SAND2020-6203C

Industry's Use of Deep Learning, and Implications for International Safeguards

Michael Higgins, Zoe N. Gastelum

Sandia National Laboratories*

Albuquerque, NM, USA

Abstract

Deep Learning (DL) has been a concept since the late 1950s and has advanced significantly in the last twenty years due to enhanced computational capabilities such as graphical processing unites (GPUs) and cloud-based computing. DL is now capability that is within reach across many industries. Following this trend, the International Atomic Energy Agency (IAEA) has started to explore the use of DL to enhance its ability to collect, process, and analyze international safeguard-relevant information. The ways industry has successful utilized DL can give insight into potential applications for international safeguards. In this paper, we will review the use of DL across multiple industries. The paper organized based on the different ways industries are using DL, and the potential transfers of DL capabilities to international safeguards activities. By understanding current applications of DL across multiple industries, the international safeguards research community can start to focus on realistic potential applications in our domain and learn from both successes and failures from other domains.

Introduction

Deep learning (DL) has become a fast-expanding area for research and a significant part of business and industry operations. Over the last twenty years, leveraging increases in computing power, DL has been successfully incorporated into many data processing and analysis situations, succeeding in computer vision, pattern recognition, speech recognition, text processing, and recommendation systems. DL's automation of previously labor-intensive or in some cases infeasible processes has increased the speed at which organizations are able to process and analyze a wide variety of datasets.

While international nuclear safeguards data may not be "big" by data science standards, there are many facets of safeguards verification that currently require significant manual (i.e. human) data processing and analysis effort. For that reason, the International Atomic Energy Agency (IAEA) Department of Safeguards has expressed interest in potential applications of DL as part of a broader interest in data analytics, evidence of which can be seen in recent forums such as the IAEA's 2017 Workshop on Emerging Technologies and the 2018 IAEA Safeguards Symposium.

In the following sections we will describe some of the most common DL classes of algorithms, provide examples of how they are being used in some industries, and suggest potential applications to international nuclear safeguards activities.

Classification

Classification is a technique in which an algorithm determines to which pre-defined category an unclassified datapoint belongs. (Khan and Madden 2010). Classification may be applied to text, images, video, audio, or other sensor data. Examples of industry use of classification include:

- *Medicine.* Dermatology utilizes image classification to automate the identification of skin conditions, with a focus on finding cancer skin lesions (Esteva et al. 2017). Figure 1 show

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.

examples of the various skin conditions are classified in this model, along with their saliency maps (visual indications of the pixels that were most “important” to the model in making the classification).

- *Law*. (Remus and Levy 2015) describe how classification is being used in the legal domain to predict case outcomes, thereby supporting lawyer decisions on whether they want to take on a case.

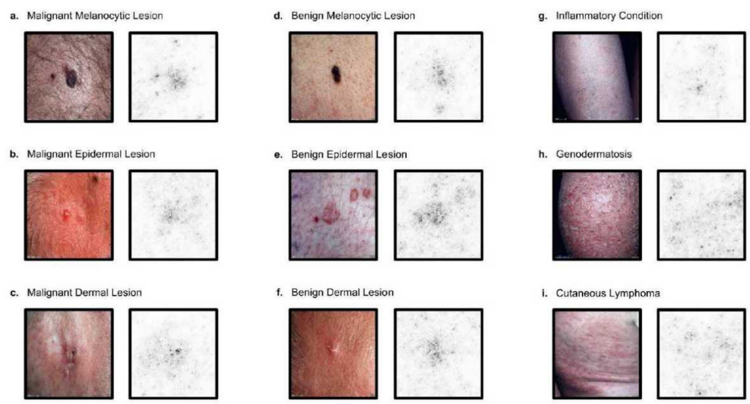

Figure 1 Examples of the various skin conditions classified via a DL model. With each image paired with their respective saliency map. (Esteva et al. 2017)

There are many potential applications of classification for international nuclear safeguards. Image classification could be developed as a preliminary assessment of the integrity of tamper indicating devices, to flag frames of surveillance footage when a specific object appears, or to automate tagging to organize internal image collections by fuel cycle step or equipment. Further, text classification tools could be developed to identify the step of the Physical Model to which technical publications are most relevant. This is the intent behind the IAEA Content Reification Engine (ICORE) tool (Crowley et al., n.d.).

Object Detection and Segmentation

Object detection algorithms identify an object of interest, and place a bounding box, segmentation mask, or other visual indicator around the object, along with the object’s classification (Druzhkov and Kustikova 2016). Object detection goes a step further than image classification by segmenting an image to indicate where in the image the object class sits and can be applied to just a few objects of interest that are being searched for, or to the segmentation and classification of an entire image. Object detection can also include facial recognition models with the ability to differentiate characteristics of a face (Z. Liu et al. 2015) and detect emotion within someone’s face (Bartlett et al. 2004). Examples of industry use of object detection and segmentation include:

- *Agriculture.* Image segmentation has been used for apple sorting and grading (Mizushima and Lu 2013), and to identify potato blight by identifying the health of the plant’s leaves(Islam et al. 2017).
- *Security.* Significant press has recently covered controversial uses of facial recognition software by police forces(Valentino-DeVries 2020), as well as the Chinese government to monitor oppressed minorities(Paul Mozur 2019).

- *Transportation.* The automotive industry has been implementing DL models to support self-driving cars. Tesla's video ROIs detection (shown in figure 3) can acquire the lane it is driving in, surrounding lanes, different types of vehicles, and pedestrians.

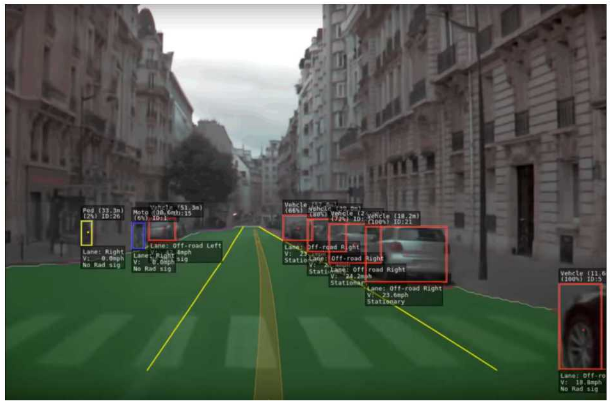

Figure 3 A image from the self-driving model used by Tesla, which identifies the lane it is driving in, pedestrians, other cars and vehicle types, and other lanes on the road, to see the whole video go to https://www.youtube.com/watch?v=_1MHGUC_BzQ("Paris Streets in the Eyes of Tesla Autopilot - YouTube" n.d.).

For safeguards, object detection models could be developed for safeguards surveillance camera data which could be used to identify the presence and movement of key pieces of equipment or containers housing nuclear materials. Research on this use case is currently being pursued by Cui et al(Cui 2018). At IAEA Headquarters, satellite imagery analysts could implement object detection models to locate and segment buildings or externally visible equipment(Warner et al. 2018), and open source analysts could use object detection to locate key indicators in highly complex, cluttered, or obscured visual fields.

Regression

Deep regression models are used to "predict continuous values."(Lathuiliere et al. 2020). While this can be used for applications such as counting crowds, Lathuiliere uses head pose estimation, facial landmark detection, and human-body pose estimation. Various industries have used regression models:

- *Transportation.* Regression can be used to understand road congestion, an imperative step for both self-driving cars, fine-tuning hybrid vehicles and designing smarter traffic systems (Devi and Neetha 2017; Park et al. 2009; Murphey et al. 2012).
- *Cyber security.* Regression has been used to identify spam email based on counting word use in segments of up to three words, that form dictionaries. Expert labeled spam emails define dictionaries, and weights within the DL model are applied based on the number of phrases that show up within an email message (Fette, Sadeh, and Tomasic 2007; Barreno et al. 2010; Lowd and Meek n.d.).

Safeguards satellite imagery analysts could use deep regression models to perform estimate counts of drums in uranium tailings yards. Regression models could also be applied to surveillance footage to track positioning of large equipment such as facility cranes.

Pattern Recognition and Anomaly Detection

Pattern recognition and anomaly detection algorithms learn what is considered normal through provision of many examples and highlight activities that are outside of those patterns. Pattern recognition is similar to classification and clustering algorithms in its assignment of new data to a community (either within the pattern, or abnormal) based on feature space that is has learned. Pattern recognition and anomaly detection algorithms are able to learn patterns over multiple data types in a single pattern ("What Is Pattern Recognition in Machine Learning" n.d.). Examples of industry use include:

- *Medicine.* In the medical field, pattern recognition is used to search for cancer, using inputs from human white blood cells and five types of cancer cells (Ozaki et al. 2019).
- *Finance.* The banking industry uses pattern recognition and anomaly detection models to detect fraudulent activities on bank accounts such as purchases made on stolen credit card information, based on learned patterns of account activity. (Maes et al. 2002).

Safeguards inspectors could use anomaly detection algorithms to support review of surveillance data, learning patterns of activity within a facility and flagging when there is a change such as where or how equipment is being used, or hours of operation. Current research is being conducted in this area (Smartt et al. 2019). Pattern recognition algorithms might also be applied to sensors deployed in a facility that are monitored remotely to detect drift from calibration or inconsistencies between disparate sensors in a facility. Safeguards analysts could potentially use anomaly detection algorithms to monitor news sources or technical publication streams for new topics, changes in word use, or arrivals and departures of new authors(Whitney, Engel, and Cramer 2009).

Image Generation and Enhancement

Deep image enhancement algorithms use deep learning to "understand" what is likely missing or pixelated in an image and improve the quality. Image enhancement can change the style of an image (for example, making an image look like the style of famous painters), add color to an image, or reconstruct an image (Brownlee 2019). In contrast, image generation uses deep generative models to make new images from scratch, such as generating faces(Karras, Laine, and Aila 2019). Examples of how industry has used these techniques include:

- *Medicine*. Medical images have benefited from image enhancement. Various reconstruction techniques are proposed for analysis. The first technique synthesizes a ROI and the second process uses a dictionary to improve image quality (Ravishankar, Ye, and Fessler 2020). Synthesis is used on MRI images of the brain, and the ROI is identified, and surrounding areas are removed to better visualize the ROI (shown in Figure 4).
- *Security.* Security has proposed using enhancement and style transfer techniques to improve the quality of images. Image enhancement of luggage is going through the X-ray machines to aid TA's search has been proposed(Singh and Singh 2005).

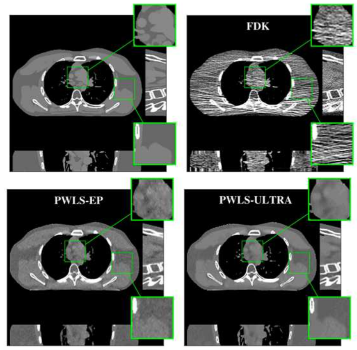

Figure 4 Image enhancement strategies on CT scans of a phantom model. (Ravishankar, Ye, and Fessler 2020).

Safeguards analysts could use image enhancement to improve image quality and resolution of open source images, for example to enhance a facility map or nuclear core diagram in the background of a publicly released photo. Satellite imagery analysts might also be able to use image enhancement to improve the quality of grainy images or support analysis when atmospheric conditions make it difficult to see.

Natural Language Processing and Speech Recognition

Natural Language Processing (NLP) consists of the application of statistical and computer algorithms in order to process human-understandable language in a way that computers can interact with the data and return results in natural language. The aim of NLP is understanding, analyzing, manipulating, and generating natural language (Shetty 2018). Some aspects of NLP fall within DL, and examples from machine translation, captioning, and question and answer capabilities will be provided below.

<u>Machine translation</u>

Machine translation refers to the use of computer algorithms to translate one natural language to another, for example from Spanish to English. Some machine translation tools use DL models to translate through a broader understanding of meaning of full text (e.g., entire book) human translations, rather than the traditional approach of translating individual words or phrases(Lewis-Kraus 2016).

- *Customer Service.* (Abacha and Zweigenbaum n.d.) have proposed DL translation to support multi-lingual medical chatbots that typically rely on SPARQL based queries (which are used to extract information from databases).

Safeguards inspectors working in the field might be able to use real-time and augmented reality translation tools to facilitate better communications with facility operators and awareness during on-site activities. Further, safeguards analysts use many different tools to search for and translate potentially relevant information from open sources. DL-based translation capabilities offer an improvement on translation, and could potentially be trained on specific libraries of technical words and phrases to further improve their capabilities.

<u>Captioning</u>

Captioning algorithms create natural language descriptions (captions) describing the content of data such as an image or video(You et al. 2016). Examples of industry applications of captioning include:

- *Medicine. The* medical community has proposed using captioning for automatic detection and description of tumors. Figure 5 shows an example output of the proposed captioning model. Biomedical engineers have proposed using captioning as a tool to aid physicians with post-appointment notes(Kisilev et al. 2011).
- *Journalism.* Proposals for context-aware captioning capabilities(Tran, Mathews, and Xie n.d.) would incorporate context from written news articles into captions of images, that would provide additional detail not available from the visual image alone.

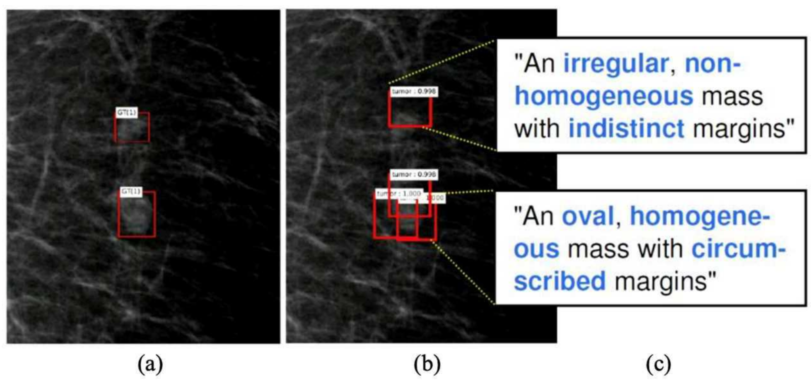

Figure 5 Captioning of a medical image, where masses are first identified (a), and then captions are added (b), and (c) the selected captions(Kisilev et al. 2011).

Captioning could potentially also be applied (not as an exhaustive measure) to images an inspector takes in the field, or to images from safeguards sensors such as Cherenkov viewing devices. Open source safeguards analysts confront the challenge of mislabeled or unlabeled images, both in the wild and within internal collections. Captioning algorithms could support better search and processing of open source images.

<u>Question and Answer</u>

Question and answer algorithms use natural language queries in text (e.g., a search box) or audio (e.g., human vocalization) to search a defined data target such as a database or the Internet, to return a result in natural language – such as a paragraph explanation, not merely a list of search results. Question and answer algorithms support some functions of a “smart” assistant like Siri (Feng et al. 2015), or can provide recommendations for anything from Jeopardy questions to what to cook for dinner (“The Era of Cognitive Systems: An Inside Look at IBM Watson and How It Works | IBM Redbooks” n.d.). Examples of question and answer use in industry include:

- *Law*. Remus and Levy describe the use of IBM Watson for legal applications to answer legal questions, and mention several companies that offer automated legal services such as contracts, wills, and divorce agreements based on similar capabilities (Remus and Levy 2015).
- *Customer Service*. The banking industry has implemented voice- and text-based chatbots that can guide users to the correct application or service for their banking need (Okuda and Shoda 2018).

International safeguards inspectors could potentially benefit from the use of a question and answer algorithm in the form of an automated inspection assistant, who could look up information, record inspector observations, or even help fill out required inspection paperwork.

Safeguards analysts could use an IBM Watson-type question and answer capability to help summarize current information from a state file, query their internal safeguards information database, or identify related information that wasn't part of an explicit search term or query based on the system's broader understanding of natural language meaning. Such a system could be fine-tuned on safeguards-related topics to make it more efficient.

Information Search and Retrieval

Information retrieval is a process of classifying a query, using that query to search a database or other information source, and returning relevant results (X. Liu et al., n.d.). Examples of information retrieval use in industry include:

- *Law.* Markoff describes how the process of finding documents relevant to a lawsuit ("discovery", in legal terms) is being automated using natural language processing to find relevant documents (Markoff 2011).

Conclusion

Enhancements to computing power and architecture and the associated increase in performance of DL models has brought us into the current age of AI popularity. Deep models have been adopted across many industries for various tasks, to the point where we as ordinary citizens may be interacting with DL-based systems multiple times in a given day without even realizing it. The broad capabilities and utility of DL models have been demonstrated in the fields of medicine, law, and security to name a few. In this paper, we explored several common DL capabilities and their uses in various industries. Based on these advances, we believe that international safeguards may be able to adopt some of the capabilities that DL supports to reduce inspector and analyst workloads, increase effectiveness on current tasks, and support new or novel workflows such as voice assisted inspection. We recognize the many nuances required for safeguards practitioners at the IAEA to be able to adopt these emerging capabilities. While many systems would likely require significant customization for application to safeguard, we see this work as the first step to marrying existing and emerging DL capabilities with safeguards workflows for more effective and efficient use of human safeguards expert time and effort.

References

Abacha, Asma Ben, and Pierre Zweigenbaum. n.d. *Medical Question Answering: Translating Medical Questions into SPARQL Queries*. Vol. 12. Accessed June 1, 2020. http://clef.isti.cnr.it/.

Barreno, Marco, Blaine Nelson, Anthony D. Joseph, and J. D. Tygar. 2010. "The Security of Machine Learning." *Machine Learning* 81 (2): 121–48. https://doi.org/10.1007/s10994-010-5188-5.

Bartlett, Marian Stewart, Gwen Littlewort, Claudia Lainscsek, Ian Fasel, and Javier Movellan. 2004. "Machine Learning Methods for Fully Automatic Recognition of Facial Expressions and Facial Actions." In *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 1:592–97. https://doi.org/10.1109/icsmc.2004.1398364.

Brownlee, Jason. 2019. "9 Applications of Deep Learning for Computer Vision." *Machine Learning*

Mastery, 1–23. https://machinelearningmastery.com/applications-of-deep-learning-for-computer-vision/.

Crowley, Jack, Danny Gagne, Dan Calle, John Murray, Remzi Kirkgoeze, and Frank Moser. n.d. “Computational Methods for Physical Opening the Aperture.”

Cui, Y. 2018. “Using Deep Machine Learning to Conduct Object-Based Identification and Motion Detection on Safeguards Video Surveillance.” https://www.osti.gov/servlets/purl/146602.

Devi, Suguna, and T Neetha. 2017. “Machine Learning Based Traffic Congestion Prediction in a IoT Based Smart City.” *International Research Journal of Engineering and Technology*. www.irjet.net.

Druzhkov, P. N., and V. D. Kustikova. 2016. “A Survey of Deep Learning Methods and Software Tools for Image Classification and Object Detection.” *Pattern Recognition and Image Analysis* 26 (1): 9–15. https://doi.org/10.1134/S1054661816010065.

Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. 2017. “Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks.” *Nature* 542 (7639): 115–18. https://doi.org/10.1038/nature21056.

Feng, Minwei, Bing Xiang, Michael R. Glass, Lidan Wang, and Bowen Zhou. 2015. “Applying Deep Learning to Answer Selection: A Study and An Open Task,” August. http://arxiv.org/abs/1508.01585.

Fette, Ian, Norman Sadeh, and Anthony Tomasic. 2007. “Learning to Detect Phishing Emails.” In *16th International World Wide Web Conference, WWW2007*, 649–56. https://doi.org/10.1145/1242572.1242660.

Islam, Monzurul, Anh Dinh, Khan Wahid, and Pankaj Bhowmik. 2017. “Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine.” In *Canadian Conference on Electrical and Computer Engineering*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/CCECE.2017.7946594.

Karras, Tero, Samuli Laine, and Timo Aila. 2019. “A Style-Based Generator Architecture for Generative Adversarial Networks.” In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June:4396–4405. https://doi.org/10.1109/CVPR.2019.00453.

Khan, Shehroz S., and Michael G. Madden. 2010. “A Survey of Recent Trends in One Class Classification.” In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6206 LNAI:188–97. https://doi.org/10.1007/978-3-642-17080-5_21.

Kisilev, Pavel, Eli Sason, Ella Barkan, and Sharbell Hashoul. 2011. “Medical Image Captioning: Learning to Describe Medical Image Findings Using Multi-Task-Loss CNN,” 1. https://dlpm2016.fbk.eu/docs/kisilev_learning.pdf.

Lathuiliere, Stephane, Pablo Mesejo, Xavier Alameda-Pineda, and Radu Horaud. 2020. “A Comprehensive Analysis of Deep Regression.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://doi.org/10.1109/tpami.2019.2910523.

Lewis-Kraus, Gideon. 2016. “The Great A.I. Awakening - The New York Times.” New York Times. 2016. https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

Liu, Xiaodong, Jianfeng Gao, Xiaodong He, Li Deng, Kevin Duh, and Ye-Yi Wang. n.d. "Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval."

Liu, Ziwei, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2015. "Deep Learning Face Attributes in the Wild." In *Proceedings of the IEEE International Conference on Computer Vision*, 2015 Inter:3730–38. https://doi.org/10.1109/ICCV.2015.425.

Lowd, Daniel, and Christopher Meek. n.d. "Good Word Attacks on Statistical Spam Filters." *Ix.Cs.Uoregon.Edu*. Accessed February 24, 2020. https://ix.cs.uoregon.edu/~lowd/ceas05lowd.pdf.

Maes, Sam, Karl Tuyls, Bram Vanschoenwinkel, and Bernard Manderick. 2002. "Credit Card Fraud Detection Using Bayesian and Neural Networks."

Markoff, John. 2011. "Armies of Expensive Lawyers, Replaced by Cheaper Software." *New York Times*, 1–6. https://www.nytimes.com/2011/03/05/science/05legal.html.

Mizushima, Akira, and Renfu Lu. 2013. "An Image Segmentation Method for Apple Sorting and Grading Using Support Vector Machine and Otsu's Method." *Computers and Electronics in Agriculture* 94 (June): 29–37. https://doi.org/10.1016/j.compag.2013.02.009.

Murphey, Yi Lu, Jungme Park, Zhihang Chen, Ming L. Kuang, M. Abul Masrur, and Anthony M. Phillips. 2012. "Intelligent Hybrid Vehicle Power Control Part I: Machine Learning of Optimal Vehicle Power." *IEEE Transactions on Vehicular Technology* 61 (8): 3519–30. https://doi.org/10.1109/TVT.2012.2206064.

Okuda, Takuma, and Sanae Shoda. 2018. "AI-Based Chatbot Service for Financial Industry." *Fujitsu Scientific and Technical Journal* 54 (2): 4–8.

Ozaki, Yusuke, Hidenao Yamada, Hirotoshi Kikuchi, Amane Hirotsu, Tomohiro Murakami, Tomohiro Matsumoto, Toshiki Kawabata, et al. 2019. "Label-Free Classification of Cells Based on Supervised Machine Learning of Subcellular Structures." *PLoS ONE* 14 (1): 1–20. https://doi.org/10.1371/journal.pone.0211347.

"Paris Streets in the Eyes of Tesla Autopilot - YouTube." n.d. Accessed March 2, 2020. https://www.youtube.com/watch?v=_1MHGUC_BzQ.

Park, Jungme, Zhihang Chen, Leonidas Kiliaris, Ming L. Kuang, M. Abul Masrur, Anthony M. Phillips, and Yi Lu Murphey. 2009. "Intelligent Vehicle Power Control Based on Machine Learning of Optimal Control Parameters and Prediction of Road Type and Traffic Congestion." *IEEE Transactions on Vehicular Technology* 58 (9): 4741–56. https://doi.org/10.1109/TVT.2009.2027710.

Paul Mozur. 2019. "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority - The New York Times." 14 April. 2019. https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

Ravishankar, Saiprasad, Jong Chul Ye, and Jeffrey A. Fessler. 2020. "Image Reconstruction: From Sparsity to Data-Adaptive Methods and Machine Learning." *Proceedings of the IEEE* 108 (1): 86–109. https://doi.org/10.1109/JPROC.2019.2936204.

Remus, Dana, and Frank S. Levy. 2015. "Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law." *SSRN Electronic Journal*, December. https://doi.org/10.2139/ssrn.2701092.

Shetty, Badreesh. 2018. “Natural Language Processing(NLP) for Machine Learning.” Towards Data Science. 2018. https://towardsdatascience.com/natural-language-processing-nlp-for-machine-learning-d44498845d5b.

Singh, M., and S. Singh. 2005. “Optimizing Image Enhancement for Screening Luggage at Airports.” *Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, 131–36. https://doi.org/10.1109/CIHSPS.2005.1500627.

Smartt, Heidi A., Dennis Lee, LaCasse Charles, Jessica Bobeck, Heinrich Bornhorst, and Allison Puccioni. 2019. “Detection via Persistence: Leveraging Commercial Imagery from Small Satellites.” *INMM*. https://doi.org/10.1017/CBO9781107415324.004.

“The Era of Cognitive Systems: An Inside Look at IBM Watson and How It Works | IBM Redbooks.” n.d. Accessed March 4, 2020. http://www.redbooks.ibm.com/abstracts/redp4955.html?Open.

Tran, Alasdair, Alexander Mathews, and Lexing Xie. n.d. “Transform and Tell: Entity-Aware News Image Captioning.” Accessed June 5, 2020. https://github.com/alasdairtran/transform-and-tell.

Valentino-DeVries, Jennifer. 2020. “How the Police Use Facial Recognition, and Where It Falls Short.” *The New York Times*, 2020. https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html.

Warner, Timothy, Antero Keskinen, Joshua Rutkowski, and Sam Duckworth. 2018. “Exploitation of High Frequency Acquisition of Imagery from Satellite Constellations within a Semi-Automated Change Detection Framework for IAEA Safeguards Purposes.” In .

“What Is Pattern Recognition in Machine Learning.” n.d. Accessed June 1, 2020. https://huspi.com/blog-open/pattern-recognition-in-machine-learning.

Whitney, Paul, Dave Engel, and Nick Cramer. 2009. “Mining for Surprise Events within Text Streams.” *Society for Industrial and Applied Mathematics - 9th SIAM International Conference on Data Mining 2009, Proceedings in Applied Mathematics* 2 (April): 613–23. https://doi.org/10.1137/1.9781611972795.53.

You, Quanzeng, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. 2016. “Image Captioning with Semantic Attention,” March. http://arxiv.org/abs/1603.03925.