25/01/2022

CONCLUSION PAPER

RAN C&N 'What's going on online? Dealing with potential use of deepfakes by extremists'
10-11 November 2022, Helsinki, Finland

What's going on online? Dealing with (potential) use of deepfake technology by extremists

Key outcomes

Being online, and doing online work, has become an integral part of everyday life. More and more work in the field of preventing and countering violent extremism (P/CVE) is also taking place online. But the online landscape changes and develops at a fast pace. In an earlier RAN C&N Working Group meeting, practitioners doing online work expressed the need to stay up to date regarding online developments. To cater to this need, the C&N Working Group held a meeting on a topic novel to the field of P/CVE: **What's going on online? Dealing with (potential) use of deepfake technology by extremists**.

Artificial intelligence (AI) technology is able to let a computer create video, audio or other media content: synthetic media. AI-created audio-visual content impersonating a person is called 'deepfake'. But what are the potential usages of this technology by extremists, and what are the implications for the field of P/CVE? And, could P/CVE practitioners potentially use this emerging technology 'for good'? On 10 and 11 November 2022, practitioners with experience in working online gathered with (tech) experts and researchers experienced in the topic of deepfakes and synthetic media to discuss these questions.

Several experts presented their views on the topic, and all participants took part in a 'red team/blue team' exercise to step into the shoes of either extremist groups or P/CVE practitioners to brainstorm about potential usage of synthetic media technology. This meeting was highly exploratory and covered hypothetical scenarios of *potential* (mis)usage. This allows practitioners to be ahead of the curve and anticipate problems that might emerge soon.

The key recommendations of this meeting are:

- Be aware that attempting to use deepfake or other synthetic media technologies 'for good' **can be counterproductive**. They could easily be used by extremist groups to undermine your goals.
- Take into account **ethical and moral considerations**. In this paper, you can find a set of questions that can guide you through these considerations.

- Keep in mind that the **emotional aspect is key** in communicating, both when working to counter or prevent extremist content and when you want to use synthetic media in a positive way.
- Consider **working together with emerging tech companies**, for example, in deepfake recognition software.
- Incorporate awareness and knowledge about deepfakes and synthetic media in **media literacy training**.

The remainder of this paper describes the highlights of the discussions held during the meeting, covering the keynote presentations by experts, the red team/blue team exercise, and several current initiatives dealing with the topic at hand. Subsequently, the recommendations following these discussions are formulated as well as potential follow-ups.

Highlights of the discussion

1 Deepfake and other synthetic media: current state and malicious use

Synthetic media is: video, image, text or voice that has been fully or partially generated by computers ([1]). Deepfakes, a portmanteau of 'deep learning' and 'fake media', are a type of synthetic media invented in 2017. They involve the use of AI techniques to manipulate or generate fake visual and audio content that humans or even technological solutions cannot immediately distinguish from authentic content ([2]). At the start of the meeting, two presentations regarding the current state of deepfake and synthetic media technology were given.

The technology to produce synthetic media and mostly deepfakes is becoming more and more widespread and is being used for a diverse array of media. However, according to the presentations given, the most prevalent (more than 95 %) type of deepfake videos are of a pornographic nature ([3]). When looking at the way deepfake pornographic content is being (mis)used, new trends can be observed. These videos are being used for extortion, blackmail, cyberstalking and 'digital trafficking' (making money by using someone's image). Today, it is even possible to produce 'custom products' where a deepfake porn video of someone can be made on request for USD 50 per minute.

Politicians are being used to produce deepfake videos, sometimes to raise awareness about the existence of deepfakes, sometimes to depict the politicians as heroes, and sometimes to make a politician look ridiculous and undermine their reputation. During the first months of the Russian war of aggression against Ukraine, deepfakes were spread suggesting Ukrainian President Volodymyr Zelensky would surrender or suggesting that Zelensky's wife had left the country and fled to France.

Synthetic media (including deepfakes) can also be used for spreading propaganda or disinformation, as is described in a publication by the United Nations Interregional Crime and Justice Research Institute (UNICRI) ([4]). Spreading propaganda or disinformation can for instance be part of a terrorist campaign, aimed at manipulating public opinion and undermining people's confidence in state institutions. One example is deepfaking political figures to make offensive remarks against a specific community to increase outrage and influence the number of sympathisers. In this way, a video can quickly reach millions of people as users spend little time authenticating the content before sharing. Even with a short lifespan (as the video will probably be debunked), such media can create momentary panic and confusion, especially when they go viral.

Other (potential) malicious uses of deepfakes by terrorists named in the report by UNICRI are less visible to the public. For instance, there have been multiple examples of deepfake voice fraud. With this type of fraud, someone's voice is automatically changed in a way that it resembles the voice of someone else. In this way, criminals have been mimicking the voices of CEOs of companies to steal their money. This method could also be used to finance terrorism. Another form of synthetic media production is to create 'morphed photos'. These AI-generated photos

([1]) https://www.synthesia.io/post/the-future-of-synthetic-media
([2]) https://unicri.it/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes
([3]) https://www.tilburguniversity.edu/sites/default/files/download/Deepfake%20EN.pdf
([4]) https://unicri.it/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes

can be matched with more than one individual, without being detected. This can enhance the ability of terrorists to move undetected through border controls or public spaces by using morphed passports.

Side effects of this technology becoming more prevalent can be that the trust levels for information will further decline. This could lead to an 'information apocalypse' or 'reality apathy': anything can be fake, so trust nothing. This situation can also be referred to as a post-truth world. This is especially worrying when combined with the fact that trust in democracies is already falling, as shown in the 2022 Edelman Trust Barometer ([5]). In many of the democracies studied, democratic institutions are trusted by less than half of their people, including only 46 % in Germany, 45 % in Spain, 44 % in the United Kingdom and 43 % in the United States. The same survey showed fake news concerns are at an all-time high. Concern about fake news or false information being used as a weapon has risen to 76%.

In the near future, as the technology to produce synthetic media will get cheaper and more accessible, 80-90 % of all media (i.e. on social media) are expected to be manipulated ([6]). This trend is already visible on social media platforms such as TikTok, where a lot of content is manipulated with real-time filters. As more and more (visual) media content will be manipulated, it will further blur the lines between 'real' and 'fake' content.

2 How could extremist groups and P/CVE practitioners potentially use deepfakes and other synthetic media?

During the meeting, participants took part in a **red team/blue team exercise**, stepping into the shoes of extremist groups (red team) and P/CVE practitioners (blue team). In a first round, the teams brainstormed about potential ways to use deepfakes or other synthetic media from their perspective (now or in the future). In a second round, the teams discussed how they would respond to the opposing team's ideas.

DISCLAIMER: please note that the following 'red team' and 'blue team' ideas presented below are purely hypothetical and based on a brainstorm session. It serves to give readers an idea of what could be possible, and it in no way reflects what is already being done or what should be done.

Potential malicious use by extremist actors and possible responses from P/CVE

Participants from the **red teams** identified several ways that deepfakes and other synthetic media could potentially be used for malicious purposes by extremist groups:

- In campaigns to spread disinformation, conspiracy theories or extremist narratives, **deepfake influencer social media profiles** could in the future be used. Right now, fake profiles are already being generated but usually only use a fake profile picture. Future use of deepfake technology could create completely fake influencer accounts containing convincing deepfake video material.
- Using AI technology, the **alteration of existing videos** could help strengthen extremist narratives. Video content of violence and hate could be altered to make the perpetrator appear to belong to a certain group (i.e. an ethnic minority), in order to blame this group and hereby strengthen extremist narratives. For example, changing the skin colour of the perpetrator in a video with senseless violence can strengthen racist narratives against the ethnic minority in question.
- Deepfake content can be used to **undermine the democratic process**, especially during election times. For example, non-consensual deepfake pornography (currently the most-used form of malicious use of deepfakes) of political candidates can bring them into discredit. Radical or extremist groups can use this to strengthen their anti-government narratives. Another possible application could be threatening public figures by showing them deepfake footage that could be spread via social media.
- In relation to other technological advancements, **bot armies** (large numbers of fake social media accounts) can be used to **push deepfake content** like the above suggestions, making them more believable or likely to be perceived as being real.

Subsequently, participants from the **blue teams** discussed how they would respond from the perspective of P/CVE practitioners to this potential malicious use of deepfakes and other synthetic media.

([5]) https://www.edelman.com/trust/2022-trust-barometer
([6]) https://futurism.com/the-byte/experts-90-online-content-ai-generated

- In addressing the **idea of deepfake influencer profiles**, participants suggested several ways to **prevent** and **counter** this:
  - Preventing: creating awareness of influencer profiles being potentially fake; introducing solid reporting mechanisms on platforms to report fake accounts and deepfake content; detecting deepfake accounts early.
  - Countering: exposing the people behind these deepfake profiles and having legislation to punish or fine them; infiltrating organisations behind the deepfake profiles (by the police); blocking malicious deepfake content.
- In addressing the **idea of using deepfakes to undermine the democratic process** (during election time), participants suggested multiple ways to **prevent** and **counter** this:
  - Preventing: pre-bunking deepfake content; investing extra resources during election time to work on prevention; strengthening legislation aimed at platforms and content creators.
  - Countering: having legislation in place to punish or fine perpetrators; strengthening legislation for accountability; investing resources to hold official investigations (i.e. by the police) and to create mechanisms to expose the people/organisations behind deepfakes.

Potential use by P/CVE practitioners and possible responses by extremist groups

Participants from the **blue teams** identified potential ways that P/CVE practitioners could use deepfakes or other synthetic media 'for good':

- **Role models:** Creating **positive role models** using synthetic media. This idea comes from the viewpoint that the youth is one of the most vulnerable groups to be affected by malicious use of synthetic media or deepfakes, so creating positive role models as alternatives could address this.
- **Education:** Deepfakes and synthetic media could be used as an **educational tool**, for example, by creating deepfakes of interesting people who witnessed important historical events. One specific idea here was to create deepfakes of holocaust survivors who can answer questions from the audience about their experience.
- **Counselling:** Deepfakes of **victims or perpetrators (or formers) in a therapeutical/counselling setting** ([7]) could help in rehabilitation. For example, if a victim does not want to talk to a convicted extremist themselves (or vice versa) but does agree for a deepfake of that victim to have a conversation with the convicted extremist, this could be used in the rehabilitation process of the extremist by creating a common ground or shared experience.

Subsequently, participants from the **red teams**, thinking as extremist groups, discussed how they would oppose these potential ways that deepfakes and other synthetic media could be used 'for good':

- To counter using deepfakes as an educational tool: it would be just as easy to create deepfakes of the person used in an educational setting but let them say contradictory things to **undermine the educational goal**. The same goes for deepfakes of positive role models: it is easy to **make the role model look 'bad'**, appear negative, etc., if the initially used role model is already a deepfake.
- In general, the fact that deepfakes are being used 'for good' can be used to **undermine the trustworthiness of practitioners and governments**. This can only fuel narratives that the government is lying, that is, it plays into anti-authorities and conspiracy narratives.

3 Ethical considerations for using deepfake technology by P/CVE practitioners

One of the outcomes of the red team/blue team exercise is that it appears that using deepfakes 'for good' can be counterproductive, as it can be easily countered and misused by extremist groups adhering to different rules and morals. Following up on especially the blue teams' ideas, the question arises as to whether it would be possible for P/CVE practitioners to use deepfake technology 'for good' in a way that is *not* counterproductive. To do this, it is important to reflect on the **ethics** around using emerging technologies like deepfake technology. Based on a presentation during the meeting about the ethics involved, the main considerations here are:

([7]) See also https://beeldengeluid.nl/en/knowledge/blog/deepfakes-mental-health-using-artificial-intelligence-good

- Taking ethics into account, and being able to judge what is 'right' or 'wrong', when dealing with **emerging technologies in general** can be difficult because:
    - traditional moral principles might not fit well with recent developments of these technologies. For example, accountability for spreading disinformation would normally lie with the author. But what if the author is an AI?;
    - it is difficult to **foretell what the intentions** of the possible use of emerging technologies will be, as their (potential) applications are not fully known yet;
    - it is difficult to **foresee the consequences** of using emerging technologies if their (potential) applications are not fully known yet; and
    - it is difficult to **weigh against potential alternatives** if there is a scarce availability of knowledge about the pros and cons of emerging technologies.
- When discussing **deepfake technology** in particular, it can be said that the technology is not **inherently bad**, as there are ways this technology can be used to do good. Therefore, deepfake technology can be compared to a knife: it can aid, for example, serve as a tool for a surgeon, or injure, by being used as a weapon.

A useful model with a **set of questions** practitioners can ask themselves when they want to work with deepfake or synthetic media (and potentially other emerging technologies as well) was presented. These thought-provoking questions centred around misrepresentation, deception, manipulation and mistrust.

This model served to reflect on the ideas the blue teams came up with earlier. Are there any promising ideas to use deepfake technology or other synthetic media 'for good', in an ethical way? First of all, it was concluded that using an actual deepfake is morally dubious in itself (because you make someone say or do things that have not taken place in the real world). Therefore, any P/CVE intervention attempts using deepfakes would inherently be more susceptible to possible backlash, especially if it is not clear upfront that a deepfake person is being used. Second, the ethical side around using deepfakes *can* be addressed, but this would impact the effectiveness. For example, if you make it clear upfront that you are using a deepfake, you might not appeal to the target group you are trying to reach. Third, even if all ethical boxes are checked, the mere fact that a deepfake is being used can still be counterproductive, for example by sowing more mistrust in media. It can therefore be concluded that using deepfakes in general P/CVE interventions will not work that well. Maybe only in closed (therapeutic) settings might it be effective, when applying the ethical standards that rise from the above-stated questions. However, using deepfakes in counselling settings still needs to be evaluated.

On the other hand, playing into emotions can be effective — this is exactly the strategy used in extremist narratives. Synthetic media using AI-generated 'people' can be useful in this sense. While *deepfakes* might have too many ethical (and practical) obstacles as they are duplicating an existing person, other types of synthetic media might be more useable. Using synthetic media can help to have emotionally charged communication to deliver the (counter) message that P/CVE practitioners may want to bring across.

Moreover, if deepfakes cannot be broadly used in a preventive way, it might be more fruitful to invest in finding ways to prevent or counter malicious use rather than invest in using this technology as a P/CVE practitioner.

4 Current initiatives dealing with deepfakes and other synthetic media

Many participants at the meeting are actively working on topics related to synthetic media. Three of them presented their initiatives to deal with deepfakes and other synthetic media at the end of the conference. The three topics were: 1) deepfake recognition software, 2) consequences of synthetic media for (digital) media literacy training, and 3) the prevalence of synthetic media on TikTok. These are the results from their presentations and the ensuing discussions:

Deepfake recognition software

One of the participants is working at a start-up aimed at developing automated recognition of deepfake content. Certain alterations to the manipulated content that are not visible to the human eye can be recognised by machine learning software. Besides being relevant to private companies, for instance when checking someone's identity, the detection software that is being developed could also be applicable for individual or civil society use. A possible way of organising this could be through a browser plug-in. In order to be transparent about the workings of this technology, it is important to offer insights as to why something is classified as a deepfake.

In the discussions that followed, some challenges for P/CVE applicability were identified. Firstly, as this is now a reactive tool (identifying deepfake content after it exists/is published online and damage has already been done), could there be a way to make it more preventive (for instance, by blocking the uploading or spreading of such content in the first place)? Secondly, a more general challenge is to search for the best way to work together with big tech companies to make more impact. As legislation is lagging behind, currently no judicial incentives for big tech are in place to force them to do something about the spread of malicious use of synthetic media.

AI in media literacy

The second initiative that was discussed is a media literacy training delivered in Serbia called 'new literacy'. In this project, media professionals, other organisations and influencers are involved. This is especially important in Serbia, as it is among the lowest ranking countries on the Media Literacy Index 2022 ([8]).

During the discussion, the following points were brought to the table:

- It is difficult to integrate information about deepfakes into media literacy training, as for an individual it is hard to distinguish deepfakes from regular media. There is no step-by-step approach to be offered to the public for this. That is why we need technological solutions for this, like a browser extension to flag if media are fake.
- At the same time, we shouldn't refrain from taking deepfakes into account in media literacy training as a part of critical thinking. We should be mindful of the fact that it might be too much to ask from the public to critically assess every image, video or audio clip they come across.
- One participant raised the point that we should not tell everyone (especially children) not to trust anything they see. We should still be able to develop trust.
- Someone said we should focus on adults, who are hard to reach, compared to children who can be reached via schools, etc. Recent attacks show conspiracy thinking can lead to extremism in adults.
- 'Being critical and making your own judgement' is used in media literacy training, but it is also a rhetoric often used by extremists. We should be aware when promoting digital media literacy not to play into that rhetoric.
- People are not necessarily looking for the truth when they look for information. People look for ideas and facts that fit their own opinions instead of the truth as they want their own world views to be confirmed. This is something to take into account when speaking about truth and manipulation in media literacy training.

Manipulated media on TikTok

The third project that was featured in this round of initiatives is a project aimed at studying risks of spreading mis- and disinformation through TikTok, including the use of manipulated content. For many young users, TikTok is their primary source for news consumption (so it is not only about dances on TikTok). The way the platform works makes the spreading of disinformation a rather big risk. There are several aspects that can make TikTok content easily misleading:

- **Time stamp:** in the design of the platform, TikTok videos do not have an original time stamp when they appear on an individual's personal page/timeline. If you want to know the date it was uploaded, you will have to find the owner of the clip.
- **Recycling of content:** if a video went viral once, TikTok tries to make it viral again after some time. Within the app, people reuse audio or video clips from other sources to make their own video. So, recycling is an integral part of the process to create content.

([8]) See also https://osis.bg/?p=4243&lang=en

- **Use of filters:** the software contains filters and other possibilities to alter one's voice and face. TikTok won't tell you if a video uses filters.

As TikTok features short videos to a primarily young audience, users fail to capture and process content correctly and are not able to assess questions of whether a clip is real or synthesised in any way.

Recommendations

The discussion during the meeting brought together different perspectives on synthetic media and deepfake technology, resulting in several recommendations for practitioners and policymakers.

Recommendations for practitioners

One of the conclusions of the meeting is that **deepfakes themselves are less usable in P/CVE work** due to their inherently misleading nature. **Other types of synthetic media** (or other emerging technologies) have more potential, as long as they are used in a **transparent** and **ethical** way. Before you start using such technologies, take into account the following considerations:

- Be aware that using deepfake or other synthetic media technologies 'for good' **can work counterproductively**. It could easily be used by extremist groups to undermine your goals.
- Take into account **ethical and moral considerations** if you want to work with emerging technologies. This is not only relevant for deepfakes or other types of synthetic media, but also for other emerging technologies.
- When using synthetic media, it is preferable to be **transparent beforehand**. If this works counterproductively in achieving your goal, consider if you could include this transparency in the medium (audio or video clip, for instance) itself.
- Keep in mind that the **emotional aspect is key** in communicating, both when working to counter or prevent extremist content and when you want to use synthetic media in a positive way.
- Keep an eye on developments regarding the use of these types of technology in **adjacent fields**. One potential way to use synthetic media and/or deepfake technology in P/CVE is to use it in a **therapeutic/counselling setting**, for instance when dealing with trauma ([9]); another could be to use this technology to create testimonials from contemporary witnesses such as victims or formers.
- When developing a new campaign on either the prevention or counter side, a **red team exercise** can be a useful working method to brainstorm. If you work with a multidisciplinary team, an accompanying **blue team exercise** can help to get everyone on the same page ([10]).

Looking at the recommendations coming out of this meeting, there is also potential to **update the GAMMMA+ model** ([11]) for developing counter- or alternative narrative campaigns:

([9]) Although not mentioned in the meeting, one application of this has already been seen. In a treatment of victims of sexual assault trauma in the Netherlands, a set-up using live deepfake technology was used. Besides the victim, two therapists are involved in this. One of the therapists is accompanying and guiding the victim in talking to the deepfake version of the perpetrator on screen. The second therapist is in another room, using voice alteration and video software in such a way that what they are saying and how they are moving is being presented to the victim on the screen as a deepfake of the perpetrator. This allows for confrontation and discussion with the deepfake version of the perpetrator, which can be helpful in dealing with trauma for instance. Transferred to the P/CVE setting, this approach could also be used for victims and survivors in their trauma treatment.
([10]) This method is often used in cybersecurity, see for instance: https://purplesec.us/red-team-vs-blue-team-cyber-security/
([11]) Building on https://home-affairs.ec.europa.eu/pages/page/ran-cn-effective-narratives-updating-gammma-model-brussels-14-15-november-2019_en

- The ethical and moral considerations around using deepfakes or other synthetic media can be applied in a broader sense within the GAMMMA+ model. Are your **goals**, the **medium** you are using, the **message** you are sending and the **messenger** you want to use all within ethical boundaries? Challenging practitioners to think about their project/campaign in this way can be a useful addition.

- In the **Audience** part of the GAMMMA+ model, undergoing a **red team exercise** can be a helpful suggested tool to deepen the understanding of the target audience.

There are also ways for P/CVE practitioners to **prevent or counter potential extremist use** of deepfakes and synthetic media:

- Consider **working together with emerging tech companies**, for example, in deepfake recognition software. There are promising start-ups that build AI deepfake recognition software. This is a new field in the technology sector. This software can be used for incorporation in P/CVE work, for example, to show fake content and how it is recognised by this software in media literacy training.

- **Incorporate deepfake and synthetic media technology in media literacy training.** While media literacy in general has been identified as key in dealing with disinformation and conspiracy narratives, deepfakes and synthetic media add another layer to this. As mentioned above, working together with deepfake recognition software could be a concrete way to do this.

- **Invest time in the media platforms that are used the most.** There is already a lot of manipulated content and sometimes even deepfake or synthetic content on big social media platforms, especially TikTok. This allows for a huge spread of disinformation. Understanding platforms like TikTok and how they can spread mis- or disinformation through manipulated content can help practitioners in finding ways to prevent or counter this.

Recommendations for policymakers

While this meeting was focused on practitioners, several discussions also led to identifying needs and challenges that could be tackled on the policy level. The following recommendations for policymakers can be formulated:

- **Look ahead and anticipate future (technological) developments.** While this was an exploratory meeting on a novel topic for practitioners, the importance of looking ahead and anticipating was stressed during the meeting. This is also true for policymakers (also underlined by the recent United Nations Operations and Crisis Centre expert panel on Safeguarding the Metaverse: Countering Terrorism and Preventing Violent Extremism in Digital Space). While deepfakes are currently not yet often used by extremist groups, there is potential in this.

- **Invest resources and create capacity** to deal with technological developments. Deepfakes pose a risk that will only grow in the future (in the metaverse, for instance).

    - In negotiations with **big tech companies**, address the risks of AI-manipulated media and demand for these companies to develop methods of detection, flagging and/or removal of such content when it is harmful.

    - Consider **working together with tech initiatives** that work in countering or recognising deepfake content.

    - Consider **working together with civil society** to find ways to educate society in dealing with manipulated media in general and deepfake/synthetic media content specifically.

- **Anticipate future developments in the legal framework** as well. Legislation is always lagging behind compared to technological developments. The upcoming Digital Services Act (DSA) is a first step, but it is still unclear how this will be concretely applied. A clear interpretation of the DSA, including what it means for future technological developments, and the implications for P/CVE practitioners are needed. Adding on this, it is important to study possible consequences of using manipulated media for criminal prosecution. When a deepfake video of a person saying illegal things (like denying the Holocaust in Germany) surfaces, that person can claim that is was not them who said those things.

Relevant practices

1. DuckDuckGoose is a tech start-up based in the Netherlands. They offer deepfake detection software, aiming to create a digital environment where we still can believe what we perceive.
2. Propulsion is an organisation specialising in social impact communications and delivering projects around media literacy in the Western Balkans. They incorporate deepfakes and synthetic media in their projects.
3. Visualising Democracy is a project by the German Amadeu Antonio Foundation. Through workshops, training courses and educational materials, they present the challenges of dealing with social media and focus on ideas for positive uses.

Follow-up

This meeting resulted in several follow-up suggestions:

- Further **update the GAMMMA+ model** to include working with online tools and emerging technologies like synthetic media. Moreover, building in some ethical or moral considerations to supplement the 'do-no-harm' principle can be a useful addition to the GAMMMA+ model.
- In future C&N meetings, keep **including tech initiatives** that might have relevant/useful tools for practitioners.
- A **future meeting could focus** more on the **policy level and the implications of policy and legislation for P/CVE practitioners**, especially around technological developments and online (social) media. Specifically, a clear interpretation of the Digital Services Act and its implications for P/CVE practitioners is useful for practitioners doing online prevention or countering work.

Further reading

- Europol Innovation Lab (2022). Facing reality? Law enforcement and the challenge of deepfakes
- UNICRI & UNCCT (2021). Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes
- UNICRI & UNCCT (2021). Countering Terrorism Online with Artificial Intelligence
- de Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology, 34*, 1311-1332. https://doi.org/10.1007/s13347-021-00459-2