Kaushal Bhat
Carlos Field-Sierra
Anthony Bisgood
Chase Klapperich
Joseph Clancy

1. Pu, Jiameng, et al. “Deepfake Videos in The Wild: Analysis and Detection.” *Proceedings of the Web Conference 2021*, 2021, https://doi.org/10.1145/3442381.3449978.
    a. Pu Jiameng explores how deepfake detection defenses do against real-world deepfakes, as compared to pre-existing datasets. To do this researchers collected “the largest dataset of deepfake videos in the wild " and analyzed the popularity and growth of that deepfake content.
2. Pu, Jiameng, et al. “Deepfake Text Detection: Limitations and Opportunities”, 2022, https://doi.org/10.48550/arXiv.2210.09421.
    a. Researchers explore current text deepfake detection by creating several novel ways of circumventing deepfake text detection technology and propose a new way of text detection by “tapping into the semantic information in the text content”.
3. D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6, doi: 10.1109/AVSS.2018.8639163.
    a. This paper proposes a new way to automatically detect deepfake videos using a “temporal-aware pipeline”. Researchers use a recurrent neural network that learns to classify if a video is a deepfake or not. Videos that are used to train this neural network are collected from “multiple video websites".
4. Li, Yuezun, and Siwei Lyu. "Exposing deepfake videos by detecting face warping artifacts." arXiv preprint arXiv:1811.00656 (2018).
    a. In Li Yuezun’s paper, researchers observe that deepfake algorithms can only generate low resolution images that must be scaled up and warped, which creates “artifacts". These artifacts are used to distinguish between real and fake videos.
5. Diakopoulos, Nicholas, and Deborah Johnson. "Anticipating and addressing the ethical implications of deepfakes in the context of elections." *New Media & Society* 23.7 (2021): 2072-2098.
    a. Diakopoulos and Johnson make the case that deepfake technology could have severe impacts to the integrity of many social domains including elections. Harm can be separated into three distinct categories: (1) harm to viewers/listeners, (2) harms to subjects, and (3) harms to social institutions. The possible vectors of attack are laid out; from intimidation of a viewer to reputational harm of candidates.
6. Matern, Falko, Christian Riess, and Marc Stamminger. "Exploiting visual artifacts to expose deepfakes and face manipulations." *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*. IEEE, 2019.
    a. Researchers used a machine learning-based method for detecting deepfakes developed by training a model with YouTube videos, focusing on identifying visual artifacts present in the videos. The artifacts analyzed include lack of reflections and insufficient details in the eyes and teeth regions.

7. P. Korshunov and S. Marcel, "Vulnerability assessment and detection of Deepfake videos," 2019 International Conference on Biometrics (ICB), Crete, Greece, 2019, pp. 1-6, doi: 10.1109/ICB45273.2019.8987375.
    a. In this paper, researchers use a GAN (generative adversarial network) and emphasize how the blending and training parameters can greatly impact the quality of deepfakes. Researchers then tested high quality deepfakes against detection methods and concluded that state of the art recognition systems are vulnerable to deepfake videos.
8. Zhao, Hanqing, et al. "Multi-attentional deepfake detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
    a. In this paper, researchers developed a new method to detect deepfake face forgery as an alternative to the standard “vanilla binary classifier” counterparts that have been traditionally used. The reason they are doing this is because the tells of a deepfake are largely subtle and local, so their approach uses three main components to more precisely detect deepfakes. With their method, they have been able to achieve state-of-the-art performance. The model has been released to the public and is available on GitHub, so we may be able to make use of this model to assist in quantifying the emergence of deepfakes.
9. Lyu, Siwei. "Deepfake detection: Current challenges and next steps." 2020 IEEE international conference on multimedia & expo workshops (ICMEW). IEEE, 2020.
    a. This paper outlines several methods to detect deepfakes, including traditional image processing techniques, deep learning based methods, and multimedia forensics. Also addressed in the paper are challenges faced by these current techniques and possibilities for future research in deepfake detection. This article may help to point us in the direction of methods that we can evaluate the efficacy of in terms of quantifying the emergence of deepfakes on social media, as well as clue us in on the main challenges that we / our methods will face in doing so.
10. Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "Deepfake detection by analyzing convolutional traces." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020.
    a. The researchers who wrote this article developed a new technique to detect deepfake by analyzing the convolution traces generated by each layer of a Convolutional Neural Network; they found that the information contained in the representations created by each of these layers differed when analyzing deepfake vs. authentic videos. This may represent a method that we can use to detect deepfakes in our analysis.
11. Rana, Md Shohel, et al. "Deepfake detection: A systematic literature review." IEEE Access (2022).
    a. This article provides a summary of 112 articles between 2018 and 2020 that present various methods to detect deepfakes. The researchers grouped them into four categories: deep learning-based, classical machine learning-based, statistical, and blockchain-based techniques. The performance, strengths, and weaknesses of different methodologies are discussed in this paper; this may help us select techniques that we will employ in our research.

12. Mitra, Alakananda, et al. "A novel machine learning based method for deepfake video detection in social media." 2020 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS). IEEE, 2020.
    a. These researchers developed a novel technique for detecting deepfake videos in social media. Their technique utilizes a pre-trained convolutional neural network trained on large-scale deepfake datasets; it outperforms state-of-the-art current detection methods and thus may prove to be helpful for us to quantify the emergence of deepfakes on social media.
13. Koopman, Marissa, Andrea Macarulla Rodriguez, and Zeno Geradts. "Detection of deepfake video manipulation." *The 20th Irish machine vision and image processing conference (IMVIP)*. 2018.
    a. Researchers analyzed videos by cropping faces into 8 sub-sections, then calculating an average PRNU (Photo Response Non Uniformity) score. Analysis was conducted on the calculated PRNU scores, revealing that the mean PRNU for an authentic video was significantly higher than a deepfaked one, demonstrating an example of a statistical measurement based method of detection.
14. Yadlin-Segal, Aya, and Yael Oppenheim. "Whose dystopia is it anyway? Deepfakes and social media regulation." *Convergence* 27.1 (2021): 36-51.
    a. Yadlin-Segal and Oppenheim peek at the potential implications of deepfake technology through 3 different lenses, the gendered perspective, the factual perspective, and the professional perspective, all three of which culminate in “a techno-dystopian future in one way or another”, begging the question of social media content regulation. They suggest both regulation enacted by social media platforms as well as needing a more positive spin on related subjects by journalists to mitigate the felt effects and help facilitate an important understanding of “the reciprocal relationship between humans and machines”.
15. Ahmed, Saifuddin. "Navigating the maze: Deepfakes, cognitive ability, and social media news skepticism." *new media & society* (2021): 14614448211019198.
    a. The researcher conducted and analyzed primary data from an online survey to test the relationships between citizen concerns regarding deepfakes, exposure to deepfakes, and skepticism of the news among other factors. The study concluded that frequent exposure to deepfakes are positively correlated with social media news skepticism. Ahmed makes the inference that when citizens are exposed to deepfakes, it induces uncertainty and makes them more likely to question the authenticity of news on social media.
16. Westerlund, Mika. "The emergence of deepfake technology: A review." *Technology innovation management review* 9.11 (2019).
    a. Westerlund dives into the history of deepfakes in addition to the potential benefits that they pose throughout multiple industries, including healthcare, where it might help an amputee digitally recreate a limb or help people with Alzheimer’s interact with a younger, more memorable face. The researcher identified four major types of deepfake producers, including communities of deepfake hobbyists, political players such as an adversarial country, those malevolently acting in bad faith, and legitimate actors.

17. Vaccari, Cristian, and Andrew Chadwick. "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news." Social Media+ Society 6.1 (2020): 2056305120903408.
    a. After surveying a large sample of people in the United kingdom on deep fakes. The study observed that the presence of deep fakes and the challenge of identifying them resulted in widespread cynicism toward real content and a lack of trust in news, ultimately causing a detachment from the concept of truth. Increasing the challenges of a civic culture in democracies.
18. Hancock, Jeffrey T., and Jeremy N. Bailenson. "The social impact of deepfakes." Cyberpsychology, behavior, and social networking 24.3 (2021): 149-152.
    a. This article highlights the limited research that has been conducted on the cultural and psychological effects of deepfakes as of 2021. It outlines some areas of research that have been covered. On topics such as the impact of deepfakes on self-perception, strategies to combat their effects, and policy implications, there remains a significant gap in our understanding of the broader impacts of this technology.
19. Kwok, Andrei OJ, and Sharon GM Koh. "Deepfake: a social construction of technology perspective." Current Issues in Tourism 24.13 (2021): 1798-1802.
    a. This article explores the history of deep fakes and their current and potential future uses, including their ability to create photo-realistic images of destinations for tourism and to generate lifelike depictions of historic art and events in museums. It also touches on the technical details of GANs.
20. Greengard, Samuel. "Will deepfakes do deep damage?." Communications of the ACM 63.1 (2019): 17-19.
    a. The paper touches on the societal issues of deepfakes. Such as the growing concern the deepfakes can be weaponized to depict false new events, or blackmail victims. It also touches on the advancements researches are doing to detect deep fakes, and help stop there rapid spread in social media
21. Kietzmann, Jan, Adam J. Mills, and Kirk Plangger. "Deepfakes: perspectives on the future “reality” of advertising and branding." International Journal of Advertising 40.3 (2021): 473-485.
    a. The paper touches on how deep fakes will deeply affect advertising and branding because they present both opportunities and threats in how they influence tangible ads, consumer perception, and sociocultural context. It also touches on how propaganda, much like an advertisement, will be deeply affected by deep fakes.
22. Dong, Xiaoyi, et al. "Identity-driven deepfake detection." arXiv preprint arXiv:2012.03930 (2020).
    a. The paper discusses issues with current methods of deepfake video detection as well as their solution to this problem. Their proposed algorithm, which detects whether a person in a deepfake is really the person using a sample video, shows promise in being able to accurately detect deepfakes while being resistant to efforts made to intentionally mask deepfakes.
23. Korshunov, Pavel, and Sébastien Marcel. "Deepfake detection: humans vs. machines." arXiv preprint arXiv:2009.03155 (2020).
    a. The paper studies the differences in how well deepfakes are detected between deepfake detection algorithms and human beings by presenting deepfake videos

and real videos and testing the subject as to whether a given video is real or not. An interesting result of this study is that humans are more skilled in detecting some types of deepfakes that the tested algorithms fail to detect, indicating potential weaknesses with current detection algorithms.

24. Das, Sowmen, et al. "Towards solving the deepfake problem: An analysis on improving deepfake detection using dynamic face augmentation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
    a. The paper analyzes some current deepfake detection algorithms and datasets and finds glaring weaknesses caused by the datasets. The paper finds that current detection algorithms are overfitted as a result of inadequate datasets and weak in detecting deepfakes of people that are not in the dataset.
25. Neekhara, Paarth, et al. "Adversarial threats to deepfake detection: A practical perspective." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
    a. The paper conducts research in developing a method that can be used in deepfake videos that can avoid being detected by several deepfake detection algorithms and expose their weaknesses. The paper finds that this can be done by introducing perturbations across every frame of the video since common deepfake video detection algorithms operate based on individual frames.
26. de Lima, Oscar, et al. "Deepfake detection using spatiotemporal convolutional networks." arXiv preprint arXiv:2006.14749 (2020).
    a. The paper studies weaknesses in current common deepfake detection algorithms that operate based on data taken from individual frames. The paper finds that temporal data can be useful in detecting deepfake videos and shows how a deepfake detection algorithm which operates on multiple frames (allowing operation based on temporal data) outperforms individual frame based methods.