This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.
SAND2022-7446C

Performance Testing of Commercially Available Tools for Logo and Text Detection

Sydney Dorawa,[1] Zoe Gastelum,[1] Yana Feldman,[2] and Teagan Zuniga[2]

1. Sandia National Laboratories. Albuquerque, New Mexico, USA

2. Lawrence Livermore National Laboratory. Livermore, California, USA

Abstract: Recent increases in performance of open-source computer vision models and the availability of enterprise or cloud-based computational power has made computer vision more tenable than ever before. Furthermore, the volume and diversity of open-source information to be reviewed by the International Atomic Energy Agency (IAEA) for potential safeguards relevance continues to grow – both providing an opportunity for deeper safeguards evaluation and creating an overwhelming burden on safeguards analyst resources. For the past five years, Lawrence Livermore National Laboratory and Sandia National Laboratories have been evaluating opportunities, technical trends, and capabilities related to the use of computer vision tools to support open-source information collection and analysis for IAEA safeguards. New tools to detect and identify logos and text within images could provide a new level of support to safeguards analysts seeking clues regarding a state's nuclear activities beyond more traditional capabilities like image classification or object detection. In this paper, we will present findings from a recent evaluation of an open-source computer vision platform on logo and text identification, including evaluation on specially curated nuclear-relevant images containing text and logos.

Introduction

Open-source information is an important data stream for the International Atomic Energy Agency (IAEA) Department of Safeguards that provides context, background, and additional indicators of nuclear activities at the state level. One of the challenges associated with the collection and analysis of open-source information is the large quantity of visual, audio, and text data that is produced every day. Analysts have limited time to sort information and determine what might be relevant. Furthermore, retrieval of non-textual information such as photographs and videos still largely relies on text-based search strings and image/video labels and meta data, which we know might be incorrectly labeled (which could result from lack of expertise, translation, or intentional mislabeling) or incompletely labeled (not containing information relevant for safeguards if relevant objects are in the background, or not fully describing all content present in a long video clip). This poses the risk for safeguards-relevant information to be overlooked.

Recently, artificial intelligence (AI) has been proposed to sort and prioritize large volumes of information to ease analysts' burden. One such capability is logo detection. Logo detection is a subset of computer vision that can support analysts in locating and identifying commercial logos in still images and video. Multiple commercial platforms have developed logo detection capabilities that are available today. A related capability is text detection within images. Like optical character recognition that is used to digitize historical documents, text detection can

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

locate and identify text within real-world environments such as text on signage, clothing, vehicles, etc.

In this paper, we will describe a recent test conducted as a collaboration between Lawrence Livermore National Laboratory (LLNL) and Sandia National Laboratories (SNL) of Google's text and logo detection capabilities that are available via the free Vision AI platform. We specifically evaluated the platform's performance on nuclear-related logos and text compared to general information to determine the level of development that might be required to deploy the current capabilities for an open-source safeguards use case.

Data

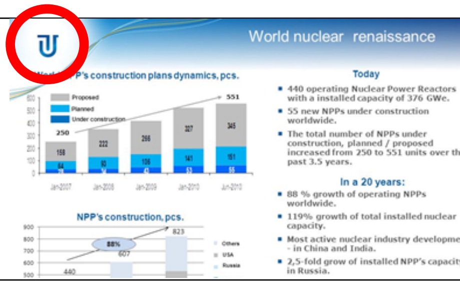
The research team collected and labeled 283 images for this logo detection testing activity. The images were collected from a variety of sources, such as the websites of nuclear-related and utility companies, nuclear-related journals and advertisements, and pages or publications from international nuclear-focused meetings or conferences. A "ground truth," or a human-generated label of the image against which Vision AI's predictions were compared, was created for each image. The labeling included information about the logo's position in the image, its classification as general or nuclear-related, the presence of text in an image, and a transcription of the text.

Of the 283 images, 59 were general domain images from areas such as transportation, medicine, dining, and entertainment. The remaining 224 images were nuclear-relevant. Both general and nuclear-relevant images include English and non-English language examples. For nuclear and general-domain images, we had four broad classes of images:

1. Isolated logos (images cropped to the edge of a logo) that included text
2. Isolated logos (images cropped to the edge of a logo) that did not include text
3. Images that contained logos ("environmental logos") that contain text in the logo. Most of these images also included text elsewhere in the image.
4. Images that contained logos ("environmental logos") that did not contain text in the logo. Most of these images also included text elsewhere in the image.

Example images from each category are provided in Table 1Table 1 Description of Logo Image Classes. Due to the manner in which our ground truth data was labeled, we conducted our analysis across these four categories for nuclear-relevant logo images and conducted a compiled analysis for general domain images across the four categories. In Table 1 we define the image count for both nuclear and general images to illustrate the distribution in our test set. Note that this table counts images in which different types of logo classes appear, while some of our analyses instead count the total number of potential logos to detect within the image.

Table 1 Description of Logo Image Classes

Description & Image Count	Nuclear Example	General Example
Logo without text Nuclear: 50 General: 5		
Logo with text Nuclear: 100 General: 13		
Environmental logo with text Nuclear: 51 General: 30		
Environmental logo without text Nuclear: 23 General: 11		

Methods

The team evaluated Google's Vision AI (hereafter Vision AI) platform[1] via the web browser interface. We uploaded one image at a time into the platform and collected the output from two different computer vision capabilities offered within Vision AI: text identification and logo identification.

We evaluated Vision AI on two metrics. First, we evaluated the platform's ability to detect the presence and location of a logo or text in an image. Then, we evaluated the platform's ability to

[1] https://cloud.google.com/vision

correctly identify or transcribe text. We performed this analysis across three categories of image content: logos, text-in-logos, and environmental text (i.e., outside of logos, in the broader image). The metrics for each type of image content are detailed below.

Logo Detection and Identification

The first metric we assessed was the platform's ability to correctly detect and locate a logo. For each trial in which the platform detected a logo, it placed a bounding box around the logo with a label of the predicted identification. It also produced a list of detected logos with predicted identifications and associated confidence for each prediction. See Figure 1.

A trial was coded as correctly *detecting* the logo if it accurately identified the location of the logo of interest in an image, or if in images with multiple instances of a logo of interest, Vision AI identified the location of at least one of them. A trial was coded as an unsuccessful attempt if Vision AI did not recognize that there was a logo in the image, if Vision AI identified a part of the image as a logo that was not, or if it missed the logo of interest but located a different, secondary logo.

A trial was coded as correctly *identifying* a logo if Vision AI accurately identified the logo of interest via the bounding box label and logo list. In cases where more than one logo of interest was present, we coded the response as correct if Vision AI successfully identified at least one of them.[2] We coded the response to this question as incorrect if Vision AI did not locate a logo in the image at all or if Vision AI located the logo(s) but did not correctly identify any of them.

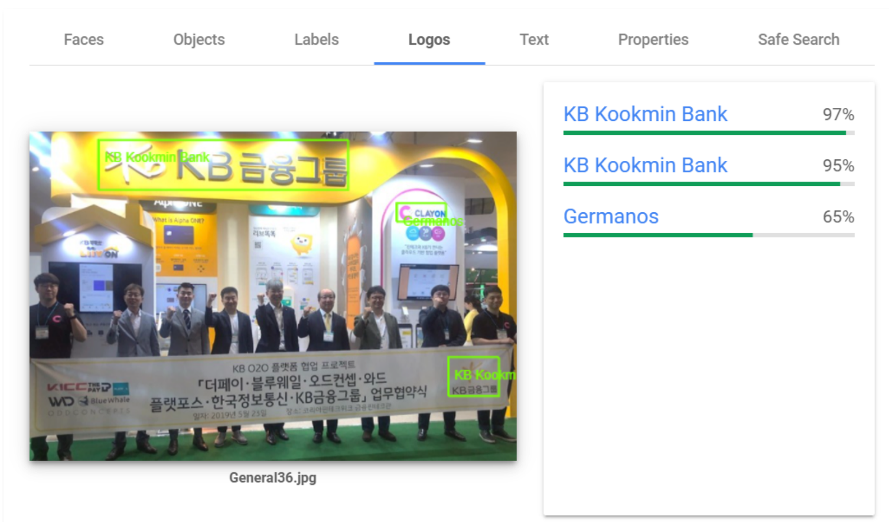


Figure 1 Example Environmental Logo Detection using Vision AI

Text-in-Logo Detection and Identification

Text-in-logo refers to logos which contain text as an integral part of the logo (i.e., directly adjacent to or part of a logo symbol). When Vision AI detected text anywhere in an image,

[2] If the name of the company was partially or fully correctly identified, it was considered correct.

including text within a logo, the platform highlighted the text in-situ and transcribed the text identification. See Figure 2.

We coded a correct response for a text-in-logo *detection* trial if Vision AI highlighted the text within the logo. We coded an incorrect response for a trial if Vision AI did not detect the text in the logo, regardless of other text detections within the image.

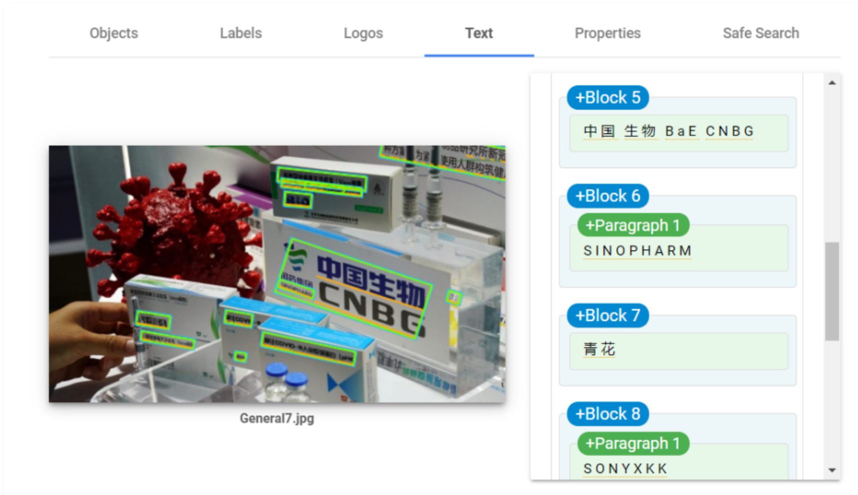


Figure 2 Example Environmental Text Detection and Identification

The evaluation of text identification – both in-logo and environmental – required more subjectivity than other assessments based on the way the data was presented. In its text identification, Vision AI often added spacing between letters, sometimes translated text from foreign languages, and in some cases languages with non-Latin alphabets required close visual inspection by the researcher to determine if the characters in the test image matched the platform's response.[3] In-logo text *identification* trials were coded as correct if Vision AI identified all or a majority of the letters or words in a logo (even when the text did not match exactly but the words or meaning of the text was still understandable), or, for non-Latin alphabets (e.g., Korean, Mandarin, Arabic, and Hebrew), the text in the logo appeared to match what Vision AI identified. If Vision AI mistook one letter for another, such as a C for an O, but got the other letters correct, we coded the trial as correct unless there were so few letters that such a mistake made the text unidentifiable as being part of a specific logo.[4] Trials were coded as incorrect if Vision AI incorrectly identified the text, including if the Vision AI-produced text was unreadable, or for non-Latin alphabets, if the text did not appear to match the original characters.

[3] The evaluator was a native-English speaker, and thus was more easily able to discern the Latin alphabet characters.
[4] For example, in one instance, Vision AI read the name "Orano" in the logo as "Areno," which was different enough that it could not be considered correct, so we coded it as an incorrect text identification. As another example, Vision AI identified the text "GP GENERAL PLASTICS MANUFACTURINO COMANY" from a logo for General Plastics Manufacturing Company. In this case, the mistaken and missing letters did not significantly impact our ability to understand the text, and this was considered an acceptable error and coded as a correct text identification.

Environmental Text Detection and Identification

Environmental text refers to text present in an image that was not associated with a logo in the image. In many cases, environmental images contained both text associated with logos and non-logo text. The Vision AI platform labels text with bounding boxes and text transcriptions, and it also transcribes text in a separate Text tab. See Figure 2.

Environmental text *detection* trials in which Vision AI correctly placed bounding boxes around environmental text and made a transcription (regardless of accuracy) were coded as correct. Incorrect trials were those for which the environmental text was not highlighted or transcribed, including trials when Vision AI detected other text in logos but did not detect the environmental text.

Trials in which Vision AI correctly *identified* most of the environmental letters or words were coded as correct. We coded trials as incorrect if the platform failed to identify significant portions of text, misread more than approximately a quarter of the letters and words, or re-ordered the text beyond recognition.

Environmental text detection and identification assessments were more challenging than the logo detection and identification because there was more text in the images, the text was typically smaller, and the text sometimes was present in difficult-to-interpret locations (e.g., at an angle, or partially obscured). In this assessment, we observed that Vision AI sometimes failed to identify text on low-quality images.

The images that did not contain text are included in our testing as a baseline. For these, no detection of text is the correct model response. In selected cases where the logo itself appeared as text (e.g., the Westinghouse logo is a stylized “W”), we did accept text detection and identification as correct if the associated text was correct. Other detection of text in images with no apparent text were coded as incorrect.

Results

Logo Detection and Identification Performance

The results of our assessments for logo detection and identification are presented in Figure 3. Average logo *detection* accuracy across all categories was 68.9%, and average logo *identification* accuracy was 35.5%.

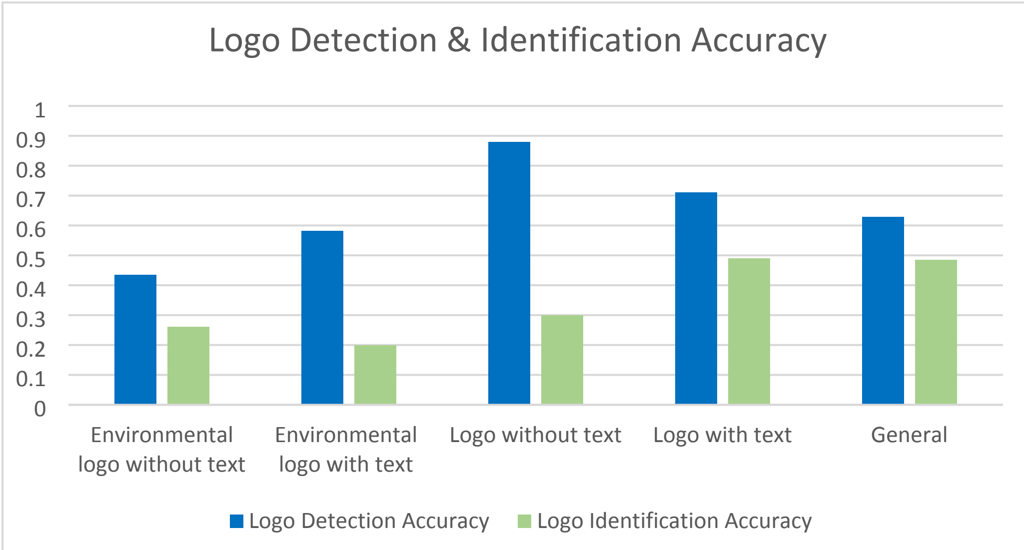


Figure 3 Vision AI Performance on Logo Detection and Identification

Text-in-Logo Detection and Identification Performance

Vision AI's performance on text detection and identification for text within logos is shown in Figure 4. Three of the image classes (environmental logo without text, isolated logo without text, and general) had smaller sample sizes for the text-in-logo assessment than the logo detection/identification task because not all images had text. Average text-in-logo detection accuracy across all nuclear-relevant image classes was 95.2%; average text-in-logo text identification accuracy was 74.9%.

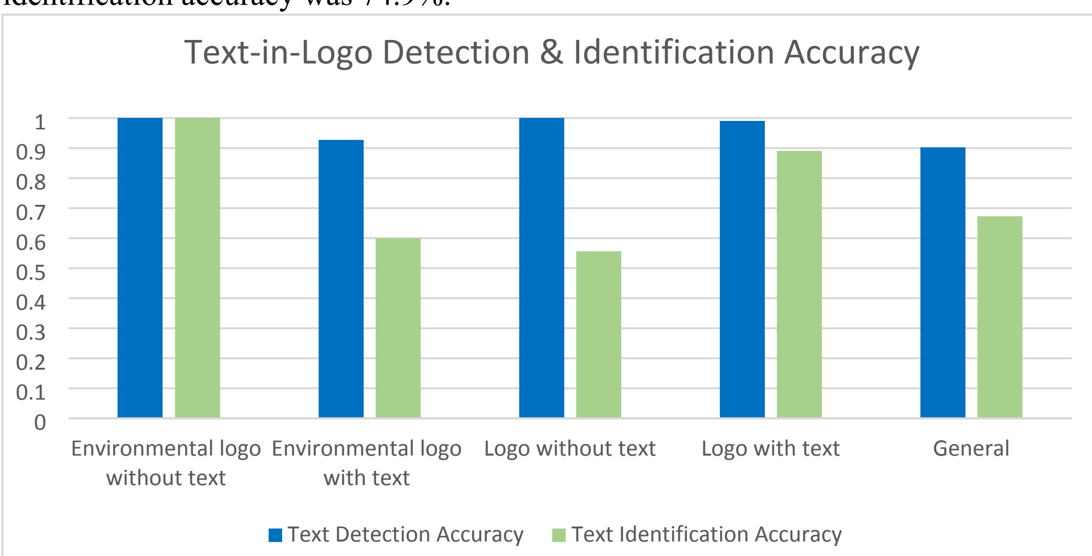


Figure 4 Vision AI Performance on Detection of Text within a Logo

Environmental Text Detection and Identification Performance

The results of Vision AI's performance on environmental text detection and identification are presented in Figure 5. This test had a smaller sample size because we excluded images of isolated logos from the test set. Average environmental text *detection* accuracy among the nuclear-relevant image classes was 97%; average environmental text *identification* accuracy was 77%.

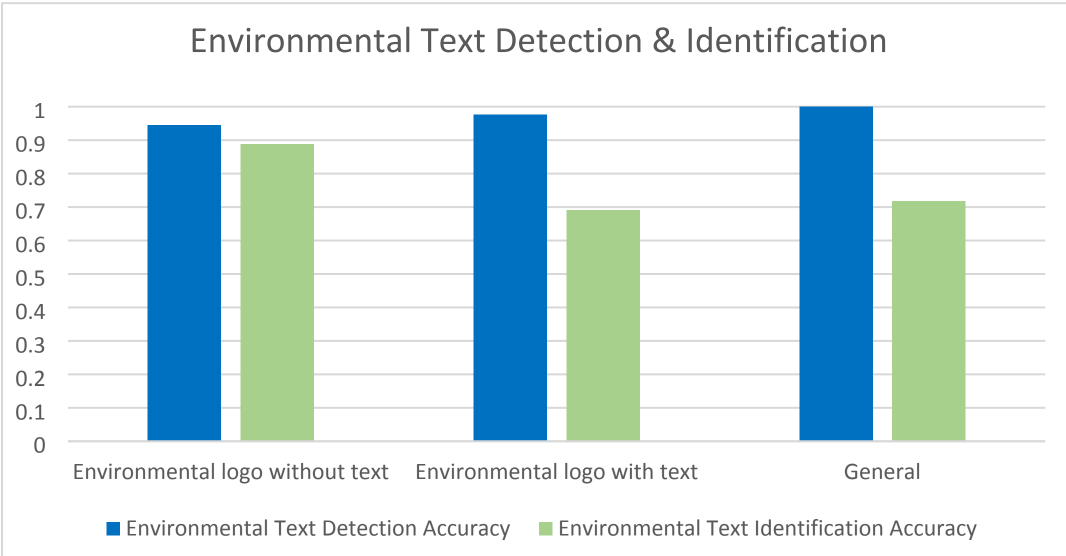


Figure 5 Vision AI Performance on Detection and Identification of Environmental Text

Discussion

Performance Observations

We observed several broad trends in Vision AI's performance throughout testing. Vision AI was better at logo detection when the logos were isolated. When Vision AI misidentified a logo, it often identified the input logo as a more common one with similar features like color, shape, or font. It was rare that Vision AI misidentified a logo as one that looked totally different. One example of a logo misidentification is shown in Figure 6. An unexpected anecdotal result of this testing was that Vision AI did not appear to perform better on general logos than nuclear-specific logos.

Figure 6 Example Logo Misclassifications with the Input Logo on the Left (Altius Materials) and Predicted Logo on the Right (American Eagle Outfitters)

Vision AI was highly successful at detecting text in logos and environmentally in images, but it was less successful at accurately identifying and transcribing the text. Most often, text transcription errors appeared to be due to language, image quality, and font. Figure 7 shows an example of highly stylized font where Vision AI misidentified text.

We observed that Vision AI performed better identifying some alphabets than others; for example, it performed well on Latin alphabets, Russian, Mandarin, and Korean, but it performed worse with languages like Arabic, Hindi, and Bengali.

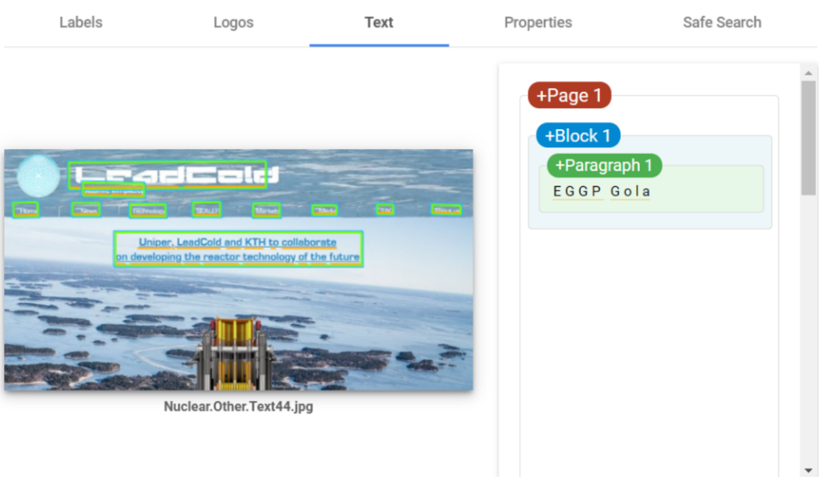


Figure 7 Example Text identification using Vision AI. The stylized text within the LeadCold logo may have contributed to a poor transcription. The text "LeadCold" was identified as "EGGP Gola."

Safeguards Impact

Through our testing, we observed that while Vision AI is generally successful at detecting and transcribing text, the platform is less able to detect or identify logos.

Vision AI correctly detected 67.4% of the logos tested (68.9% of the nuclear-relevant logos, and 62.8% of the general logos). As expected, the detection rate was much higher for isolated logos, with or without text, than for environmental logos (77% versus 54%); however, the detection of logos within a busy environment of other unrelated content is a more realistic safeguards use case. While the logo detection rate was reasonable, the poor identification performance (35.5% correct identifications for nuclear-relevant logos, and 48.5% correct for general logos) indicates that logo detection and identification still need to improve before they can be useful in an operational safeguards analysis setting. As with many safeguards applications for computer vision, the lower performance on nuclear-relevant logos is likely due to their lower relevance to the broader commercial uses for these large, open-source models. Additional fine-tuning of the logo identification model with nuclear-relevant logos would likely improve performance to the level of broader logo identification levels (though that is still fairly low).

Vision AI's text detection and identification capabilities had high levels of performance. For text in logos, Vision AI had an average detection rate of 95.2% (97.0% for text in nuclear-relevant logos, and 90.1% for text in general logos) and identification rate of 74.9% (77.8% for text in nuclear logos, and 67.2% identification for text in general logos). For environmental text, Vision AI had an average text detection rate of 98.0% (96.7% in nuclear-relevant images, and 100% in general images), and average text identification rate of 73.7% (75.0% in nuclear-relevant image text, and 71.8% in general image text). This means that with a few exceptions, if there was text in an image, Vision AI recognized it as text. Transcription accuracy for within-logo text and environmental text dropped, and we hypothesize that the drop in performance was related to font or language. Vision AI's fairly high accuracy in text detection and transcription for nuclear and non-nuclear images suggests that these capabilities could be immediately useful for IAEA safeguards.

Conclusion

In this work, we evaluated Vision AI's performance at detecting and identifying logos and text in different types of images. Although Vision AI's text detection and transcription capabilities could be immediately useful for IAEA safeguards analysis, the logo detection and identification performance would require additional development prior to safeguards use.

Future work in this area could involve testing other platforms' logo and text detection and identification capabilities, for example Amazon Rekognition or Hive AI. However, any future comparison of performance would also require re-testing of Vision AI given the rapid pace of development and continuous training for many of these platforms.

While Vision AI might not have sufficient performance to fully meet an IAEA safeguards analysis needs today, Vision AI or similar platforms could be used as an additional tool to identify or recognize nuclear-relevant images in open sources that might not otherwise be detected or provide an early level of triage support to analysts trying to identify logos in images.