Manipulated VISUAL content: tools & techniques

Extracts

From: Master ADVANCED Digital Tools for Research (author: Christine Gardiol), ***available on AMAZON marketplaces***

Check our tools and techniques to debunk manipulated visual content

Audio-visual content manipulation has become the strongest digital concern of this decade, as Gretel Kahn (from the Reuters Institute for the Study of Journalism) emphasized it: "*for those who work to debunk disinformation, the rise of AI generated images is indeed a growing concern since a big proportion of the fact-checking they do is image or video-based… disinformation is a particular concern since images are especially compelling and they can have a strong emotive impact on audience's perceptions*" [i].

There are basically **two forms of visual content manipulations**: on the one hand, **finely manipulated images**, on the other, **artificially generated images** (also called, synthetic images). AI image generators like DALL-E and Midjourney are getting very popular and easy to use. According to Kahn, G. (2023) in a Reuters Institute for the Study of Journalism post, "*anyone can create new images through text prompts. Both applications are getting a lot of attention. DALL-E claims more than 3 million users. Midjourney has not published numbers, but they recently halted free trials citing a massive influx of new users*".[ii] Professionals in the information field face tough challenges.

Arivazhagan, R. and Chandran, S. in a post from September 2023 *"The Quest to Detect and Prevent Image Manipulation"*[iii], identified four main **challenges in detecting image manipulation** in research publications. These can be applied to a wide range of data and information:

1. *The sheer volume of published content: With the increasing number of published papers, it is difficult to manually scrutinize every image in every paper.*
2. *Limitations of manual detection: Visual scrutiny is currently the standard for analyzing images in scientific papers, but it is time-consuming and prone to errors.*
3. ***Lack of standardized image manipulation detection tools**: There is currently no standardized method or tool for detecting image manipulation in scholarly publications. Many of the available tools are experimental and not user-friendly.*
4. *Guidelines specific to the scholarly information industry: There is a lack of guidelines specific to the scholarly information industry for detecting image manipulation.*

Tools to DETECT fraudulent dealings in images: an overview

Reverse image search engines are excellent tools to identify the origin of some visual content. Rather than inputting a search string, you search uploading an image or copying its URL. These engines allow to **investigate images**, see where an image appears on the web and if it is from a suspicious source. You can also verify if the image has been modified, stolen, changed, where it originally comes from, etc. *Google Image Search* (https://images.google.com/) as well as *Bing Image* (https://www.bing.com/images/), and most of the major general-purpose search engines offer this "reverse-image" function; click on the little camera icon in their search box (you must be in their image search engine to see this icon).

In addition, there are **some specialized engines**, such as https://lenso.ai/en and the Canadian TinEye at https://tineye.com/, among others. TinEye is a well-known image search and recognition company that constantly crawls the web and adds images to its index. Today the TinEye index is over 71.2 billion images (end of October 2024). For more, websearch *reverse image search tools* and *best.*

Besides these reverse image search engines, Google announced (October 2023), working on three ways to check images and sources online[iv]. Similar to its "*About the source*" pop-up window for many of its search results, an "*About this image*" will allow users to check the credibility and context of online images, the image's history, other sites that use and describe the image, and the image's metadata. In addition, Google mentioned expanding its Check Explorer tool to provide fact-checks published by fact-checking entities around the world. Quite some promising features. *"AI-generated images are becoming more popular every day. But how can we better identify them, especially when they look so realistic?".* With this issue in mind, Google DeepMind presented mid-2023, SynthID[v] to "*help watermark and identify synthetic images [AI-generated images] created by Imagen".*

InVid at https://www.invid-project.eu/ is a European Union project (2016), which offers an easy-to-use Chrome plugin that allows to verify and debunk videos files and content which is spread via social media. "*In video veritas…The InVID innovation action develops a knowledge verification platform to detect emerging stories and assess the reliability of newsworthy video files and content spread via social media".*

https://fotoforensics.com is a forensic tool that allows users to see if a picture has been manipulated. The picture must be original and of high quality. *« FotoForensics provides budding researchers and professional investigators access to cutting-edge tools for digital photo forensics"*[vi]. Along this line, are imageforensic.org and fakeimagedetector.com.

More and more tools are coming onto the market for tracking visual content manipulation. Some entities also propose their "*image manipulation detection services*". Such as enago.com, *"Author First, Quality First".* According to MarketsandMarkets[vii], the global

market of fake image detection tools is expected to grow at a CAGR of 41,6%. Development is thus in full swing in this field. For more, websearch *tools detect image manipulation* or *edited* or *fake.*

Tools to PREVENT fraudulent dealings in images

Besides the many tools to detect fraudulent dealings in images, a few are being developed to **prevent** images being manipulated by artificial intelligence. Researchers from the MIT CSAIL[viii] proposed (early 2023) an AI tool, PhotoGuard, to **protect images against AI manipulation**[ix]. According to a post from David Curry[x], PhotoGuard is a *"protective shield which alters photos at a level unnoticeable to the human eye. With this alteration, images which are then run through a generative system will look unrealistic or warped, which should prevent them from being used for indecent purposes."* This is a promising development.

In a May 2024 blog post, Gelato[xi] (the world's largest print-on-demand network) proposed **four strategies to protect its clients' artworks** from AI, namely:

1. **watermarking and digital signatures** (watermarking are subtle, semi-transparent logos embedded in one's artwork);
2. **opting out** (some AI platforms now allow to register an artwork, explicitly requesting it not to be included in the datasets used to train their AI models);
3. **cloaking images** with Glaze or Nightshade (these introduce subtle modifications, not seen by the human eye, which confuse AI algorithms);
4. taking **legal action**.

The number of tools that **prevent visual content** from being manipulated is still limited. It is expected that more and more of these tools to detect whether a photo has been generated or edited by AI will come onto the market. The stakes are large, especially when it comes to people's image. Identity theft has grown into a serious issue[xii], especially, when the person is a politician, a religious leader, any influential person. Tools at both ends of the spectrum are being launched: upstream, to help protect images from being manipulated, downstream, to help users detect manipulated or artificially created images.

Tools are not enough …

Besides the tools, a systematic **information evaluation** is essential to any research process. High-quality reliable information is what makes a difference. In a Reuters Institute of Journalism's post, De Marvel, V., a university professor offers some basic and sensible tips on assessing visual content[xiii]: "***to look at the context around the image and question who*** *is distributing these 'news': the more politically incendiary an image is, the more hesitant we should be about its veracity".* A tip from a HuffPost online workshop in June 2020, "*Think before you link: How readers can help fight misinformation online*"[xiv], goes in a similar

direction: any information that makes you jump, think about it, and check further before forwarding. It is a modest start to applying critical thinking to online content, which anyone can easily relate to.

Our first handbook presented various **information evaluation and assessment techniques.** Being able to identify any information disorders and assess the reliability of a piece of information has become a vital skill for anyone relying on the internet for its research projects, whether academically, professionally or privately. Besides demonstrating its relevance, we provide various techniques and recommendations. One is the "4R" which allows to drill down into any information in an organized way. Here's a quick reminder.

Recency → My results are timely and current.

Relevancy → They answer my research questions or I can justify why they don't.

Reliability → They are trustworthy and reliable. I trust their authors, their intentions and their author-ity.

Richness → I provide a variety of contextualized perspectives and points of view.

From: Master ADVANCED Digital Tools for Research, and Master BASIC Digital Tools for Research, **available on AMAZON marketplaces**

---

[i] Kahn, G. (2023). *Will AI-generated images create a new crisis for fact-checkers? Expert are not so sure.* reutersinstitute.politics.ox.ac.uk. (Reuters Institute for the Study of Journalism, University of Oxford). 11 April 2023. Available at: https://reutersinstitute.politics.ox.ac.uk/news/will-ai-generated-images-create-new-crisis-fact-checkers-experts-are-not-so-sure [Viewed 20 July 2024]

[ii] Kahn, G. (2023). *Will AI-generated images create a new crisis for fact-checkers? Expert are not so sure.* reutersinstitute.politics.ox.ac.uk. (Reuters Institute for the Study of Journalism, University of Oxford). 11 April 2023. Available at: https://reutersinstitute.politics.ox.ac.uk/news/will-ai-generated-images-create-new-crisis-fact-checkers-experts-are-not-so-sure [Viewed 20 July 2024]

[iii] Arivazhagan, R. & Chandran, S. (2023). *The Quest to Detect and Prevent Image Manipulation*. straive.com. 21 September 2023. Available at: https://www.straive.com/blogs/the-quest-to-detect-and-prevent-image-manipulation/ See as well strive.com white paper: protecting-the-integrity-of-scientific-images - Straive Website [Viewed 17 July 2024]

[iv] Hebbar, N. and Savcak, C. (2023). *3 new ways to check images and sources online*. Google blog – search. 25 October 2023. Available at:

https://blog.google/products/search/google-search-new-fact-checking-features/ [Viewed 18 July 2024]

[v] Gowal, S. and Kohli, P. (2023). *Identifying AI-generated images with SynthID*. Google DeepMind. 29 August 2023. Available at: https://deepmind.google/discover/blog/identifying-ai-generated-images-with-synthid/ [Viewed 18 July 2024]

[vi] FotoForensics. *Tutorial: Using FotoForensics*. https://fotoforensics.com/tutorial-about.php [Viewed 20 July 2024]

[vii] MarketsandMarkets (2024). *Fake Image Detection Market (Solutions and Services)*..marketsandmarkets.com. April 2024. Available at: Fake Image Detection Market Size, Share and Global Forecast to 2029 | MarketsandMarkets [Viewed 21 July 2024]

[viii] Salman, H. & all (2023). *Raising the Cost of Malicious AI-Powered Image Editing*. arxiv.org. 13 February 2023. Available at: 2302.06588 (arxiv.org) . See as well, Using AI to protect against AI image manipulation | MIT News | Massachusetts Institute of Technology

[ix] . PhotoGuard code is available at : https://github.com/MadryLab/photoguard

[x] Curry, David (2023). *Researchers Create Tool To Prevent AI Image Manipulation*. From rtinsights.com. 28 August 2023. Available at : Researchers Create Tool To Prevent AI Image Manipulation (rtinsights.com) [Viewed 17 July 2024]

[xi] How To Protect Your Art From AI In 2024 (gelato.com) [Viewed 17 July 2024]

[xii] If the subject is of interest, websearch *identity theft* and *statistics*. And check from scantionscanner, *Identity Theft Statistics.* 17 September 2024. Available at: https://www.sanctionscanner.com/blog/identity-theft-statistics-587 [Viewed 24 October 2024]

[xiii] Kahn, G. (2023). *Will AI-generated images create a new crisis for fact-checkers? Experts are not so sure.* reutersinstitute.politics.ox.ac.uk. (Reuters Institute for the Study of Journalism, University of Oxford). 11 April 2023. Available at: https://reutersinstitute.politics.ox.ac.uk/news/will-ai-generated-images-create-new-crisis-fact-checkers-experts-are-not-so-sure [Viewed 20 July 2024]

[xiv] Huffpost Editors (2020). *Think Before You Link : How You Can Help Fight Misinformation Online.* huffpost.com. 8 June 2020. Available at: https://www.huffpost.com/entry/how-to-fight-misinformation-online_n_5ed140f5c5b68d76d74cf196. [Viewed 17 October 2020].