Guidebook for Teachers

10

DISINFORMATION TYPES

USED IN SOCIAL MEDIA

Chryssanthopoulou, K., Gotsi, I., López, B., Neuvonen, M., Ražinskaitė, R., Salo, M., Varanauskas, A., Zinkevičiūtė, G.

European Commission

This project is co-funded by the European Commission under the preparatory action "Media Literacy for All 2018".

CONTENTS:

INTRODUCTION

This document is intended to familiarise teachers with ten types of disinformation which are usually encountered in social media. It provides dossier on each selected disinformation types in which it's level of deception, working principle, examples, checking methods and other relevant information is presented.

These dossiers provide only essential information related to concrete disinformation type because this document is intended to be a first stepping stone for teachers on the pathway to media and information literacy avenue. Under each disinformation type or at the end of the document reader may find links to other useful materials which we wholeheartedly encourage to read and use as this one as well.

Ten types of disinformation in social media were selected through an iterative process. In the beginning analysis of existing material and research were analysed and long-list of the disinformation types were developed. Afterwards, experts from each partner's done mapping and prioritising activities and selected 15 disinformation types to work further. In the last step, teacher and student feedback were collected, and it allowed to narrow down the list to precisely ten disinformation types used in social media.

This educational resource can be used as a standalone material (together with prepared slides and poster-type visuals). Yet, much great material by other organisations is also freely available. Thus consortium partner strongly encourages to use them in order to provide the best experience for one student. Especially we commend European Association for Viewers Interest (EAVI) developed ten types of misleading news: https://eavi.eu/beyond-fake-news-10-types-misleading-info/ and First Draft identified misinformation and disinformation types: https://firstdraftnews.org/fake-

news-complicated/). You will find more thoughts for inspiration at the end of the document.

Moreover, under this project (gaMEdia (co-funded by the European Commission under the preparatory action "Media Literacy for All 2018", LC-01234900) card game for 12-15-year-old students will be developed and should be used as a complementary tool for media and information literacy education. More information about the project and deliverables is accessible on www.checkorcheat.eu.

Before proceeding to read the information on ten selected disinformation types, it is necessary to clarify a few concepts as they sometimes by different actors are used differently. In this project, we use the following definition and distinctions between misinformation, disinformation and malinformation:

Misinformation:

When false information is shared, but no harm is meant;

Disinformation:

When false information is knowingly shared to cause harm;

Malinformation:

When genuine information is shared to cause harm.

Some topics in this document are not presented in that extent it would require to do so. Nevertheless, they are left out not because they are unimportant; on the contrary, because of their importance, they would require separate publication. One of the topics is misleading YouTube-content. Kids of all ages watch many videos, and due to suggestion algorithms, they are widely exposed to Crazy conspiracy theories and fringe ideas, nutcase health claims, radicalised content (hate speech). Topics mentioned above are covered in this document, but no attention is given to" weaponisation" of the comment sections, which is done usually by alt-right groups. Another not presented topic is PsyOps (stands for psychological operations: military operations usually aimed at influencing the enemy's state of mind through non-combative means (such as the distribution of leaflets) which is very complex and would require extensive research and focus on technicalities. The last topic which we feel needs mentioning is cyber-bullying because it is very widely spread problem and many students face it in an online environment. Many techniques are very similar to those used in disinformation (shareability to spread rumours, manipulated images and videos, nudes, use of funny memes). We do believe that cyber-bullying is quite extensively covered the topic in contemporary education discussion, and many authors have done it better than we could do.

Before you go and read about disinformation types, here is a small

reminder of how each of us can contribute to not spreading wrong or wrong-intended information.

1. Don't just read the headline. Read the whole story before you share it — including the name of the author and the date it was published. If the story is fake, that's where you'll find your first clues — especially if the headline doesn't match the content.

2. Check to see if other outlets are reporting on the same story. If only one outlet is reporting a story, there might be a good reason — it could be an exclusive or something based on leaked information or sources who request anonymity. In such cases, other media outlets are usually quick to try to verify the story or will cover reaction to the story, which gives you more confidence that the story is true. But if no mainstream outlets have picked up on the story, and you only see it on blogs or niche outlets, wait to share the story or video.

3. Do a quick Google search. You may learn new details, or find out that the story was published five years ago, or that the headline is false.

4. When people tell you something or pass on something that seems like it could be false, **ask them more about where they heard, saw or read it**.

If in doubt, don't share it. It never hurts not to share something, but it can be damaging to share something that isn't true.

Check or Cheat is part of the project gaMEdia – My Media Literacy Class, carried out by four partners:

DEEPFAKE

Who is able to produce:
Professional – **Amateur** – Anyone.

However, technology is galloping and in the near future predicted that anyone would be able to produce it. There are applications available that users can download and start experimenting.

Level of deception:
Low – Average – High – **Very high**.

Detecting deepfakes is a challenging problem. Amateurish deepfakes sometimes can be detected by the naked eye, but deepfakes are getting better all the time, and soon we will have to rely on digital forensics to detect deepfakes — if we can detect them at all.

Deepfake is an AI-based technology used to produce or alter video content by editing faces (face-swapping or creating new face expressions). First deepfakes were created by changing people faces in videos by celebrity faces, particularly in pornographic video clips. It was done in December 2017 by Reddit user known as deepfakes (a portmanteau of "deep learning" and "fake" after whom this type was named) who used deep learning technology to edit the faces. For spreading false information, deep learning technology is used to create a new facial expression of celebrities which simulates facial muscle movements to represent saying fabricated text which was newer told by that person.

Working principle (what and how does it do):

Deepfake video is created by using two competing AI systems - one is called the generator, and the other is called the discriminator. The generator creates a

fake video clip and then asks the discriminator to determine whether the clip is real or fake. Each time the discriminator accurately identifies a video clip as being fake, it gives the generator a clue about what not to do when creating the next clip.

As the generator gets better at creating fake video clips, the discriminator gets better at spotting them. Conversely, as the discriminator gets better at spotting fake video, the generator gets better at creating them.

Together, the generator and discriminator form something called a *generative adversarial network (GEN)*. The first step in establishing a GAN is to identify the desired output and create a training dataset for the generator. Once the generator begins creating an acceptable level of output, video clips can be fed to the discriminator.

Source: https://whatis.techtarget.com/definition/deepfake.

Example:

Many examples (some fitting for education as well) can be found in YouTube channel TheFakening: https://www.youtube.com/c/TheFakening/videos.

Gizmodo article about deepfake: https://gizmodo.com/insanely-accurate-lip-synching-tech-could-turn-fake-new-1796843610.

Synthesising Obama video: https://www.youtube.com/watch?time_continue=62&v=MVBe6_o4cMI&feature=emb_logo.

Checking method:

If deepfake is not professional one can spot that shadows do not fall as they should be falling, or the person is not blinking. But if deepfake is of higher quality, there is no way to recognise it using your eyes. Many firms are trying to develop software which could help to identify deepfakes: https://techcrunch.com/2020/09/14/sentinel-loads-up-with-1-35m-in-the-deepfake-detection-arms-race/. The US military is also funding an effort to catch deepfakes: https://www.technologyreview.com/s/611146/the-us-military-is-funding-an-effort-to-catch-deepfakes-and-other-ai-trickery/.

PSEUDOSCIENCE

Who is able to produce:
Professional – Amateur – **Anyone**.

It is very easy to spread pseudoscience, and it does not require any skills. But the creation of pseudoscience theory is way harder than one can expect.

Level of deception:
Low – Average – **High** – Very high.

Promoters of pseudoscience often adopt the vocabulary of science, describing conjectures as hypotheses, theories, or laws, providing "evidence" from observation and "expert" testimonies, or even developing what appear to be mathematical models of their ideas.
Because of that, it might be hard to tell is information legit or not, especially without additional verification.

Pseudoscience consists of statements, beliefs, or practices that are claimed to be both scientific and factual but are incompatible with the scientific method.

Pseudoscience is often characterised by contradictory, exaggerated or unfalsifiable claims; reliance on confirmation bias rather than rigorous attempts to disproof; lack of openness to evaluation by other experts; absence of systematic practices when developing hypotheses; and continued adherence long after the pseudoscientific hypotheses have been experimentally discredited.

Source:
https://www.youtube.com/watch?v=-X8Xfl0JdTQ.

Working principle (what and how does it do):

Pseudoscience suggests something is being presented as science inaccurately or even deceptively.

When something is popular, yet wrong, it can often become an established "fact" merely through virtue of being repeated so many times. Sometimes this misinformation is due to popular science fiction/fantasy which either is based on old obsolete concepts or just plain poor present science.

Pseudoscientific claims very rarely have specific, testable scientific predictions, and instead rely on vague and ambiguous language, often encompassing grandiose claims.

In quack medicine a pseudoscience promoter might claim a given treatment "removes toxins from your system", never saying what toxins, or how they will be removed, or how you can tell if they have been removed. The toxins are the real cause of disease, never saying how they cause disease, and that removing them will cure you of all known afflictions.

Lists of pseudoscience examples: https://examples.yourdictionary.com/examples-of-pseudoscience.html and https://rationalwiki.org/wiki/List_of_pseudosciences.

Checking method:

The easiest way to distinguish the pseudoscientific method from the scientific method is to look at whether there are testable predictions, and see whether the experiments set out to test the theory or simply to confirm it.

It might be helpful to distinguish science and pseudoscience

Table 1).

Two videos on spotting pseudoscience: https://www.youtube.com/watch?v=gaDvroATyjw, https://www.youtube.com/watch?v=h-agrL1gS4c.

Example:

Table 1 *Differences between science and pseudoscience.*

SCIENCE	PSEUDOSCIENCE
The primary goal of science is to achieve a more complete and more unified understanding of the physical world.	Pseudosciences are more likely to be driven by ideological, cultural, or commercial goals. Some examples: astrology (from ancient Babylonian culture,) UFO-ology (popular culture and mistrust of government), Creation Science (attempt to justify a literal interpretation of the Bible), "structure-altered" waters (commercial quackery.)
Most scientific fields are the subjects of intense research which result in the	The field has evolved very little since it was first established. The small amount of research and experimentation that is carried

continual expansion of knowledge in the discipline.	out is generally done more to justify the belief than to extend it.
Workers in the field commonly seek out counterexamples or findings that appear to be inconsistent with accepted theories.	In the pseudosciences, a challenge to accepted dogma is often considered a hostile act if not heresy, and leads to bitter disputes or even schisms.
Observations or data that are not consistent with current scientific understanding, once shown to be credible, generate intense interest among scientists and stimulate additional studies.	Observations or data that are not consistent with established beliefs tend to be ignored or actively suppressed.

Source of the table: http://www.chem1.com/acad/sci/pseudosci.html.

MANIPULATED CONTENT

Who is able to produce:
Professional – Amateur – **Anyone**.

People share millions of photos and videos on social media every day. Some of that content is manipulated, often for benign reasons, like making a video sharper or audio clearer. But some people engage in media manipulation in order to mislead. Manipulations can be made through simple technology like Photoshop or through sophisticated tools that use artificial intelligence or "deep learning" techniques to create videos that distort reality – usually called "deepfakes" (deepfakes are presented separately in this analysis).

Level of deception:
Low – Average – **High** – Very high.

Identity is becoming harder to validate, as trolls and bots adopt new ways to mask their real source. Video editing, manipulated images are much harder to detect, compared to textual disinformation.
Platform companies are coming under increased pressure and scrutiny to respond. Some of them take action and removes such content. But even in best circumstances, it takes time during which harm already is done.

"Manipulated content" is when an aspect of genuine content is altered, relating most often to photos or videos. Visual media can be transformed through photo manipulation, commonly called "photoshopping". Video manipulation targets digital video using a combination of traditional video processing and video editing techniques and auxiliary methods from artificial intelligence like face recognition.

Working principle (what and how does it do):

New techniques to modify images, audio, and video enable the creation of manipulated content. "Photoshopping" can make a product, person, or idea seem more appealing. It is done by highlighting certain features. Also, other techniques, such as framing (showing only part of the picture, taking it out of context) can be used to distort reality as well.

In typical video manipulation, the facial structure, body movements, and voice of the subject are replicated in order to create a fabricated recording of the subject. The applications of these methods range from educational videos to videos aimed at (mass) manipulation and propaganda, a straightforward extension of the long-standing possibilities of photo manipulation.

Example:

Picture manipulation example: https://spotlightstories.co/32-examples-media-manipulating-truth/.

The Washington Post guide on manipulated video: https://www.washingtonpost.com/graphics/2019/politics/fact-checker/manipulated-video-guide/.

Checking method:

How to verify viral social media videos: https://www.youtube.com/watch?v=e91IGj_apsY.

Online tools to verify photo: https://www.stopfake.org/en/13-online-tools-that-help-to-verify-the-authenticity-of-a-photo/.

CONSPIRACY THEORY (FRINGE SCIENCE)

Who is able to produce:
Professional – Amateur – Anyone.

Most people are consumers rather than producers of conspiracy theories. They don't come up with their own conspiracy theories but endorse those that are already in circulation.

Level of deception:
Low – Average – **High** – Very high.

Belief in conspiracy theories is generally based not on evidence, but in the faith of the believer. Conspiracy theory conversely posits the existence of secretive coalitions of individuals and speculates on their alleged activities which might be hard to disprove.

A conspiracy theory is an explanation of an event or situation that invokes a conspiracy by sinister and powerful actors, often political in motivation when other explanations are more probable. However, unlike pseudoscience, fringe science conducts itself using the scientific method. The ideas studied by fringe scientists do not receive mainstream support.

Conspiracy beliefs have the potential to cause harm both to the individual and the community. Conspiracy endorsement is associated with lowered intention to participate in social and political causes, unwillingness to follow authoritative medical advice, increased willingness to seek alternative medicine, and a tendency to reject critical scientific findings.

The origin of countless conspiracy theories:

https://www.youtube.com/watch?v=88_C-fogY40.

Working principle (what and how does it do):

Conspiracy theories are widely present on the web in the form of blogs and YouTube videos, as well as on social media.

Conspiracy theories are first and foremost forms of political propaganda. They are designed to denigrate specific individuals or groups or advance a political agenda. The theory that the Clintons were somehow involved in the Epstein suicide denigrates the Clintons. The notion that the US government staged the 2012 mass shooting at Sandy Hook Elementary School helped the pro-gun lobby to deflect arguments for greater gun control. What better way to pre-empt calls for greater gun control in the wake of a school shooting than to claim that it never happened?

Conspiracy theories appear to provide broad, internally consistent explanations that allow people to preserve beliefs in the face of uncertainty and contradiction.

Example:

A conspiracy theory may take any matter as its subject, but certain subjects attract greater interest than others. Favoured subjects include famous deaths and assassinations, morally dubious government activities, suppressed technologies, and "false flag" terrorism (pinning the blame on a second party).

Among the longest-standing and most widely recognised conspiracy theories are notions concerning the assassination of John F. Kennedy, the 1969 Apollo moon landings, and the 9/11 terrorist attacks, as well as numerous theories about alleged plots for world domination by various groups both real and imaginary.

WIRED list of articles on different conspiracies (up-to-date): https://www.wired.com/tag/conspiracy-theories/page/1/.

Five fact-checked tech conspiracies: https://www.businessinsider.com/facebook-microphone-listening-for-ads-other-tech-conspiracy-theories-explained-2019-9.

YouTube video on different conspiracies: https://www.youtube.com/watch?v=_53cGxAUuDk.

Fact check: A guide to 9 conspiracy theories Trump is currently pushing: https://edition.cnn.com/2020/09/02/politics/fact-check-trump-conspiracy-theories-biden-covid-thugs-plane/index.html.

Checking method:

List of fact-checking websites: https://en.wikipedia.org/wiki/List_of_fact-checking_websites.

Worth-while EU funded initiative oriented at disinformation (including conspiracy theories): https://euvsdisinfo.eu/ .

HOAX
(URBAN LEGENDS AND CHAIN EMAILS)

Who is able to produce:
Professional – Amateur – **Anyone**.

Everyone can perpetrate a hoax by making only factual statements using unfamiliar wording or context, such as in the dihydrogen monoxide hoax (dihydrogen monoxide is water. Example: "The atomic components of DHMO are found in a number of caustic, explosive and poisonous compounds such as Sulfuric Acid, Nitroglycerine and Ethyl Alcohol "- http://www.dhmo.org/facts.html).

Level of deception:
Low – **Average** – High – Very high.

Hoaxes have staying power because of the peculiar way people process information and arrive at beliefs. When confronted with new information, humans don't always do the logical thing and evaluate it on its own merits. Instead, we often make snap decisions based on how the information adheres with our existing worldviews. Although small google research through fact-checkers internet pages might show quite fast if a piece of information is a hoax.

A hoax is a falsehood deliberately fabricated to masquerade as the truth.

A common aspect that hoaxes have is that they are all meant to deceive or lie. For something to become a hoax, the lie must have something more to offer. It must be outrageous, dramatic, but also has to be believable and ingenious. Above all, it must be able to attract attention from the public.

Fake news (also referred to as hoax news) deliberately publish hoaxes which may serve the goal of propaganda or disinformation — using social media to drive web traffic and amplify their effect.

The urban legend is a modern genre of folklore. It often consists of fictional stories associated with the macabre, superstitions, cryptids, creepypasta, and other fear generating narrative elements.

Urban legends are often rooted in local history and popular culture.

A chain letter is a message that attempts to convince the recipient to make some copies and pass them on to a certain number of recipients (same applied to emails).

Working principle (what and how does it do):

As mentioned hoax is false or half-truth information which is presented as accurate and factual with the intention to deceive other persons. Usually, it is "sensational", so it helps for it to spread because people tend to not check information reliability before sharing and liking.

Example:

When a newspaper or the news reports a fake story, it is known as a hoax. Misleading public stunts, scientific frauds, false bomb threats and business scams are examples of hoaxes.

One of the earliest recorded media hoaxes is a fake almanac published by Jonathan Swift under the pseudonym of Isaac Bickerstaff in 1708. Swift predicted the death of John Partridge, one of the leading astrologers in England at that time, in the almanac and later issued an elegy on the day Partridge was supposed to have died. Partridge's reputation was damaged as a result, and his astrological almanac was not published for the next six years.

Few contemporary examples: https://www.mentalfloss.com/article/49674/14-greatest-hoaxes-all-time.

Checking method:

List of fact-checking websites: https://en.wikipedia.org/wiki/List_of_fact-checking_websites.

Worth-while EU funded initiative oriented at disinformation (including hoaxes): https://euvsdisinfo.eu/.

CLICKBAIT

Who is able to produce:
Professional – **Amateur** – Anyone.

For sites that thrive on thousands of click-throughs to content, many authors see the use of clickbait as a means to tap into the human psyche by crafting these eye-catching headlines. Sometimes clickbaiting is also used by journalists. Amateurs can produce good clickbaits occasionally, but great and consistent clickbaiting requires professional skills.

Level of deception:
Low – Average – **High** – Very high.

"Clickbait" has become a dominant form of online media, with headlines designed to entice people to click becoming the norm. Resisting clickbait is problematic because it is exploiting the neural circuitry that evolved over millions of years. Our brains were not designed to be exposed to the variety of temptations that are found in this hyper-connected world.
A more concerning form of clickbait is one that appeals directly to people's fears, especially as it relates to a threat to a social group to which they belong - emotional clickbait. This form of clickbait serves the twin purposes of inducing excitement by appealing to group competition and being easily spread among online social networks.

Clickbait is a form of false advertisement which uses hyperlink text or a thumbnail link that is designed to attract attention and entice users to follow that link and read, view, or listen to the linked piece of online content, with a defining characteristic of being deceptive, typically sensationalised or misleading (source: https://www.cyber.gov.au/acsc/view-all-content/glossary/clickbait). Clickbait as the effect is also sometimes seen with journalistic headlines which exaggerate or scandalise content.

In some cases, clickbait is simply used to generate income; more clicks means more money made with advertisers. But these headlines and articles can also be used to influence a group of people on social media. They are constructed to appeal to the interest group's pre-existing biases and thus to be shared within filter bubbles.

Working principle (what and how does it do):

A "teaser" aims to exploit the "curiosity gap", providing just enough information to make readers of news websites curious, but not enough to satisfy their curiosity without clicking through to the linked content. Clickbait headlines add an element of dishonesty, using enticements that do not accurately reflect the content being delivered. The "-bait" part of the term is used in analogy to fishing, where a hook is disguised by an enticement (bait), presenting the impression to the fish that it is a desirable thing to swallow.

Sometimes clickbait is more like bait and switch. That is, we read a catchy headline or link, click it, only to find ourselves enveloped within an ad. There is content when we click on the link, but it is heavily wrapped in advertisements. Thus, the article or video is in actuality a lure that exposes us to the ad, which is the true purpose of the content. When enough people are exposed to the ads, there will be a percentage who become buyers. It works because elsewhere; it would not be used so widely.

Source:
https://www.youtube.com/watch?v=qskqM9O0FC0 .

Example:

https://adespresso.com/blog/clickbait-facebook-advertising-examples/

https://www.bluleadz.com/blog/the-scientific-reasons-why-clickbait-actually-works

https://www.reputationx.com/orm/techniques/process/content/orm-guest-posts/click-bait

https://medium.com/zerone-magazine/you-wont-believe-how-these-9-shocking-clickbaits-work-number-8-is-a-killer-4cb2ceded8b6

Checking method:

Tools have been developed to address the clickbait problem. Clickbait detection has been integrated with browser applications while digital platforms where contents are shared, such as Twitter have updated their respective algorithms to filter clickbait contents. Social media groups, such as Stop Clickbait, combat clickbait by giving a summary of the clickbait article, closing the "curiosity gap". The research community has also developed clickbait reporting browser plug-ins in order to report clickbait links for further advances in the field based on supervised learning algorithms.

Here are a few tips that might help to resists clickbait:

1. Think of strategies outside of when the problem is happening. Come up with some ideas on how to resist clickbaiting when the problem is not happening. Put some of those ideas into place and assess the results. Start with the simplest, easiest to implement strategies.

Sometimes even small changes yield significant returns.

2. Notice your patterns and replace them with more adaptive ones. Perhaps through a little data collection, you realise that you tend to go down the YouTube wormhole (arguably, a subtype of clickbait) at work later in the day. What purpose is it serving? Maybe you need a break? Is there something else you can do to relieve boredom or angst?

3. Consider using some website blocking tools. Many tools can help save ourselves from ourselves. For instance, if we keep checking a particular website for news updates (and get hooked by clickbait), we can install a tool that will limit our access to those tempting sites during periods that we define.

YouTube video on how to spot clickbait: https://www.youtube.com/watch?v=8IzfzoZsa-Q.

ADVERTISING (NATIVE, POLITICAL, MISLEADING, INFLUENCER MARKETING)

Who is able to produce:
Professional – Amateur – Anyone.

Successful advertisement requires financial investment and often uses sophisticated techniques. Certain skills are needed to create the visual or audio content of the ad. There is many research data in the fields of neuroscience, psychology and data analysis.
It's important to note that with several free or paid advertisement making tools and easy to use social media advertisement and targeting platforms the process of making ads is becoming easier and more available to the broader audience.

Level of deception:
Low – **Average** – High – Very high.

Most ads are marked as sponsored content or placed in a way that lets the consumer know that the media is an advertisement. Some forms of advertising (native and influencer advertisement) are more challenging to recognise. Even when identified, claims, presented in ads, can be quite misleading.
Laws on advertisement limit the level of manipulation; however, there are several different methods for attempting to deceive consumers that are not permitted under advertising law.

Advertising is a marketing tactic involving paying for space to promote a product, service, or cause. The actual promotional messages are called advertisements or ads for short. The goal of advertising is to reach people most likely to be willing to

pay for a company's products or services and entice them to buy.

Advertisements can be placed nearly anywhere, including roadside billboards, sides of buildings, websites, electronic newsletters, print newsletters, inside bills, product packaging, restaurant placemats, event bulletins, store windows, the sides of cars and trucks, subway car walls, airport kiosks, sporting arenas, YouTube videos and many more.

Working principle (what and how does it do):

Marketers and advertisers have a surplus of tools to help nudge, persuade and even influence a person's buying habits. From the classics like data derived from demographic, geographic and ethnographic sources to more forward-thinking solutions like facial recognition, body language biometrics or microtargeting based on psychographic information (see for more info: https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie).

Advertisement can use deception by photobleaching (portray false and unobtainable results), omitting information, having hidden fees and surcharges, manipulation of measurement units and standards, fillers and oversized packaging, misleading health claims.

Ads can also exaggerate a product's worth through the use of meaningless unsubstantiated terms, based on opinion rather than fact, and in some cases through the manipulation of data.

Microtargeting is a powerful advertising tool that allows aiming ads at specific groups of people, or even individuals, online. For example, it permits politicians to target very narrow groups of voters with tailored messages that have the potential to manipulate the political debate. Here's how it works: Political campaigns create databases about voters that include information about whether a person is registered to vote, how often they vote, their party affiliation, their mailing address, their email address, and their phone number. They can then upload those voter files to Google and Facebook to find those people's online profiles, and then advertise specifically to them (source: https://www.vox.com/recode/2019/11/27/20977988/google-facebook-political-ads-targeting-twitter-disinformation).

Bait and switch a sales tactic in which a customer is attracted by the advertisement of a low-priced item but is then encouraged to buy a higher-priced one. Or the ploy of offering a person something desirable to gain favour (such as political support) then thwarting expectations with something less desirable (source: https://www.merriam-webster.com/dictionary/bait%20and%20switch).

Native marketing is the use of paid ads that match the look, feel and function of the media format in which they appear. The word "native" refers to this coherence of the content with the other media that appear on the platform. These ads that blend most easily into digital content and are harder to identify as ads (source: https://www.outbrain.com/native-advertising/).

Influencer marketing is a type of social media marketing that uses endorsements made by people, organisations, and groups seen as influential or experts in a particular area (source: https://entrepreneurship.babson.edu/what-is-influencer-marketing/).

Advertising aims to present a product in the best light possible. There is some leeway in the creative process. The problem arises when the dramatisation crosses the line into falsely representing a product.

Political advertising attempts to influence or comment upon a matter which is currently the subject of extensive political debate (source: https://adstandards.com.au/issues/political-and-election-advertising).

Widespread concern exists about the potential effects that media portrayals of drinking, alcohol product placements and alcohol advertising may have on alcohol consumption and problems among young people. Television, radio, film, and popular music are often identified as potential sources through which young people learn about alcohol and as possible influences on young people's drinking and drinking problems.

Example:

Ribena-maker fined $217,500 for misleading vitamin C ads.

Airbrushed make-up ads banned for 'misleading'.

Climbing rope not suitable for climbing.

General election 2019: Ads are 'indecent, dishonest and untruthful'.

Influencer marketing:

- TikTok
- Instagram
- Instagram (2)

Native advertisement:

https://www.paldesk.com/spot-native-advertising/

https://www.wordstream.com/blog/ws/2014/07/07/native-advertising-examples

Checking method:

Facebook Political Ad Collector: This tool shows users the advertisements on their Facebook feeds and guesses which ones are political. It also shows users political advertisements aimed at other users. All political ads that are collected are put into a publicly available database.

Who Targets Me: This tool allows users to create an anonymous profile, then collect information about the political and other ads that they see, along with details about why they were targeted with those

ads. The tool can provide users with statistics on who/what has been targeting them and uses this information to build a database of political advertising and targeting.

TV News Fact Check: The TV News Archive is an initiative developing an archive of digital media, ranging from web pages, books and texts, audio recordings, videos, images, and software programs. One of its projects, the "Political TV Ad Archive" is an archive of 2016 political ads combined with fact-checking from a variety of sources (e.g. Politifact, Factcheck.com).

SATIRE (PARODY)

Who is able to produce:
Professional – Amateur – Anyone.

Although great humour and satire often feel effortless to the reader, for the writer it takes effort and practice and requires care and revision, especially as satire is said to be one of the most challenging types of humour to write. Typically, in order to take a serious subject and make a serious comment on it in a way that comes across with a note of humour, the author has to succeed in a couple of areas: he or she must be intelligent, well-read and informed, and relevant.

Level of deception:
Low – Average – High – Very high.

Satire should not be deceptive – when creating a satire piece, one seeks to make it, so the reader understands that this is satire. However, there have been numerous cases when even governments, politicians, mainstream media, or news outlets are fooled by satire and represent it as credible news.

Satire: the use of humour, irony, exaggeration, or ridicule to expose and criticise people's stupidity or vices, particularly in the context of contemporary politics and other topical issues.

Parody is a form of satire that exaggerates the notable features of a public figure, artist, or genre, intentionally copies the style of someone famous or copies a particular situation, making the characteristics or qualities of the original more noticeable in a humorous way.

Satiric comedy ridicules policies or philosophical doctrines or else attacks deviations from the social order by making ridiculous the violators of its standards of morals or manners.

Irony describes situations that are strange or funny because things happen in a way that seems to be the opposite of what you expected (the difference between what is said or done, and what is meant).

Working principle (what and how does it do):

Satire is a powerful art form which can point out the deficiencies in certain human behaviours and the social issues which result from them in such a way that they become absurd, even hilarious, which is therefore entertaining and reaches a broad audience. Satire also can protect its creator from culpability for criticism because it is implied rather than overtly stated; in this way, it becomes a powerful tool for dissenters in difficult or oppressive political and social periods.

Satire has endured as a storytelling technique for centuries because it offers a brilliant mix of comedic relief and social critique. It combines entertainment with a purpose.

Source: https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1065&context=srhonorsprog.

Tools of satire:

1. Exaggeration: hyperbole or understatement. To enlarge, increase, or represent something beyond normal bounds so that it becomes ridiculous and its faults can be seen.

2. Irony: To present things that are out of place or are absurd in relation to the surroundings.

3. Reversal: To present the opposite of the regular order (e.g. the order of events, hierarchical order).

4. Parody: To imitate the techniques or style of some person, place, or thing.

5. Cynicism: The ability to look with suspicion at something, or someone and offer an opinion contrary to the status quo is an excellent tool for satire

6. Double entendre: saying one thing and (clearly) meaning another.

Example:

https://www.thedailybeast.com/fooled-by-the-onion-9-most-embarrassing-fails

https://www.theonion.com/

https://www.currantdaily.com/

https://babylonbee.com/

https://theconversation.com/too-many-people-think-satirical-news-is-real-121666

https://preview.redd.it/18bwg09g3zn11.jpg?width=640&crop=smart&auto=webp&s=88def0be2595c89ef9ce12cc2a625218d3fa371f

TV shows, videos:

LastWeek Tonight, The Daily Show with Trevor Noah, The Late Show with Stephen Colbert, Late Night with Seth Meyers.

Memes:

https://i.chzbgr.com/full/9233889792/hF2226249/meme-about-the-first-

picture-of-earth-taken-from-space-being-of-a-turtle-with-grass-on-its-back

http://www.electomatic.com/political-meme-tracker/

Checking method:

Most satire has the following characteristics in common:

- Satire relies on humour to bring about social change.
- Satire is most often implied. The reader has to pick up on the humour, or he/she will miss the satirical nature of the writing.
- Satire, most often, does not go over individual people. Instead, satire is directed at society as a whole, or types of people in a society-the politician, the adulterer, the prideful, etc.
- The wit and irony of the satire are exaggerated - it is in the exaggeration that people are made aware of their foolishness.

TROLLS, BOTS AND FAKE OR PUPPET ACCOUNTS

Who is able to produce:
Professional – **Amateur** – Anyone.

Level of skill needed to produce active troll, bot, fake or puppet accounts varies. Anyone can create and utilise a simple fake account, use trolling techniques or buy bots for click and like farming. There are online tools that can generate all kinds of counterfeit personal information, needed to create fake accounts – from phoney names to temporary email addresses to National ID Number generation and validation.
At least a minimum amount of programming skills is needed in order to create social media bots. The most harmful effects of troll, bot or fake accounts are usually conducted by people who have professional skills: some bots employ advanced AI techniques in order to look more realistic; some trolls use compelling storytelling and manipulation techniques in order to get a needed reaction. The creation of fake social media profiles (or buying 'likes') is now an industry worth over 700 million €.

Level of deception:
Low – Average – **High** – Very high.

Level of deception heavily varies – while some trolls, bots or fake accounts can be identified easily, others look like accounts of real people, and need a more serious investigation to identify.
A study from the University of Reading School of Systems Engineering found that 30% of people in the study could be deceived into believing a real person ran a social media bot account. Trolls usually mislead other social media users by posting harmless content, creating realistic profiles and stories.

A troll is a person who deliberately tries to upset or start an argument, especially by posting offensive or unkind things on the internet (source: https://www.collinsdictionary.com/dictionary/english/troll).

A bot is a software application that runs automated tasks over the Internet (in this case follow social media accounts and interact by like, comment, share or other platform functions). Bots behave in an either partially or fully autonomous fashion and are often designed to mimic human users.

A puppet account is an account someone sets up to act in ways they either can't publicly, or to support them (to upvote their own material, and give positive comments, praise, or advertise my work).

Working principle (what and how does it do):

Some **tactics trolls use** (source: https://medium.com/better-humans/the-complete-guide-to-understanding-and-dealing-with-online-trolls-4a606ae25c2c):

Refusing to back down on known fallacies: when one troll tells a lie (either directly or through the use of hyperbole, omission, or twisting facts), many will repeat it - even if it can be easily disproved.

Troll telephone. A troll in one forum says something flip, and another troll takes it as truth and repeats in another forum. Then it becomes a lie that gets told repeatedly.

Sea-lioning: Repeated and relentless questioning, often after the question has been explained in detail multiple times. The sea lion will insist they are acting perfectly civilly, but they are just trying to delay you as long as possible and derail the conversation. The name comes from a webcomic frame: http://www.muddycolors.com/wp-content/uploads/2017/12/81acd-a5b.jpg.

Flaming: Bringing up incendiary and controversial topics to overwhelm a post or moderator, who must deal with finding and policing every post.

Grammar police: Not caring about the content of your post or comment but insisting your spelling and grammar must be perfect, or you can't possibly make a valid argument.

Boomerang: Someone who returns as much as possible to keep commenting on a thread. Even if you do block them on social media. They'll make new accounts and keep making comments to follow you until you are convinced, they are right.

Flooding: When someone posts on your page, but they repeat the same thing over and over, to destroy the ability to have a conversation with anyone else. Usually, it's something like "lol" or something NSFW, or just childish and taunting.

Hatemonger: That person that goes straight for the incendiary words and name-calling—or right for the death thrusts and rape threats—even when the thread or comments didn't warrant that

level of response. Drives all your sane commenters into a raging frenzy and the conversation immediately turns into a melee.

For social bots to be applied to a specific (social media) channel, the platform has to be accessible through an Application Programming Interface (API), as offered, e.g. by Twitter and Facebook. By using APIs, a large number of bot accounts can be controlled simultaneously with little effort. With simple keyword searches, they scan Twitter timelines and Facebook posts for specific terms or hashtags. As soon as they find what they are looking for, they comment, share links or start a fictive discussion. Or they comment directly on specific topics. In combination with other bots (forming a botnet), their noise becomes even louder and can mislead other users.

Malicious **social media bots** can be used for a number of purposes (source: https://www.cloudflare.com/learning/bots/what-is-a-social-media-bot/):

Artificially amplifying the popularity of a person or movement: A person or organisation with millions of social media followers can be seen as important or influential. A primary use case of social media bots is to boost the apparent popularity of other accounts.

Influencing elections: A study by First Monday, a peer-reviewed journal found that in the day before the 2016 U.S. presidential election, as much as 20% of political discussion on social media was generated by about 400,000 social media bots.

Manipulating financial markets: Social media bots can also be used to influence financial markets. For example, bot accounts can flood social media with manufactured good or bad news about a corporation, in an attempt to manipulate the direction of stock prices.

Amplify phishing attacks: Phishing attacks rely on an attacker gaining their victim's confidence. Fake social media followers and social engagement can help convince a victim that their scammer can be trusted.

Spreading spam: Social media bots are often used for illicit advertising purposes by spamming the social web with links to commercial websites.

Shutting down free speech: During the 2010-2012 Arab Spring movement, government agencies used Twitter bots to overwhelm social media feeds. These bots were used to push down the messages of protestors and activists deliberately.

More about trolls: https://www.lifewire.com/types-of-internet-trolls-3485894.

More about bots: https://niccs.us-cert.gov/sites/default/files/documents/pdf/ncsam_socialmediabotsoverview_508.pdf?trackDocs=ncsam_socialmediabotsoverview_508.pdf.

Example:

Trolls:

- https://www.rollingstone.com/politics/politics-features/russia-troll-2020-election-interference-twitter-916482/
- https://motherboard-images.vice.com/content-images/contentimage/32137/1459536705346993.png?resize=664:*
- https://motherboard-images.vice.com/content-images/contentimage/32137/1459536758030213.png?resize=638:*
- https://imgur.com/gallery/INtY5SB
- Fake customer support troll account: https://imgur.com/t/trolling/4yTQWCo

Bots:

- https://www.targetinternet.com/social-media-spam-bots-and-fake-engagement/

Fake accounts:

- https://socialmediarevolver.com/fake-facebook-accounts-attacking-facebook-groups/
- https://www.hackread.com/google-image-search-social-media-profiles/
- https://www.sbs.com.au/news/thai-click-farm-raided-over-300-000-sim-cards-found
- https://techcrunch.com/2018/08/27/twitter-suspends-more-accounts-for-engaging-in-coordinated-manipulation/

Checking method:

Recognising and Dealing with different types of trolls: https://www.teamtechnology.co.uk/troll-tactics.html.

While some of the most advanced social media bots can be hard to spot even for experts, there are a few strategies to identify some of the less sophisticated bot accounts. These include:

- Running a reverse image search on their profile picture to see if they are using a photo of someone else taken off the web.
- Looking at the timing of their posts. If they are posting at times of day that don't match up with their time zone or are making posts every few minutes every single day, these are indications that the account is automated.
- Using a bot detection service such as botcheck.me that uses machine learning to detect bot behaviour. Cloudflare Bot Management also uses machine learning to identify bots.
- https://botometer.iuni.iu.edu.

Recognising a fake social media account: https://smallbusiness.chron.com/spot-social-media-fake-46150.html.

BEFRIENDING (AND IMPERSONATION)

Who is able to produce:
Professional – **Amateur** – Anyone.

Good impersonation requires high skills, but even amateurs can create believable impersonation and trick other people. Befriending requires basic psychological knowledge and being good at reading other people.

Level of deception:
Low – **Average** – High – Very high.

Some impersonations are easy to spot. Some criminals will pretend to be a large organisation you likely are doing business with. In contrast, others will do more in-depth research into you and the company you work for and attempt to fool you into believing they are a company executive. It is hard to spot befriending at the beginning of such a process because it is no different from a friendly relationship. In later stages, when befriending person will try to use this relationship to his/her advantage, it becomes easier to spot.

Impersonation – imitation of someone actions, behaviour. Pretending to be someone else.

Befriending – posing as a friend (or friend to be) in social media with a purpose to deceive or to take advantage of (i.e. get personal info, photo, video).

Working principle (what and how does it do):

Usually, fake accounts are used for impersonation. These accounts imitate celebrities, existing brands or organisations, or random people. At times, the accounts can imitate friends, relatives or others who are close to the potential victim. Sometimes, instead of creating fake accounts, hackers target accounts of inactive users and use them to target the friends who are still active on the platform.

When creating accounts that impersonate celebrities or organisations, various social media platform loopholes are used. E.g. it's possible to imitate a popular YouTube channel since the name displayed on YouTube channels, and YouTube accounts can be different from the actual account name. Within YouTube, users can send friend requests to anyone on the platform. Once accepted, they can send that person direct messages. This way, someone impersonating a famous YouTuber can send messages to subscribers, fooling people that the said star themselves contacted them.

Sometimes they send elementary messages informing the recipient that they've won something, inviting them to click links that potentially lead to scam or malicious sites. Other times these threat actors leveraged a combination of creative impersonation techniques, which boosted the legitimacy of their messages and improved the likelihood that users would click their links.

For befriending, both fake and real accounts may be used. But it depends on the medium in which befriending is happening, i.e. in online video games usually nicknames are used which does not give any information about the true identity of the person.

By using impersonation or befriending scammers can also trick people into:

- giving away money (by transferring them or "making a donation ");
- giving away sensitive information;
- downloading malicious software;
- visiting scam sites.

A typical impersonation attempt by cybercriminals is for them to pretend to be with one of the principal online players that you may pay a regular subscription fee to. Apple Music, Spotify, Netflix, and others are commonly seen. You'll receive a breathlessly-worded message in your inbox warning you of some problem with your account. And if you don't click right this second, they'll have no choice but to lock you out of your account and block any further access. If you do click, you will be sent to a copycat website that looks similar (if not identical) to the impersonated company, and you will be asked to provide your login credentials.

Once you "log in" to the fake site, you will be asked to confirm all your billing details – but the criminals ask for far more information than you should be providing. They'll ask for your complete mailing address, your full credit card details, including expiry and CVV code. Some will ask for other incredibly personal information like your mother's maiden name and your Social Security Number. Everything a cybercriminal needs to steal your identity, open new accounts in your name, or take over some of your other accounts. Other cybercriminals will use similar techniques but claim to be from your bank or your cell phone carrier.

Example:

https://www.riskiq.com/blog/labs/youtube-impersonation-scams/

Quiz "Find the Fake ": https://www.zerofox.com/find-the-fake/

Checking method:

If someone is trying to convince you that they're a celebrity, take the following precautions:

- Check out the identity of the person contacting you. Can you verify that they are who they say they are? If not, or if you're unsure, stop corresponding and don't do what they're asking of you.
- If you're contacted by a celebrity from their own social media account, carefully examine the account. Does it include the blue checkmark that verifies they are who they say they are? Does the information in the account correspond with news stories about this celebrity?
- Google the celebrity's name plus the word "scam" to see what comes up.
- Consider reporting the matter to the social media site where you encountered this person.

Check the profile of new requests to connect or be friends, especially if you have only met the person online. Look out for:

- new profiles with limited content
- hidden friend or network lists or lists full of people of the opposite gender
- Don't send money to someone you've never met in person.
- Be cautious when sharing personal pictures or videos, especially if you've never met them before in person. Scammers are known to blackmail their targets using compromising material.
- Don't share personal information with someone you have never met in person.
- Do an image search of your admirer to see if they are who they say they are. Use image search services such as Google or TinEye.

RESOURCES

We draw inspiration from:

- https://eavi.eu/ and specifically https://eavi.eu/beyond-fake-news-10-types-misleading-info/;
- https://firstdraftnews.org/ and specifically https://firstdraftnews.org/latest/fake-news-complicated/;
- https://euvsdisinfo.eu/;
- https://newslit.org/;
- https://groundviews.org/2018/05/12/infographic-10-types-of-mis-and-disinformation/;
- https://en.unesco.org/sites/default/files/f._jfnd_handbook_module_2.pdf;
- https://misinfocon.com/catalogue-of-all-projects-working-to-solve-misinformation-and-disinformation-f85324c6076c;
- https://www.ifla.org/publications/node/11174;
- https://faktabaari.fi/assets/FactBar_EDU_Fact-checking_for_educators_and_future_voters_13112018.pdf;
- http://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf;
- http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624279/EPRS_STU(2019)624279_EN.pdf.