUSSION

Most of the art works in MediaFutures discussed above engage with the key existing strategies for countering misinformation online. They largely adhere to media literacy routes but also develop new, clean verifiable data sets, valuable for fact checking, and many worked with deepfakes to raise awareness of the technology’s capabilities and impact. Only one artist rejected the idea that fact-checking would work as an approach. However, an emergent

strategy appeared to be focused around collaborative intelligence, essentially a distributed network where each agent contributes autonomously to problem solving. This approach has been used in, for instance, participatory democracy. One of the challenges for traditional counteraction strategies is that they largely attempt to apply considered and direct approaches to mitigating the effects of often highly emotional misinformation. Many of the art works in MediaFutures deal with the challenge of emotion head on, whether by responding emotionally, or neutralizing the emotion in the misinformation. Simply by being art, and sitting outside of the traditional online arenas of media and social media, art is able to create more neutral ground for the discussion of politicised subjects that are vulnerable to post-truth argumentation. This neutral ground is sometimes highly structured, as in *Social Sandwich*, which reflects other attempts to create less polarised social media online. Similarly, one project used ‘standard’ classifier, clustering and neural network approaches to the detection of fake news, but added the additional element of using emotion to identify to alert people that they had encountered untrustworthy news.

Deliberately trying to engage an audience or making an intervention ‘less boring’ has not necessarily been a key goal of media literacy attempts, many of which work with captive audiences. Artists, however, are experienced in the aim of capturing the attention of potential audiences in a world competing for attention, and this has been the driver behind collaborations such as the British Antarctic Survey exhibition. Through the ability to be multi-sensory (even online, through the use of sound) art has an extra dimension through which to communicate with the audience, and be, in the words of one artist, “much more than representations of data”. This is key, as it is this appeal which enables virality. Many of the interviewees discussed narrative as a compelling technique for engaging with their audience, or for ensuring their audience engaged with each other. This narrative could then be distributed and consumed via any of the multisensory methods described above, from a brief cartoon to a virtual exhibition of refugee art, but with the ability to appeal to the natural human instinct for storytelling. This reflects the findings that artistic approaches enable interpretation by the audience, requiring engagement, rather than a one-way explanation [4].

Although the types of projects selected into MediaFutures meant that our dataset would not include any artists who were using wholly non-AI approaches to critique misinformation, there was still a vast spectrum across the use of data and technology. While there was standard use of such datasets such as LIAR and techniques such as generative adversarial networks, there was also the utilisation of more unusual software such as translation software and the dataset of the Observatory of Cultural and Linguistic Diversity on the Internet [859977]. Through such approaches, AI art may offer a way to engage more tools in the fight against misinformation. As noted, the major task of debunking fake news is establishing reliable training datasets [10]. For some parts of the world these are simply not available, and the creation of these datasets as part of these artworks is a useful contribution. Artists also promote inclusivity via awareness and addressing of the inherent biases in general datasets) that are focused on claims in majority languages (e.g. English) and the lack of representativity of certain countries.

Projects such as the British Antarctic Survey and those studied in [54] focused on educating audiences. Many of the artists we interviewed engaged specifically with educating their audiences around data, to the point that data literacy, rather than a broader media literacy, was their aim. The artists were sensitive to and responsive to the idea that some groups were excluded from the majority of media literacy approaches, and there was focus on making artworks inclusive and accessible in terms of the amount of digital awareness that was necessary to engage with them, despite their underlying use of AI tools and techniques. This is particularly clear in works such as *730 Hours of Violence*, *Soft Evidence* or *How Many Heartbeats to Send a Love Email*, where the audience can experience the art works in comparatively familiar ways.

The artworks frequently demonstrated economy of use with multiple aims, but with the audience able to engage with the art at whatever level they felt comfortable. For instance, a number of art works offer tools that allow individuals to simply engage with the art work, but then provide an opportunity to engage further, either with other individuals or by taking knowledge from the artwork into other parts of life. In this way the artwork operates on a number of levels, as a visual, tactile or sonified experience, an educational tool, and then a tool of active choice, or protest against misinformation. This demonstrates the linking of the technical to the power as described in [4]. While the majority of the artists interviewed dealt with ethics mainly in terms of ensuring their own art was ethical, rather than explicitly engaging with the ethics of AI two art projects made highlighting unethical practices the focus of their engagement with misinformation.

This therefore offers a range of considerations to take into account when designing future technologies or interventions against misinformation. The first is that, as data is an established art material, many artists are well-positioned to bring technical as well as artistic skills to their work, creating highly integrated art works. We also find that narrative is a powerful tool that can be exploited through the data/art relationship and resists easy binaries. Integrating emotion into the response to emotionally-heightened misinformation allows for engagement on a more equal footing, which may help reduce the inequity of virality. We also find the idea of engaging with misinformation not prior to or post exposure, but synchronously, via collaborative and participatory opportunities for engagement, to be compelling and worthy of more investigation.

6 CONCLUSION, LIMITATIONS AND FUTURE WORK

6.1 Conclusion

Efforts to counter misinformation online have been hampered by both an age old truism and a very contemporary concept. On the one hand, humans are emotional beings, who respond to storytelling, whether that story is objectively true or not. On the other hand, there is now, in certain circles, a reluctance to accept anything as objective fact, and a mistrust of experts. This creates fertile ground for the most appealing information to be the most widely shared, regardless of veracity, and for there to be little leverage for counter argument.

Countering misinformation with art, however, addresses both of these aspects. Firstly, we have an emotional response to art we

find compelling. Secondly, although art can certainly be argued about, it cannot be argued with. After all, “you cannot refute a work of art” [49]. AI-driven artistic interventions allow their creators to use some of the same core techniques of media literacy and fact-checking, based on the same AI approaches, but also utilise multisensory and emotional tools that have the possibility of reaching a wide range of demographics.

6.2 Limitations

The strategies outlined above are from one project. They represent only a subsection of the possible approaches. A number of these artworks are yet to be implemented or exhibited, so the aims are still largely theoretical, rather than user tested for efficacy.

6.3 Future work

As it stands, the variety of strategies available to art and artists, the difficulty in countering a work of art and the fact that art can be understood on many levels mean that this intuitively appears to be a useful addition to other forms of media literacy. In particular, more work needs to be done on understanding the role of data literacy and collective intelligence. However, the effectiveness of current interventions is unclear [38]. Taken together, these suggest that more work is necessary to measure the impacts, both immediate and long term, of artistic approaches. Assessment of this is underway with a number of MediaFutures projects.

ACKNOWLEDGMENTS

This research was supported by EU Horizon 2020 research and innovation programme grant agreement 951962. We would like to thank all our colleagues and participants in the programme.

REFERENCES

- [1] [n. d.]. <https://botometer.osome.iu.edu>
- [2] [n. d.]. <https://www.nexttrembrandt.com>
- [3] [n. d.]. <https://paglen.studio/2020/04/29/imagenet-roulette/>
- [4] [n. d.]. <https://www.bas.ac.uk/project/data-as-art/>
- [5] [n. d.]. <https://www.rand.org/about/nextgen/art-plus-data.html>
- [6] [n. d.]. www.datastories.co.uk
- [7] João Pedro Baptista and Anabela Gradim. [n. d.]. “Brave New World” of Fake News: How It Works, Javnost. *The Public* 28(4) ([n. d.]), 426–443. <https://doi.org/10.1080/13183222.2021.1861409> DOI.
- [8] Alessandro Bondielli and Francesco Marcelloni. [n. d.]. A survey on fake news and rumour detection techniques. *Information Sciences* 497(4) ([n. d.]), 38–55. <https://doi.org/10.1016/j.ins.2019.05.035> DOI.
- [9] Miriam Boon. [n. d.]. Augmenting Media Literacy with Automatic Characterization of News along Pragmatic Dimensions In. In *CSCW '17 Companion, February 25*. Portland, OR, US DOI. <https://doi.org/10.1145/3022198.3024948>
- [10] B. Brendan Nyhan, E. Ethan Porter, and Jason Reifer. [n. d.]. Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Polit Behav* 42 ([n. d.]), 939–960. <https://doi.org/10.1007/s11109-019-09528-x>
- [11] M. Buczel, P.D. Szyszka, A. Siwiak, M. Szpitalak, and R. Polczyk. [n. d.]. Vaccination against misinformation: The inoculation technique reduces the continued influence effect. *PLoS ONE* 17, 4 ([n. d.]), 0267463. <https://doi.org/10.1371/journal.pone.0267463>
- [12] Monica Bulger and Patrick Davison. [n. d.]. The Promises, Challenges, and Futures of Media Literacy. *Journal of Media Literacy Education* 10(1) ([n. d.]), 1–21. <https://doi.org/10.23860/JMLE-2018-10-1-1>
- [13] Leonardo Bursztyn, Aakaash Rao, Christopher P. Roth, and David H. Yanagizawa-Drott. [n. d.]. Misinformation During a Pandemic National Bureau of Economic Research Working Papers. <https://doi.org/10.3386/w27417> DOI.
- [14] Carlos Carrasco-Farré. [n. d.]. The fingerprints of misinformation: how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. *Humanit Soc Sci Commun* 9 ([n. d.]), 162. <https://doi.org/10.1057/s41599-022-01174-9>

- [15] Sterling Crispin. [n. d.]. Data-Masks Biometric Surveillance Masks Evolving in the Gaze of the Technological Other. https://www.sterlingcrispin.com/Sterling_Crispin_Data-masks_MS_Thesis.pdf Accessed 02/02/23.
- [16] Bernd Debusmann, Jr. [n. d.]. Deepfake is the future of content creation. <https://www.bbc.co.uk/news/business-56278411> Accessed 15/03/23.
- [17] Russian Disinformation. [n. d.]. Deepfake of Zelensky tells Ukrainian troops to surrender. <https://www.youtube.com/watch?v=X17yrEV5s14>. Online at:
- [18] Marco Donnarumma, Wesley Goatley, and Helena Nikonole. [n. d.]. Critical art and the ethics of AI. <https://cryptpad.fr/pad/#/2/pad/view/H44naOgAhHBdcF2vb2HDKtCpWs0hV2sHML8yMKIp910/> Accessed 15/03/2023.
- [19] Arianna D'Ulizia, Maria Caschera Fernando Ferri, and Patrizia Grifoni. [n. d.]. Fake news detection: a survey of evaluation datasets. *PeerJ Comput Sci* 18, 7 ([n. d.]). <https://doi.org/10.7717/peerj-cs.518> DOI.
- [20] Lee Edwards, Mariya Stoilova, Nick Anstead, Andrew Fry, Gail El-Halaby, and Matthew Smith. [n. d.]. *Rapid Evidence Assessment on Online Misinformation and Media Literacy: Final Report for Ofcom* ([n. d.]). www.ofcom.org.uk. Online at:
- [21] Olafur Eliasson. [n. d.]. Why art has the power to change the world. <https://www.weforum.org/agenda/2016/01/why-art-has-the-power-to-change-the-world/> Accessed 15/03/23.
- [22] Julie Freeman, Geraint Wiggins, Gavin Stark, and Mark Sandler. [n. d.]. A Concise Taxonomy for Describing Data as an Art Material. *Leonardo* 21, 1 ([n. d.]).
- [23] Jennifer Gradecki and Derek Curry. [n. d.]. Crowd Sourced Intelligence Agency. <https://docbase.mit.edu/project/crowd-sourced-intelligence-agency/>
- [24] Lucas Graves. [n. d.]. Understanding the promise of automated fact checking. https://ora.ox.ac.uk/objects/uuid:f321ff43-05f0-4430-b978-f5f517b73b9b/download_file?file_format=application%2Fpdf&safe_filename=graves_factsheet_180226%2BFINAL.pdf&type_of_work=Report Accessed 15/03/23.
- [25] Dejan Grba. [n. d.]. Deep Else: A Critical Framework for AI Art. *Digital* 2, 1 ([n. d.]), 1–32. <https://doi.org/10.3390/digital2010001>
- [26] Saqib Hakak, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, and Wazir Zada Khan. [n. d.]. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems* 117 ([n. d.]), 47–58. <https://doi.org/10.1016/j.future.2020.11.022> DOI.
- [27] Drew Hemment, S.J. Bennett Morgan Currie, Jake Elwes, Anna Ridler, Caroline Sinders, Matjaz Vidmar, Robin Hill, and Holly Warner. [n. d.]. AI in the Public Eye: Investigating Public AI Literacy Through AI Art. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FACT '23)* (2023). Association for Computing Machinery, New York, NY, USA, 931–942. <https://doi.org/10.1145/3593013.3594052>
- [28] H.L.E.G. [n. d.]. A multi-dimensional approach to disinformation: report of the independent high-level group (HLEG) on fake news and online disinformation. https://blog.wan-ifra.org/sites/default/files/field_blog_entry_file/HLEGReportonFakeNewsandOnlineDisinformation.pdf Online at:
- [29] Peter Jandric. [n. d.]. The Postdigital Challenge of Critical Media Literacy. *The International Journal of Critical Media Literacy DOI* ([n. d.]). <https://doi.org/10.1163/25900110-00101002>
- [30] H.M. Johnson and C.M. Seifert. [n. d.]. Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, 6 ([n. d.]), 1420–1436. <https://doi.org/10.1037/0278-7393.20.6.1420>
- [31] Jonas Juul and Johan Ugander. [n. d.]. Comparing information diffusion mechanisms by matching on cascade size. In *In, Proceedings of the National Academy of Sciences* 118.46 e2100786118 (2021).
- [32] Eleni Kapantai, Androniki Christopoulou, Christos Berberidis, and Vasilios Peristeras. [n. d.]. A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society* 23(5) ([n. d.]), 1301–1326. <https://doi.org/10.1177/1461444820959296>
- [33] Sang Ah Kim. [n. d.]. Social Media Algorithms: Why you see what you see. In *2 GEO. L. TECH. REV.*, 147.
- [34] Chris Leeder. [n. d.]. How college students evaluate and share “fake news” stories.
- [35] Stephan Lewandowsky, John Cook, Ulrich Ecker, Dolores Albarracín, Michelle Amazeen, Panayiota Kendeou, Doug Lombardi, Eryn Newman, Gordon Pennycook, Ethan Porter, David Rand, David Rapp, Jason Reifler, Jon Roozbeek, Philipp Schmid, Colleen Seifert, and Gale Sinatra. [n. d.]. Briony Swire-Thompson, Sander van der Linden. <https://doi.org/10.17910/b7.1182> The Debunking Handbook 2020. Online at.
- [36] Stephan Lewandowsky and Sander Linden. [n. d.]. Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology* ([n. d.]). <https://doi.org/10.1080/10463283.2021.1876983> DOI.
- [37] Benjamin Lyons, Vittorio Merola, and Jason Reifler. [n. d.]. How bad is the fake news problem?: The role of baseline information in public perceptions. In *The Psychology of Fake News* Routledge ([n. d.]), 11–26.
- [38] Yisroel Mirsky and Wenke Lee. [n. d.]. The Creation and Detection of Deepfakes: A Survey. *ACM Comput. Surv* 54(1) ([n. d.]).
- [39] Fernando Miró-Llinares and Jesús C. Aguerri. [n. d.]. Misinformation about fake news: A systematic critical review of empirical studies on the phenomenon and its status as a ‘threat’. *European Journal of Criminology* 20(1) ([n. d.]). <https://doi.org/10.1177/1477370821994059> DOI.
- [40] Maria Molina, Shyam Sundar, Thai Le, and Dongwon Lee. [n. d.]. Fake News” Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist* 65(2) ([n. d.]), 180–212. <https://doi.org/10.1177/0002764219878224>
- [41] moondisaster.org. [n. d.].
- [42] Ryan Moore and Jeffrey Hancock. [n. d.]. A digital media literacy intervention for older adults improves resilience to fake news. *Sci Rep* 12 ([n. d.]), 6008. <https://doi.org/10.1038/s41598-022-08437-0>
- [43] Susan Morgan. [n. d.]. Fake news, disinformation, manipulation and online tactics to undermine democracy. *Journal of Cyber Policy* 3, 1 ([n. d.]), 39–43.
- [44] Jordan Peele. [n. d.]. Obama Deepfake. <https://ars.electonica.art/center/en/obama-deep-fake/> Accessed 02/07/23.
- [45] Tal Schuster, Roei Schuster, Darsh J. Shah, and Regina Barzilay. [n. d.]. The Limitations of Stylogometry for Detecting Machine-Generated Fake News. *Computational Linguistics* 46 (2) ([n. d.]), 499–510. https://doi.org/10.1162/coli_a_00380 DOI.
- [46] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. [n. d.]. Hoaxy: A Platform for Tracking Online Misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)* (2016), 745–750. <https://doi.org/10.1145/2872518.2890098>
- [47] David Silva, Chan Chen, and Ying Zhu. [n. d.]. Facets of algorithmic literacy: Information, experience, and individual factors predict attitudes toward algorithmic systems. *New Media and Society* ([n. d.]). <https://doi.org/10.1177/14614448221098> DOI.
- [48] Marita Skjive, Petter Bae Brandtzæg, Petter, and Asbjørn Følstad. [n. d.]. Why People Use ChatGPT. <https://doi.org/10.2139/ssrn.4376834> Preprint DOI.
- [49] Tom Southern. [n. d.]. The Art of Hitting Disinformation Where it Lives. <https://www.wired.com/story/disinformation-art-science/> Accessed 15/03/23.
- [50] Kate Starbird, Jim Maddock, Marnia Orand, Peg Achterman, and Robert M. Mason. [n. d.]. Rumors, False Flags. In *and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing In, Proceedings iConference 2014* (2014). <https://doi.org/10.9776/14308> DOI.
- [51] Luke Stark and Kate Crawford. [n. d.]. The Work of Art in the Age of Artificial Intelligence: What Art Can Tell Us About the Ethics of Data Practice. *Surveillance and Society* 17, 3 ([n. d.]). <https://doi.org/10.24908/ss.v17i3/4.10821> DOI.
- [52] startseu. [n. d.].
- [53] Edson C. Tandoc, Jr., Zheng Wei Lim, and Richard Ling. [n. d.]. Defining “Fake News”. *Digital Journalism* 6, 2 ([n. d.]), 13–53. <https://doi.org/10.1080/21670811.2017.1360143> DOI.
- [54] Kathie Treen, Hywel Williams, and Saffron O'Neill. [n. d.]. Online misinformation about climate change. *Wiley Interdisciplinary Reviews: Climate Change* 11, 5 ([n. d.]), 665. <https://doi.org/10.1002/wcc.665> DOI.
- [55] Teemu Valtonen, Matti Tedre, Kati Makitalo, and Henriikka Vartiainen. [n. d.]. Media Literacy Education in the Age of Machine Learning. *Journal of Media Literacy Education* 11, 2 ([n. d.]), 20–36.
- [56] Soroush Vosoughi, Deb Roy, and Siman Aral. [n. d.]. The spread of true and false news online. <https://doi.org/10.1126/science.aap9559> Science 1146-1151 DOI.
- [57] Emily Vraga and Leticia Bode. [n. d.]. Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation. *Political Communication* 37(1) ([n. d.]), 136–144. <https://doi.org/10.1080/10584609.2020.1716500>
- [58] Patricia Vries and Willem Schinkel. [n. d.]. Algorithmic anxiety: Masks and camouflage in artistic imaginaries of facial recognition algorithms. *Big Data & Society* 6(1) ([n. d.]), 20. <https://doi.org/10.1177/2053951719851532> DOI.
- [59] Nathan Walter, R.Lance Holbert Jonathan Cohen, and Yasmin Morag. [n. d.]. Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication* ([n. d.]). <https://doi.org/10.1080/10584609.2019.1668894>

APPENDIX A: AI ART PROJECTS IN MEDIAFUTURES 1ST (2021) AND 2ND (2022) COHORTS

| Artwork | Brief Description |
|-----------------|---|
| Soft Evidence | A series of synthetic scenes intended to be as believable as possible. In the exhibition, they mixed synthetic scenes with real scenes to invite the audience to guess which scenes are fake. This interactive guessing encouraged the audience to question the images they see, who created them and for what purpose. |
| Evil Magic | In this experiment the artists create a deep fake in real time with the face of the user. They demonstrated how, in a few seconds, software can capture the audience’s face and body gestures, and turn their images into a deep fake saying words that they would never have said. |
| Social Sandwich | Offers encounters with anonymous strangers with opposing views or values. The users are invited to collaborate with one another to determine the trustworthiness of the news that appears online. During a 15 minute message-based conversation, they see the world from another perspective, and discover how to keep exchanging when they do not agree with someone’s views. |
| Chanate Machine | Quantifies and reveals the mechanisms of misinformation on global warming. It is composed of two data sets. Fake news stories are categorised and the number of the category is displayed on the artwork. In the exhibition, correspondence is established between values and categories of detected false arguments. Confronted with a landscape of numbers, the visitors are invited to evaluate the quantities of each type of false argument. |

| | |
|---------------------------------|--|
| 730 Hours of Violence | An exhibition using data to explore the new paradigms of violence in the 21st century. Each piece is based on specific data sets, with the aim of engaging the audience with data and encouraging them to understand it rather than ignore it. |
| Two Truths and a Lie | A multimedia installation exploring the relationship between foreign languages, mother tongues and trustworthiness. In the age of disinformation, distributing credible information is an increasingly complex challenge, but what exactly makes individuals ‘deem’ an information source trustworthy? Combining experimental documentary, video art, spatial audio, and assemblage, the 100-channel and 100-language installation uses the children’s game of ‘two truths and a lie’ to blend truth and fiction beyond the point of discernibility. |
| Invisible Voice | A browser plugin that empowers individuals to make informed decisions about the websites and companies that they use. It opens a pop-up, that displays data to enable the user to make informed decisions based on the company’s practices, derived from 17 databases. |
| Edit Wars | The project addresses the use of aggressive narratives in the government-controlled media that isolate public perception from the real state of affairs. Data from large datasets are analysed quantitatively and qualitatively to draw meaningful conclusions for the presentation, and the findings displayed in a multimedia interactive medium. |
| Synthetic Identity Speculations | A participatory artistic research project that monitors individual synergy effects of social network algorithms and their impact on body images. Transgressing platforms through hyperrealistic ideals and potentially momentous misclassifications, accountability for algorithmic agency is effectively shifted to users. |
| How Many Heartbeats to | Offers a new narrative about energy consumption and digital data through an interactive experience engaging our own body energy. This artwork intends to create awareness on digital pollution caused by |

| | |
|----------------------|---|
| Send A Love Email | infobesity and fake news, and to question our intimate relationship with online data. |
| Bibliograph | Bibliograph combines two micro tools for collective linked data aggregation, text annotation and voice recording. The resulting semantic layer allows greater engagement with texts. We are proposing the use of this tools in a non-technical digital environment suited for independent research, autonomy and digital literacy. |
| HyperViz | This immersive prototype turns hyperspectral data satellite data from a wide variety of sectors including environmental management, agriculture and pandemics into a digital experience for the general public in a way that raw data cannot be experienced. |
| Fragile Perspectives | A multi-sensory experience of news landscapes, information distortion and the fragility of perspectives formed by unbalanced news consumption. The audience can go further with the tool Ject-ai which was used to analyse news |
| Ponte | An online tool which allows discussions starting with an illustration. The artwork-based discussion launches the participants into a creative narrative mode. The metaphors and abstractions at the center of arts allow people to contrast perspectives without devaluing an opinion or attributing blame for being wrong. The inputs from those discussions with people around the world are then presented back in illustrations to the participants in an exciting way. |
| MUMIDIS | The Museum of mis- and disinformation» educates people about methods of brainwashing with disinformation. Using realistic visuals and audio, the audience is engaged in a gamified way to guess false and trusted news. Emotion AI technology, detects emotions & behavior of online visitors and compares the emotions of readers, while they consume news. |
| The Oracle Network | The Oracle Network is composed of two main parts: urban augmented reality art spread around the city that leads, like a |

| | |
|----------------------|---|
| | treasure hunt, to the Central Hub of interactive artificial intelligence art installations. The Central Hub is a private space where visitors interact with artificial intelligence art installations. The installations are on three levels of virtualization to gradually introduce the viewer into the abstract tech space of fake media. |
| Time-lapse Migration | A digital storytelling and exhibitions platform on the web and on the metaverse to disseminate first hand forced migration narratives and to give visibility and new market opportunities to refugee artist's artwork. This counters misinformation and disinformation on refugee phenomenon in Europe, by giving a wider context and different lights on the migration fluxes, their motivations, problems but also past episodes of forced migration in Europe and their contributions to European culture. |
| Doppelgänger | Doppelgänger aims to educate the audience on understanding of the laws and practices surrounding CCTV and draw into question the neutrality and trustworthiness of images recorded by a machine in an age where deep fake technology is widespread. |

APPENDIX B: INTERVIEW GUIDE

The first line in each question is the main question, with the following questions used as prompts or followups as necessary.

1. Can you give us a brief overview of what you set out to do in the MediaFutures support programme (MediaFutures for short), and what you ended up doing?

How is your project connected to misinformation?

2. What were your expectations going into MediaFutures?

What did you hope to learn?

What did you think are going to be the biggest challenges?

3. What was your actual experience of participating in MediaFutures?

What challenges did you face during your work in MediaFutures? Did you use any of the MediaFutures resources, like datasets, infrastructure, tools, or the toolkit?

4. What types of support did you seek outside of MediaFutures? Were there types of support that you could not find either within or outside of MediaFutures?

5. What role did citizens/users engagement play in your project? Did you engage citizens/users in the design or testing of your product/artwork?

Did you support interaction among citizens/users? How would you describe that interaction and their effect on participants?

Did you change or improve your citizen engagement practices as a result of MediaFutures, for example in terms of ethics?

6. What do you think is the impact of your project on your users?

Do you think it improved or is able to improve their knowledge on a specific subject?

Did or could it have an impact on their media literacy, how they understand, access, use media?

Did you notice or do you think it could improve users' soft skills such as interpersonal communication, collaboration, problem solving and critical thinking?

7. Do you think that MediaFutures had an impact on your visibility and recognition in your sector? How?

Did you observe an increase in your followers and/or interactions on social media?

8. How has participating in MediaFutures changed your use of data? Did you use new or different data sources? Which?

Did you change your data management practices? How?

What were your biggest challenges around use of data?

9. How relevant was interdisciplinarity and/or an intersectoral approach for your project? How did that surface?

10. (Start up/Artist only)

What do you think is the value of start-up/artist collaborations?

11. What was it like to work with artists/start-ups on this project?

12. Will you continue working with this artist/start-up after MediaFutures?

13. How did the physical distance impact the co-creation process?
AI Art and Misinformation

14. What would you recommend to others attempting such collaborations?

(Artists only)

10. Have you consulted or collaborated with scientists or technologists (internal or external) to carry out the project?

11. What were the fields of expertise that you needed to carry out the project?

12. How did you establish that collaboration? Has MediaFutures helped you to do that? Why not?

14. Which are the main challenges and opportunities of scientist/technologist and artist collaboration?

15. Which are the crucial elements of a successful artist / scientist /technologist partnership

16. What have you learned during MediaFutures? Have you changed the way you work? What is your most important take-away? What do you wish you had known before starting in MediaFutures? What did you learn about misinformation during your project?

17. What was your biggest success within MediaFutures and what did this mean for your project and for you as an artist/entrepreneur?

18. What are your plans after the end of MediaFutures?

(Follow-up questions via email)

1. What data set did you use? Please provide links to open, or a brief description for closed datasets. If the datasets are associated with a project or paper, references would be great.

2. What data tools did you use? Please provide a list of any tools you used during the project, and a brief description of what you used them for (where appropriate). This could include high-level tools for data processing (cleaning, modelling, visualization, etc.), but also data science frameworks, libraries, and models. We are especially interested in models that deal with disinformation or other core applications of projects (metaverse, community detection, cascade effects, XAI or algorithmic fairness techniques, etc.).

3. What data, code, or other outputs of your project did you or would like to publish? Please provide a link to any published resources, and some details about any you would like to publish, including whether you would like our support in doing so.