What is Real Anymore? An AI/ML Image Dataset Using Authenticity Validation and Traceable Origins for Every Data Instance

Andrew McDonald

East Tennessee State University 1276 Gilbreath Dr,
Johnson City, TN 37614 USA
mcdonaldai@etsu.edu

Abstract

This project addresses the increasing challenge of detecting AI-generated images by creating a novel dataset titled "What Is Real Anymore?" (WIRA). WIRA comprises two subsets: the first includes over 2000 images, validated as authentically real by a set criterion and sourced from photographs on Flickr. The second subset consists of hyper-realistic AI-generated counterparts for each validated Flickr image, aggregated through the Leonardo.AI commercial API. All Flickr-validated images in WIRA are credited to their respective photographers and retain their associated rights. Commercial use of this dataset requires permission from the photographers or adherence to the copyright laws of each validated Flickr image used. This document details the rationale for image authentication, image categories, the motive for category selection, authenticity validation criterion, methodology for the creation of the dataset, the computational resources used, a review of included and excluded decision records, and potential enhancements to expand WIRA.

Code — https://github.com/McDonaldAndrew-ETSU/Real-To-AI-Pipeline.git

Datasets —
https://github.com/McDonaldAndrew-ETSU/WIRA.git

1 Introduction

In recent years, the rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) technologies has led to the proliferation of AI-generated content across various domains, such as text, images, and videos. While AI-generated content has the potential to revolutionize content creation and improve efficiency, it also poses significant challenges in terms of authenticity, trustworthiness, and potential misuse. The ability to distinguish between human-generated and AI-generated content has become increasingly important to maintain the integrity of information and prevent the spread of misinformation. Thankfully, researchers have tried to tackle the problem of detecting AI-generated content with AI/ML models such as those within the first 10 references (Monkam, Xu, and Yan 2023; Luo et al. 2024; Zhang et al. 2022; Xia et al. 2022; Sun, Wang, and Tang 2014a; Anokhin et al. 2021; Zhan et al. 2023; Wang et al. 2023a; Lorenz, Durall, and Keuper 2023; Liu et al. 2015).

Mainstream image datasets used for training, such as those within the first 9 to 36 references (Liu et al. 2015; Lin, Shang, and Gao 2023; Epstein et al. 2023; Schuhmann et al. 2021; Wang et al. 2023b; Yu et al. 2016; Deng et al. 2009; Karras et al. 2018; Zhu et al. 2023; Bird and Lotfi 2023; Krizhevsky 2009; Rahman et al. 2023; Sun, Wang, and Tang 2013, 2014b; Karras, Laine, and Aila 2021; Karras et al. 2021, 2020; Aksac et al. 2019; Choi et al. 2020; Zhou et al. 2017; Russell et al. 2008; Zhou et al. 2014; Xiao et al. 2010; Lin et al. 2015; Cordts et al. 2016; Wang et al. 2020; Russakovsky et al. 2015), do not validate or authenticate any of the training images. Without recording each image's origin and due to the advancements of AI-generated content, it is impossible to conclude whether an image is truly real or not. Unfortunately, this applies to all images provided in the earlier mentioned datasets as there are not any validation methods or criteria used to determine if web-scraped images are authentically real.

This report aims to address the importance and explanation of the creation of a dataset consisting of authentically real and validated photographs along with AI-generated counterparts. The dataset created will serve as a precursor to future datasets for ensuring each data instance representing an authentically real image is verified by its origin first before its aggregation. To illustrate the complexity of distinguishing between authentically real and AI-generated images, Figure 1 proposes a challenge to identify which images are real. This visual exercise emphasizes the growing difficulty of human perception alone in validating image authenticity, further underscoring the importance of datasets with rigorous validation criteria for real images such as WIRA.

2 Rationale For Image Athentication

Many of the popular and otherwise trusted datasets used for the detection of AI-generated images such as LAION-400M (Schuhmann et al. 2021), LSUN (Yu et al. 2016), and CIFAKE (Bird and Lotfi 2023) were created over a decade from the writing of this document. These datasets contain subsets of images labeled as real. The curators of these datasets, however, did not employ any validation techniques that ensured the web-scraped images used were authentically real. During the period that these datasets were curated,

Figure 1: The challenge in identifying which images are authentically real and which are AI-generated from these shuffled pairs underscores the necessity of robust authenticity validation methods in datasets like WIRA.

AI-generated content contaminating real image aggregation within search engines was not an enormous problem as it is today, if even a problem at all. Research regarding the issue of detecting AI-generated images has been conducted only within the last decade, with an increase of related studies within the last 2 years such as references 1-5, 7-9, 11, and 12 (Monkam, Xu, and Yan 2023; Luo et al. 2024; Zhang et al. 2022; Xia et al. 2022; Sun, Wang, and Tang 2014a; Zhan et al. 2023; Wang et al. 2023a; Lorenz, Durall, and Keuper 2023; Lin, Shang, and Gao 2023; Epstein et al. 2023) due to the enormous performance gains of AI-generative image models. This performance gain is so impressive that now, most humans, even those with a trained-eye, may easily be deceived by the realism of AI-generated images. More now than ever before, AI-generated images are contaminating search engines, causing real images to be interspersed with artificial content, making it harder to find authentic visuals on real-life topics or things. If a dataset was curated today to detect AI-generated content by only web-scraping without any validation criteria, it is guaranteed any set of real-labeled images may be contaminated with AI-generated content. Due to the increasing photo-realism of the AI images, human eye validation is becoming less effective to separate what is real and what is not.

Most researchers, therefore, depend on the foundation of established datasets such as the ones mentioned previously. One could argue, however, that the real images used in these datasets are not "real" since not one of them used any validation criteria to determine the origin of each data-instance labeled real. A counter argument against this, however, can be that AI-generated content did not start proliferating sources of the web-scraped images from these datasets during the time of their aggregation. While this counterargument is plausible, it remains unprovable since no validation criteria were applied to confirm whether the images scraped in the past were genuinely real or computer altered. Some of the first AI-generated photos can be traced back to over a decade from the writing of this report. Even so, one could argue that the timestamps in the image's metadata were forged. If these datasets of 'real' images lacked records of the origin and authenticity validation, it is now impossible to confirm the true authenticity of each image, leaving room for perpetual debate over their genuineness.

3 Image Categories

3.1 Landscapes and Environments

This hub of image subcategories conveys the beauty of the natural world exploring stunning landforms, ecosystems, and biomes. From vast mountain ranges to intricate forest ecosystems, each subcategory captures unique aspects of Earth's landscapes and environments, offering a comprehensive view of nature's complexity and unique patterns. Figure 2 shows the full tree of these four main categories and all nested leaf image subcategories within WIRA.

- Cities: Offers exploration of urban life across the globe, featuring a range of subcategories dedicated to cities from every corner of the world.
- Coastlines: Showcases the intersections of land and sea, with subcategories highlighting diverse coastal landscapes from around the world.
- Deserts: Delves into desert landscapes, featuring subcategories that explore arid regions across the globe.
- Forests: Shows lush and diverse forests worldwide, with subcategories showcasing everything from rainforests to serene temperate woodlands.
- Mountains: Captures many mountain landscapes, featuring subcategories that span a range of peaks, valleys, and rugged terrains from around the world.

3.2 Life and Portraits

This hub of image subcategories captures people in their everyday lives across cultures and environments. From portraits to candid moments, each collection reveals unique stories that make up human life, offering a rich number of faces and traditions around the world.

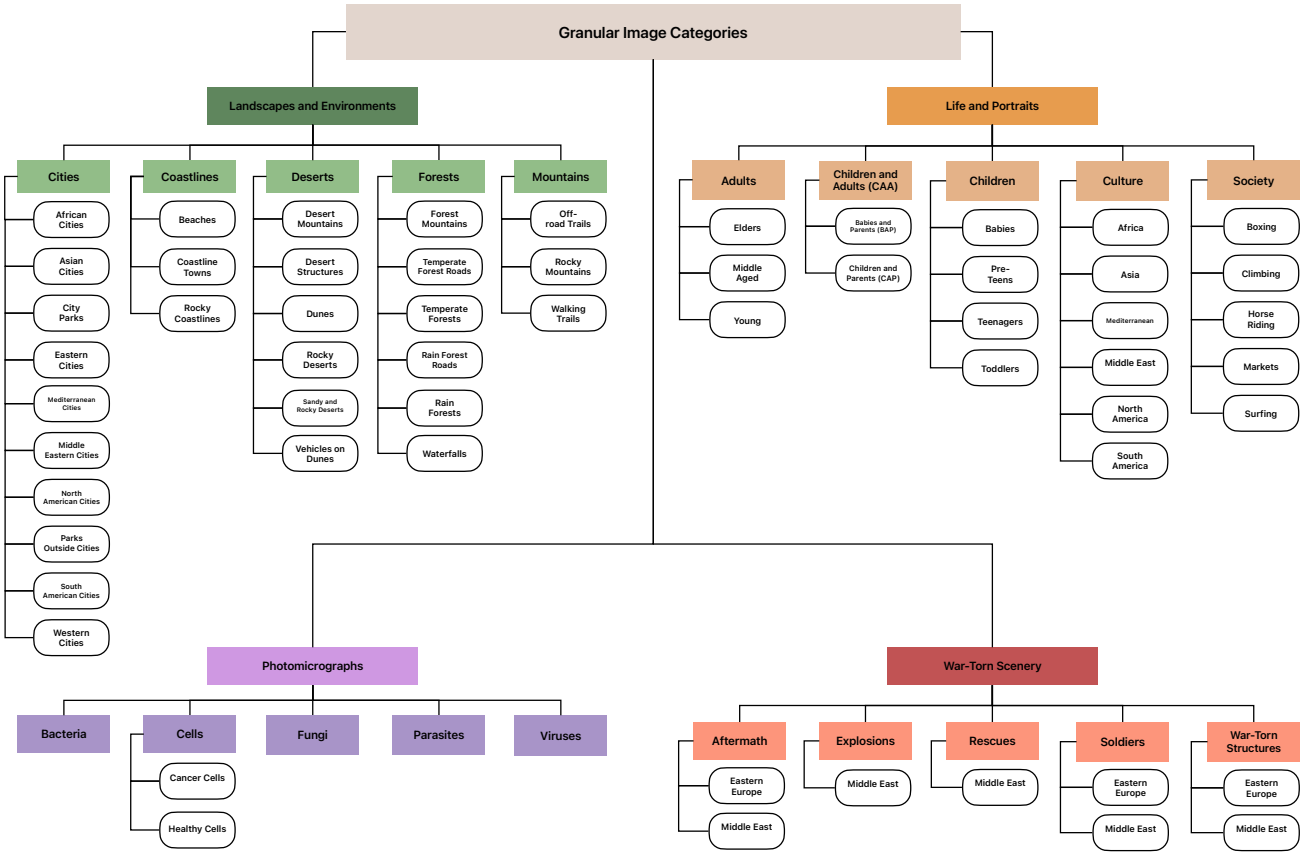

Figure 2: Granular image categories for WIRA, showcasing the main hubs and their detailed subcategories.

- Adults: Highlights adult life, showcasing individuals from diverse backgrounds and cultures.
- Children and Adults: Connects children and adults, portraying mentorship and family experiences.
- Children: Highlights adolescent life across different cultures and settings around the world.
- Culture: Explores rich culture, traditions, attire, and celebrations from unique communities around the world.
- Society: Reflects the structure of communities, capturing scenes of daily life, social interactions, and activities that illustrate how people live, work, and connect.

3.3 Photomicrographs

This hub of image subcategories delves into microorganisms. From detailed views of cellular organisms to intricate patterns in microscopic matter, each collection unveils otherwise hidden complexity.

- Bacteria: Reveals the forms and structures of bacteria.
- Cells: Contains the patterns of healthy plant cell life as well as cancerous cells and the patterns they form.
- Fungi: Captures fungi at a microscopic level, showing unique structures of spores, hyphae, and fungal networks.
- Parasites: Focuses on structures parasitic organisms.
- Viruses: Explores the varied structures of viruses, showcasing unique shapes for infecting host cells.

3.4 War-Torn Scenery

Offers an unflinching look at the devastating impact of war on societies, featuring subcategories that capture the harsh realities of conflict. These images confront viewers with graphic scenes of destruction, loss, and the human suffering that war leaves in its wake, providing a visceral portrayal of the profound toll that conflict exacts on people and places.

- Aftermath: This subcategory captures remnants of conflict, illustrating the devastation it leaves behind.
- Explosions:Focuses on the intense, destructive power of explosions, capturing their smoke clouds and fire.
- Rescues: Highlights moments of bravery and compassion amidst chaos, capturing scenes of people helping others to safety, providing aid, and showing resilience.
- Soldiers: Portrays the experiences of soldiers in various contexts, from intense moments to quieter scenes.
- War-Torn Structures: Shows buildings and infrastructure bearing the destruction inflicted upon once-thriving structures in war zones.

4 Motive For Image Category Selection

The selection of the four main image categories in WIRA; Landscapes and Environments, Life and Portraits, Photomicrographs, and War-Torn Scenery were chosen for their potential to safeguard society against the misuse of hyper-realistic AI-generated images in scenarios that directly impact public trust and safety.

4.1 Landscapes and Environments

With generative AI, a malicious user could create hyper-realistic but non-existent locations. Without verification, such synthetic images could deceive individuals, leading to belief in the existence of fabricated places, potentially endangering lives if the information is used maliciously. This poses exploitation, manipulation, and endangerment to individuals who are led to non-existent destinations.

4.2 Life and Portraits

The misuse of AI to generate non-existent individuals or to portray real people uncharacteristically can create false narratives, impacting public trust. Such synthetic images could also contribute to identity manipulation and deception in social and political arenas. Furthermore, the ability of AI to fabricate human faces could mislead viewers, leading to the belief in the existence of these fabricated entities.

4.3 Photomicrographs

Public trust in scientific imagery is vulnerable to exploitation, as AI-generated images of non-existent pathogens could incite unnecessary fear or panic. This category emphasizes the importance of image authenticity to prevent such misinformation in scientific and medical fields.

4.4 War-Torn Scenery

Synthetic images could be weaponized to mislead the public in war-related contexts. This category is critical in identifying accurate conflict reporting and helps support humanitarian accountability. Malicious uses of generative AI could generate graphic war scenes to falsely depict conflict or suppress real events by portraying peaceful settings in active war zones. This also potentially affects the preservation of accurate historical records.

5 Authenticity Validation Criterion

A comprehensive validation criterion was applied to ensure authenticity with verified origins in each Flickr image used in WIRA. This multi-step process rigorously verifies each image's source, creator information, equipment used, and metadata distinguishing it from datasets lacking any image verification protocols. If any of these items fail to meet the criterion, they each are meticulously documented for transparent analytical review. This analytical record-keeping enhances the dataset's reliability and effectiveness when training AI/ML models to accurately detect hyper-realistic AI-generated images apart from reality. The following outlines the steps in sequential order of the criterion.

5.1 Initial Download and Metadata Collection

The image is initially downloaded using Flickr's API and its associated metadata is recorded.

5.2 Creator Verification

The original creator of each image is identified using the creator's Flickr URL.

- If creator details are unavailable, the image is discarded, and the process resumes with the next image in sequence.
- If creator details are available and verified, the origin is recorded, and the image proceeds to the next stage.

5.3 Ownership Validation

The metadata obtained from the Flickr API is reviewed to confirm that the identified creator is the legitimate owner of the image as indicated by the image's origin URL.

- If ownership cannot be verified, the image is discarded, and the process resumes with the next image in sequence.
- If ownership can be verified, the creator's details are logged, and the image proceeds to the next stage.

5.4 Camera Model Verification

The image's origin URL is reviewed to confirm the inclusion of the camera model used to capture the image, a critical indicator of authenticity.

- If camera model details are missing, the image is discarded, and the process resumes with the next image.
- If the camera model is present, additional verification is performed through Flickr's camera database.
- If the camera model cannot be authenticated, the image is discarded, and the process resumes with the next image.
- If the camera model is authenticated, the information is recorded, and the image proceeds to the next stage.

5.5 Image Similarity Comparison

The downloaded image is compared with the version found on the Flickr origin page using a Structural Similarity Index Measure (SSIM). This confirms that no alterations have been made and verifies the ownership of the image.

- If the SSIM score is below 90%, the image is discarded, and the process resumes with the next image in sequence.
- If the SSIM score is above 90%, the image passes validation and proceeds to the final stage.

5.6 Final Approval and Storage

After meeting all preceding criteria, the image is designated as authentically validated and stored. This process is repeated for each downloaded image until the dataset's required thresholds are achieved.

6 Methodology For Dataset Creation

WIRA was developed through a comprehensive three-part application known as the Real-To-AI Pipeline. The pipeline is flexibly designed to aggregate authentic real images from a search engine of a user's configuration. For the basis of WIRA, the Flickr API was used. The pipeline then creates AI-generated images from Leonardo.AI's commercial API. This pipeline additionally enables users to select different AI models for generated images along with customizable hyperparameters. Beyond simple aggregation, the pipeline ensures authenticity of real images using the specific criterion detailed in section 5. Outlined below are the three main components of the Real-To-AI-Pipeline for WIRA's construction: the Real Image Scraper, the AI Image Captioner, and the Leonardo Image Generator.

6.1 Real Image Scraper

The Real Image Scraper retrieves authentically validated images from Flickr through a multi-step process. First, it queries the Flickr API using customizable search parameters, such as keywords, tags, sorting preferences, and media types. Next, the image processing phase begins, checking each image to ensure it is not a duplicate. Following this, each image's origin and original metadata are documented, providing a traceable history for every image collected. Once the origins are recorded, the previously detailed validation criteria are applied, and if the image passes, it is saved for use.

Querying the Flickr API All images that meet specified criteria are saved in a "GranularImageCategories" directory, with additional subdirectories based on the Flickr query parameters. To manage the scraping process, thresholds are set to automatically stop scraping upon reaching the desired number of validated images. The Flickr API is queried with parameters such as sorting, safe search, and media types to tailor search results. The scraper keeps track of its progress by maintaining a count of images requested from the Flickr API so that duplicate entries are not requested. Once images are returned, the owner's details and image origin URL are recorded.

Duplication Check Before an image is downloaded, it must first pass the Real-To-AI-Pipeline's duplicate image check. For each image response, the Python imagehash library calculates the average, difference, perceptual, and wavelet hashes. These hashes are then cross-checked against their respective hash logs. If no match is found, the hashes are logged respectively to prevent duplicate downloads, avoiding reprocessing the image through the intensive authenticity validation criterion.

Traceable Origins For each new image encountered by the Real-To-AI-Pipeline, all related information including origin data and camera details, is saved to a directory. A manifest is created to facilitate transparent analysis, allowing the identification of each unvalidated image. Each image's path is mapped to a JSON-formatted block containing the original metadata extracted, and organized in a "Scraped Image Manifest" file. The repository path is linked to an image's origin URL in an "All Links Checked" file. Together, these files provide complete traceability records of each image's origin. The unique and documented image is now ready for authenticity validation.

Authenticity Validation This step is guided by the detailed criteria outlined in section 5. Python Selenium is employed to access and confirm attribute values for each image on the Flickr website, ensuring the accuracy and authenticity of the data received from the API.

Validate Creator with Image First, the validation process begins with a Validator module, which creates a headless Selenium Microsoft Edge instance. The Validator uses specific attributes from the current image's response such as the image's ID, user ID, and image URL for its authentication. The creator's Flickr page is located using the user ID obtained from the Flickr API. If the creator's URL is provided, the process continues to the next step. Otherwise the process restarts from Step 1 with the next available image.

Validate Creator on Creator URL Second, the Validator verifies the creator's information on the Flickr website using Selenium. If the creator's name matches the account name displayed on the account page where the image is hosted, the validation process continues. Otherwise, the process restarts from Step 1 with the next available image.

Validate Camera from Image Metadata Third, the Validator retrieves camera EXIF data in the image's Flickr API response. If the camera attribute is present, it is recorded, and the process proceeds to the next step. Otherwise, the creator is flagged in a "Watchlist" file, recording creators who failed validation along with their culprit image URLs and reasonings. In this case the reason, "No camera listed within image metadata," is appended after the image URL, separated by " - " to ensure the URL remains intact. This entry is added to a list of failed image URL-reason pairs regarding the specific creator. The URL is also mapped to a local path in a "Failed" file, as the initial image is downloaded and saved separately from WIRA. The process then restarts from Step 1 with the next available image.

Validate Camera on Flickr Camera Database Fourth, with validated creator and recorded camera data, the Validator verifies the camera's authenticity using Flickr's camera database. Flickr maintains a verified database with detailed descriptions for each recognized camera. For images that include a camera in the Flickr API response, there is typically a link on the image's origin page the Validator searches for that leads to the camera's description. In most cases, this link is present; however, some photographers may use unverified cameras not listed in Flickr's camera database. This step ensures that only images with Flickr-validated cameras proceed to the next step. Otherwise, the creator is added to the "Watchlist" file, noting the image URL and the reason for failure "Camera could not be validated on Flickr page". The image URL is also mapped to its local downloaded path in the "Failed" file. The process then restarts from Step 1 with the next available image.

Validate Local and Creator's Images by SSIM Fifth, once the creator and camera are validated, the downloaded image must be identical to the image displayed on the creator's Flickr account. Occasionally, discrepancies arise between the image provided by the Flickr API and the one displayed on its original page, often due to slight modifications such as watermarks, borders, or minor edits. To address this, a Structural Similarity Index Measure (SSIM) is calculated between the two images using the Python Scikit-Image Metrics library. An SSIM range of 95%-100% typically signifies that the images are visually identical, with any variations likely due to minor artifacts or compression. Scores between 85%-95% suggest small edits or adjustments, while scores below 85% indicate significant structural or visual differences, suggesting the images are not the same. For accurate comparison, both images are resized to match the dimensions of the smaller image. A threshold of 90% SSIM was selected to allow for minor modifications such as watermarks or borders that photographers might add for copyright purposes. If the SSIM score meets or exceeds 90%, the validation proceeds. If not, the creator is added to the "Watchlist" file, with the image URL recorded alongside the reason for failure "Image downloaded is not visually the same as the image on Creator page based on SSIM scoring". The URL is also mapped to its local path in the "Failed" file. The process then restarts from Step 1 with the next available image.

Complete Validation and Traceable Origins Since the creator, camera, and image are now successfully validated, the authenticity criterion is fully met. The creator is appended to a file titled "Criteria Success List," with the validated image URL appended to the creator's list of previously validated image URLs. The image URL is also mapped to its local path in the "Passing" file. The image is then saved to the "GranularImageCategories" directory. This process continues until reaching the image threshold.

6.2 AI Image Captioner

The AI Image Captioner accepts image requests and returns captions, providing descriptive context for each image. The main components of the AI Image Captioner, which are further detailed in the following subheadings, include the API, the AI captioning, and the containerization processes.

Creating the Flask API The Python Flask library is used to build a simple API with two primary methods: one for general image captioning and another specifically for handling photomicrographs. The rationale for these separate methods lies in the need to provide contextual prompts. For most images, the API sends a prompt to the captioning model asking for a detailed description without specifying the image type, allowing the model to infer its content. For photomicrographs the llama-3-vision-alpha-hf model requires specific context for accurate descriptions. Otherwise it struggles to interpret the content correctly. The API operates by opening a web socket that receives HTTP POST requests with an image attachment. Upon startup, it initializes the AI Image Captioner. The API route for general image captioning is "/caption," while photomicrograph images are sent to "/caption-photomicrograph."

The Captioning Model When the Flask API initializes, it creates an instance of the AI "Captioner" class. The Captioner is configured to run offline, ensuring that the model is fully tokenized and loaded within its container without the need for internet. If a cached model is missing or corrupted, the Captioner can detect this and attempt to retrieve the latest version from its original repository. Upon successfully downloading the latest model version, it creates a new cache directory to store its safe-tensor shards. Once the model is ready, the API is ready to accept image caption requests.

Containerization Docker is used to containerize the Flask API and AI Captioner components, creating a cohesive and scalable application. The Docker container uses the official python:3.11.8-slim image. Necessary PyTorch and CUDA libraries for GPU interaction are installed to the container from PyTorch's cu124 library. To support the synchronization between the Docker container and the host machine's NVIDIA GPU, WSL2 is used for the Docker Desktop backend. The Docker Compose file is preconfigured to ensure compatibility with an NVIDIA GPU on the host OS. The project's virtual environment dependencies are defined in a requirements file, which the container uses to install the remaining needed libraries. Contents of the AI Image Captioner directory, including the cached model, are then written into the container. Once setup is complete, the container is launched, starting the Flask API and instantiating an instance of the Captioner model, which then awaits image POST requests. Upon captioning an image, the Captioner model sends a response containing the generated caption, which can be stored for future use.

6.3 Leonardo Image Generator

The third and final component of the Real-To-AI-Pipeline is the Leonardo Image Generator. Once all images are aggregated into the "GranularImageCategories" directory according to the thresholds set by the Real Image Scraper, each image is sent via HTTP POST to the AI Image Captioner container. The captions generated for each image are then recorded. After all images are captioned, each image and its corresponding caption are submitted to the Leonardo.AI commercial API to produce a hyper-realistic AI-generated counterpart. The API is polled until the AI-generated image is ready, at which point it is saved locally to a designated "AI" directory. The following subheadings provide a detailed, sequential overview of this process.

Captioning Images from Image Directory Paths All image paths for each image subcategory are recorded in a file titled "Directory Paths". To handle photomicrographs, the paths for each subcategory within the Photomicrographs directory are specifically recorded in a file named "Photomicrograph Paths". Once the Captioner generates a caption for a given image, it is saved in an "Images Captioned" file and mapped to the image's path. This mapping ensures all images are captioned and allows each image-caption pair to be sent to the Leonardo.AI commercial API.

Generating Hyper-Realistic AI Images After all images have been captioned, the Directory Path and Photomicro-

graph Path files are used to locate an image-caption pair for submission to the Leonardo.AI commercial API. First, a pre-signed URL is requested from Leonardo.AI to send an authenticated image generation request. The image-caption pair is then sent to Leonardo.AI's "Image to Image" generation feature with the caption as the prompt. To ensure consistency between images and their AI counterparts, the AI-generated image's dimensions are configured to match the original image's height and width, maintaining the aspect ratio between the two. The model selected for generating images is the Leonardo Vision XL model. Further details on model selection are provided in section 9. The Leonardo.AI API is polled to track the generation status. Once an AI-generated image is ready, it is downloaded and saved to the "AI" local directory. A "Main Manifest" file records the local path of each AI-generated image and maps it to the original image's local path, enabling analytical comparisons between paired images. This process iterates over all captioned images until each has a hyper-realistic AI counterpart. This completes the Real-To-AI-Pipeline and finalizes the WIRA dataset creation.

7 Authenticity Validation And Traceable Origins For Photomicrographs

As noted in section 4, the Photomicrographs category does not apply the main Authenticity Validation and Traceable Origins criterion described in sections 5 or 6. This decision was made due to the lack of mainstream capability on Flickr for photographers to record specific tools, such as microscopes, within Flickr's camera database. Consequently, a modified approach was applied for authenticity validation within the Photomicrographs category. For WIRA's transparency, all sources for the Bacteria, Cancer Cells, Healthy Cells, Fungi, Parasites, and Virus subcategories are all thoroughly cited for transparency, ensuring their Traceable Origins. These subcategories contain images exclusively aggregated by hand from reliable sources, including the CDC's Public Health Image Library, the Broad Institute's Broad Bioimage Benchmark Collection, the Image Data Resource for Open Microscopy, and IAQ Consultants. Where available, each image from these sources is further documented with its original publication reference. This method provides a transparent and traceable foundation for the photomicrographs included in WIRA. Please note that the citations of all individual images or image datasets including their urls and origin publication where available are cited within the GitHub repository.

7.1 Bacteria

All photomicrographs of bacteria were collected from the DAS+4tag_Trial2 images from IDR located on the DOI organization online. This subset of images originates from a study containing photomicrographs of E. Coli bacteria (Ali et al. 2020). Other individual images were hand collected from the CDC PHIL with no link to an original publication.

7.2 Cancer Cells

All photomicrographs of cancer cells were collected from the BBBC, originating from BBBC001 and BBBC0018 (Moffat et al. 2006). BBBC006 is used but does not have a direct link to an original publication.

7.3 Healthy Cells

All photomicrographs of healthy cells were collected from the BBBC along with image datasets from IDR. BBBC009 is used but does not have a direct link to an original publication. AT1G02730 and AT1G05570 are from IDR but originate from a separate study (Yang et al. 2016). Diplophyllum taxifolium and Scapania mucronate are from IDR but originate from a separate study (Peters and König-Ries 2022).

7.4 Fungi

All photomicrographs of fungi were collected from the CDC PHIL and IAQ Consultants without links to original publications.

7.5 Parasites

All photomicrographs of parasites were collected from the BBBC. BBBC010 is used and from a separate study (Moy et al. 2009). BBBC041 is used but does not have a direct link to an original publication.

7.6 Viruses

All photomicrographs of viruses were collected from the IDR, some of which do not contain any original publication. These are the Zb_BSF019089, BSF019243-1A, and preScreen datasets. BSF018307-4D image is used and from a separate study (Georgi et al. 2020).

8 Computational Resources For WIRA Construction

This section presents the specific hardware and software resources used in the construction of the WIRA dataset, which was developed entirely on a local machine. Including these details ensures transparency and supports reproducibility for researchers who may wish to replicate or extend this work without relying on cloud resources. Table 1 displays the hardware specifications while Table 2 displays the software environmeent of the machine used to construct WIRA.

9 WIRA Decsion Records

This section presents the decisions made throughout the creation of WIRA, detailing both accepted and rejected choices along with the rationale behind each. Organized chronologically, it provides explanations for each decision, offering a transparent view of WIRA's development.

9.1 Third-Party Software to Validate Images

Third-party software, such as APIs like isitai.com, was initially considered to streamline the validation of web-scraped images by detecting anomalies indicative of AI-generated content, thereby expediting the authenticity validation process. It was determined that such tools should not be part of

Component	Specification
Machine	Dell Precision 7770
Processor	12th G. Intel i7-12850HX 2.10GHz
Installed RAM	64.0 GB DDR5 4800MHz CAMM
System Type	64-bit OS, x64-based processor
Integrated GPU	Intel UHD Graphics, 32.0 GB
Discrete GPU	NVIDIA RTX A1000, 4GB GDDR6

Table 1: Hardware Specifications

Component	Specification
Operating System	Windows 11 Pro
OS Version / Build	23H2 / 22631.446
Code Editor	VS Code
Program Language	Python 3.11.8
AI/ML Backend	PyTorch 2.5.1
CUDA Version	12.4
Containerization	Docker Desktop v4.34.3
Docker Image	python:3.11.8-slim
WSL version	2.2.4.0
AI Image Generator	Leonardo.AI API v1.0
Metadata Tool	ExifTool by Phil Harvey 12.96

Table 2: Software Environment

the authenticity validation process, as they do not provide insight to an image's origin. Relying on a third party for validation could compromise the credibility of authenticity, especially as the Real-To-AI Pipeline already depends on the search engine as a third party for initial image sourcing.

9.2 Image Captioning Websites

Websites like pallyy.com can be used to automatically caption images, which is an essential component of the Real-To-AI Pipeline. These tools were found to produce subpar results when compared to the llama-3-vision-alpha-hf model.

9.3 Source of Images Scraped

Initially, Google was selected as the primary source for scraping images. It became clear that using a custom Google search engine would better streamline the web-scraping process. Despite this, challenges persisted in maintaining accountability for image sources on Google. Identifying the original author of an image was rare and verifying if an image was captured by a camera proved difficult. Reverse image searches often failed to provide the oldest publication date, as some entries lacked this data. These limitations made the web-scraping process inefficient for aggregating authentically validated images. Consequently, Flickr was chosen due to its robust API, which supports thorough investigation into the source and origin of each image. This ensured that if an image is later determined to be non-authentic despite if it were to pass the Authenticity Validation criteria, accountability would rest solely on the photographer, not the search engine. The combination of the validation criteria and Flickr's platform reinforced the authenticity of images, enabling each to be traced back to its photographer, who attests to its authenticity.

9.4 Image Metadata for Authenticity Validation

While tools like the ExifTool make it easy to access an image's metadata, they equally allow for metadata manipulation. Initially, metadata was considered a primary factor for determining an image's authenticity; however, a malicious actor could use the same tool to alter metadata on an AI-generated image. Consequently, metadata is now utilized solely for analytical purposes and does not play a role in any stage of the Authenticity Validation criteria.

9.5 Using Cloud Computing Architecture

Due to time and funding constraints during the research and development of WIRA, implementing a cloud computing architecture was not feasible. As outlined in section 10, future integration of cloud architecture could enhance WIRA.

9.6 Transparency of WIRA

WIRA is designed to maintain complete transparency, allowing users to easily critique or validate its contents. For each successful image, the photographer assumes full responsibility for ensuring the content they produce is authentically real. Researchers may use the analytical files detailed in section 6 and available in the GitHub repository, to analyze each image. Without this transparency, determining "what is real anymore" would be impossible.

9.7 Choice of Leonardo.AI Model

Through rigorous testing of various parameter settings for the Leonardo.AI commercial API, many models were tested. Figure 3 shows a comparison of real images to a sample of model and parameter combinations used to determine the final image-generation model for WIRA. A larger figure can be found on the GitHub repository. The Leonardo Kino XL and Leonardo Vision XL models performed exceptionally well showing hyper-realistic counterpart images comparatively to the real images sampled. The Leonardo Vision XL model was selected after it was found to produce fewer anomalies compared to the Leonardo Kino XL model.

10 Enhancements For Expanding WIRA

This section outlines future enhancements for the WIRA dataset, focusing on upgrades that can extend its applicability beyond AI-generated image detection to a broader range of AI/ML solutions. The following suggestions aim to increase the scalability and adaptability of WIRA, making it versatile for diverse applications.

10.1 SSIM Scoring Optimization

To improve SSIM scoring accuracy, the comparison process should resize the local image to match the dimensions of the reference image, rather than resizing both images to the smallest dimensions of either image.

10.2 Cloud Computing Integration

A cloud architecture would support the dataset's scalability, allowing large-scale image aggregation and storage.

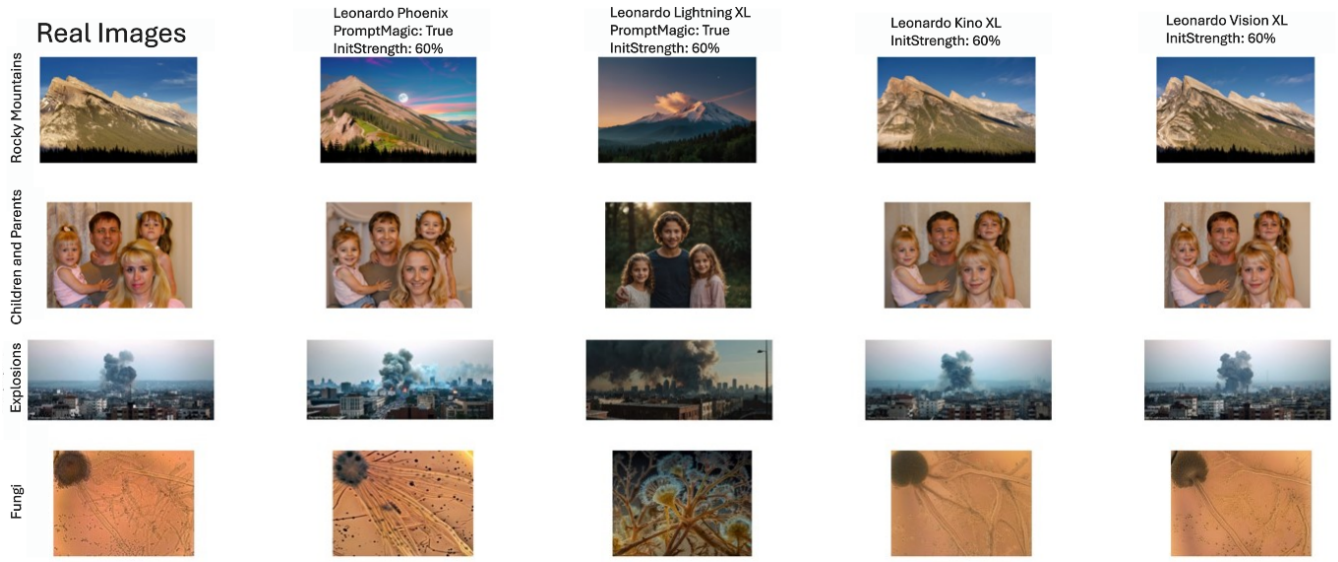

Figure 3: Showcasing different categories within WIRA, the first column represents authentically real images validated from the Authenticity Validation steps. The following columns represent different model and parameter outputs used when generating hyper-realistic AI counterpart images from the Leonardo.AI commercial API.

10.3 Image Captioning Standards

For the image captioning process, ensuring the captioning model generates clear, contextually appropriate descriptions that adhere to AI moderation standards, such as those established by Leonardo.AI, to maintain ethical and safe content generation.

11 Conclusion

The novel "What Is Real Anymore?" (WIRA) dataset is an authentic image dataset curated for AI-generated image detection. By incorporating a rigorous authenticity validation process and traceable origins, WIRA addresses critical gaps in mainstream datasets by ensuring each image is authentically validated, and its source transparent. WIRA fills a vital need in AI research, where distinguishing between authentic and synthetic imagery becomes increasingly challenging due to the sophistication and hyper-realism of AI-generated visuals. Future enhancements like cloud computing integration and refined image similarity measures will enable the expansion of WIRA, making it adaptable to diverse research applications. In conclusion, WIRA provides the AI/ML community with a trusted resource for advancing AI-generated content detection, promoting integrity of digital imagery in an era of increasing visual manipulation, and setting a benchmark for the ethical curation of authentically real data. As the digital landscape continues to evolve, datasets like WIRA will remain instrumental in upholding public trust in visual content and contribute to the defense of innocent individuals against adverse uses of generative AI across the globe.

References

Aksac, A.; J. Demetrick, D.; Ozyer, T.; and Alhajj, R. 2019. BreCaHAD: A Dataset for Breast Cancer Histopathological Annotation and Diagnosis. *BMC Research Notes*, 12(82).

Ali, M. Z.; Parisutham, V.; Choubey, S.; and Brewster, R. C. 2020. Inherent regulatory asymmetry emanating from network architecture in a prevalent autoregulatory motif. *eLife*, 9: e56517.

Anokhin, I.; Demochkin, K.; Khakhulin, T.; Sterkin, G.; Lempitsky, V.; and Korzhenkov, D. 2021. Image Generators with Conditionally-Independent Pixel Synthesis. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14273–14282.

Bird, J. J.; and Lotfi, A. 2023. CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images. arXiv:2303.14126.

Choi, Y.; Uh, Y.; Yoo, J.; and Ha, J.-W. 2020. StarGAN v2: Diverse Image Synthesis for Multiple Domains. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8185–8194.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. arXiv:1604.01685.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Epstein, D. C.; Jain, I.; Wang, O.; and Zhang, R. 2023. Online Detection of AI-Generated Images. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 382–392.

Georgi, F.; Kuttler, F.; Murer, L.; Andriasyan, V.; Witte, R.; Yakimovich, A.; Turcatti, G.; and Greber, U. F. 2020. A High-Content Image-Based Drug Screen of Clinical Compounds Against Cell Transmission of Adenovirus. *Scientific Data*, 7(1): 265.

Karras, T.; Aila, T.; Laine, S.; and Lehtinen, J. 2018. Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv:1710.10196.

Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; and Aila, T. 2020. Training Generative Adversarial Networks with Limited Data. arXiv:2006.06676.

Karras, T.; Aittala, M.; Laine, S.; Härkönen, E.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2021. Alias-Free Generative Adversarial Networks. arXiv:2106.12423.

Karras, T.; Laine, S.; and Aila, T. 2021. A Style-Based Generator Architecture for Generative Adversarial Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12): 4217–4228.

Krizhevsky, A. 2009. *Learning Multiple Layers of Features from Tiny Images*. Master's thesis, University of Toronto. [Online]. Available: https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Lin, M.; Shang, L.; and Gao, X. 2023. Enhancing Interpretability in AI-Generated Image Detection with Genetic Programming. In *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, 371–378.

Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C. L.; and Dollár, P. 2015. Microsoft COCO: Common Objects in Context. arXiv:1405.0312.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep Learning Face Attributes in the Wild. arXiv:1411.7766.

Lorenz, P.; Durall, R. L.; and Keuper, J. 2023. Detecting Images Generated by Deep Diffusion Models using their Local Intrinsic Dimensionality. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 448–459.

Luo, Y.; Du, J.; Yan, K.; and Ding, S. 2024. LaRE2: Latent Reconstruction Error Based Method for Diffusion-Generated Image Detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17006–17015.

Moffat, J.; Grueneberg, D. A.; Yang, X.; Kim, S. Y.; Kloepfer, A. M.; Hinkle, G.; Piqani, B.; Eisenhaure, T. M.; Luo, B.; Grenier, J. K.; Carpenter, A. E.; Foo, S. Y.; Stewart, S. A.; Stockwell, B. R.; Hacohen, N.; Hahn, W. C.; Lander, E. S.; Sabatini, D. M.; and Root, D. E. 2006. A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell*, 124(6): 1283–1298.

Monkam, G.; Xu, W.; and Yan, J. 2023. A GAN-based Approach to Detect AI-Generated Images. In *2023 26th ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*, 229–232.

Moy, T. I.; Conery, A. L.; Larkins-Ford, J.; Wu, G.; Mazitschek, R.; Casadei, G.; Lewis, K.; Carpenter, A. E.; and Ausubel, F. M. 2009. High-Throughput Screen for Novel Antimicrobials Using a Whole Animal Infection Model. *ACS Chemical Biology*, 4(7): 527–533.

Peters, K.; and König-Ries, B. 2022. Reference Bioimaging to Assess the Phenotypic Trait Diversity of Bryophytes Within the Family Scapaniaceae. *Scientific Data*, 9(1): 598.

Rahman, M. A.; Paul, B.; Sarker, N. H.; Hakim, Z. I. A.; and Fattah, S. A. 2023. ArtiFact: A Large-Scale Dataset with Artificial and Factual Images for Generalizable and Robust Synthetic Image Detection. arXiv:2302.11970.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; and Fei-Fei, L. 2015. ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575.

Russell, B. C.; Torralba, A.; Murphy, K. P.; and Freeman, W. T. 2008. LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision*, 77(1): 157–173.

Schuhmann, C.; Vencu, R.; Beaumont, R.; Kaczmarczyk, R.; Mullis, C.; Katta, A.; Coombes, T.; Jitsev, J.; and Komatsuzaki, A. 2021. LAION-400M: Open Dataset of CLIP-Filtered 400 Million Image-Text Pairs. arXiv:2111.02114.

Sun, Y.; Wang, X.; and Tang, X. 2013. Hybrid Deep Learning for Face Verification. In *2013 IEEE International Conference on Computer Vision*, 1489–1496.

Sun, Y.; Wang, X.; and Tang, X. 2014a. Deep Learning Face Representation by Joint Identification-Verification. arXiv:1406.4773.

Sun, Y.; Wang, X.; and Tang, X. 2014b. Deep Learning Face Representation from Predicting 10,000 Classes. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1891–1898.

Wang, H.; Fei, J.; Dai, Y.; Leng, L.; and Xia, Z. 2023a. General GAN-generated Image Detection by Data Augmentation in Fingerprint Domain. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, 1187–1192.

Wang, S.-Y.; Wang, O.; Zhang, R.; Owens, A.; and Efros, A. A. 2020. CNN-generated images are surprisingly easy to spot... for now. arXiv:1912.11035.

Wang, Z.; Bao, J.; Zhou, W.; Wang, W.; Hu, H.; Chen, H.; and Li, H. 2023b. DIRE for Diffusion-Generated Image Detection. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 22388–22398.

Xia, W.; Zhang, Y.; Yang, Y.; Xue, J.-H.; Zhou, B.; and Yang, M.-H. 2022. GAN Inversion: A Survey. arXiv:2101.05278.

Xiao, J.; Hays, J.; Ehinger, K. A.; Oliva, A.; and Torralba, A. 2010. SUN database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492.

Yang, W.; Schuster, C.; Beahan, C. T.; Charoensawan, V.; Peaucelle, A.; Bacic, A.; Doblin, M. S.; Wightman, R.; and Meyerowitz, E. M. 2016. Regulation of Meristem Morphogenesis by Cell Wall Synthases in *Arabidopsis*. *Current Biology*, 26(11): 1404–1415.

Yu, F.; Seff, A.; Zhang, Y.; Song, S.; Funkhouser, T.; and Xiao, J. 2016. LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop. arXiv:1506.03365.

Zhan, F.; Yu, Y.; Wu, R.; Zhang, J.; Lu, S.; Liu, L.; Kortylewski, A.; Theobalt, C.; and Xing, E. 2023. Multimodal Image Synthesis and Editing: The Generative AI Era. arXiv:2112.13592.

Zhang, M.; Wang, H.; He, P.; Malik, A.; and Liu, H. 2022. Improving GAN-Generated Image Detection Generalization Using Unsupervised Domain Adaptation. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; and Oliva, A. 2014. Learning deep features for scene recognition using places database. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'14, 487–495. Cambridge, MA, USA: MIT Press.

Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; and Torralba, A. 2017. Scene Parsing through ADE20K Dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5122–5130.

Zhu, M.; Chen, H.; Yan, Q.; Huang, X.; Lin, G.; Li, W.; Tu, Z.; Hu, H.; Hu, J.; and Wang, Y. 2023. GenImage: A Million-Scale Benchmark for Detecting AI-Generated Image. arXiv:2306.08571.