OpenTag: Understanding Human Perceptions
of Image Tagging Algorithms

Kyriakos Kyriakou,[1] **Pınar Barlas,**[1] **Styliani Kleanthous,**[1,2] **& Jahna Otterbacher**[1,2]

[1]Research Centre on Interactive Media, Smart Systems and Emerging Technologies (Nicosia, CYPRUS)
[2]Cyprus Center for Algorithmic Transparency, Open University of Cyprus (Latsia, CYPRUS)
{k.kyriakou, p.barlas}@rise.org.cy, {styliani.kleanthous, jahna.otterbacher}@ouc.ac.cy

Abstract

Image Tagging Algorithms (ITAs) are extensively used in our information ecosystem, from facilitating the retrieval of images in social platforms to learning about users and their preferences. However, audits performed on ITAs have demonstrated that their behaviors often exhibit social biases, especially when analyzing images depicting people. We present OpenTag, a platform that fuses the auditing process with a crowdsourcing approach. Users can upload an image, which is then analyzed by various ITAs, resulting in multiple sets of descriptive tags. With OpenTag, the user can observe and compare the output of multiple ITAs simultaneously, while researchers can study the manner in which users perceive this output. Finally, using the collected data, further audits can be performed on ITAs.

Introduction

Experts have predicted that AI and in particular, machine learning (ML), will drive the next industrial revolution, transforming all sectors of society.[1] Computer vision is clearly a success story for modern ML; along with its technological progress, it has gone beyond its early use in restricted domains (e.g., military or medical imagery) and into user-facing, social applications. In the current work, we focus on proprietary image tagging algorithms (ITAs) offered as cognitive services, including Google Vision,[2] Amazon Rekognition,[3] and Clarifai,[4] given their popularity with developers but also researchers.

Previous work demonstrated that these services often exhibit socially biased behaviors when processing images depicting people. For instance, error rates (e.g., on recognizing the gender of the depicted person) tend to be higher on darker-skinned individuals (Buolamwini and Gebru 2018). Furthermore, the descriptive tags used by ITAs can be subjective (e.g., "beautiful," "fine-looking") (Kyriakou et al.

2019; Barlas et al. 2019). With these technologies now being used extensively in the information ecosystem, it is crucial to hold them accountable for their behaviors. To this end, researchers have proposed the use of human intelligence to interpret opaque machine behavior (Hase et al. 2019) as well as to understand and explain where such systems fail (Nushi, Kamar, and Horvitz 2018).

The purpose of the OpenTag Demo is twofold: i) it serves as a research tool for understanding people's perceptions on ITA outputs and ii) as an awareness tool to help the general public in understanding the risks when using applications that are based on ITAs, by considering the results they get in response to their own images. Figure 1 summarizes the flow of interaction between participating users and researchers with OpenTag respectively. The following section provides a detailed interaction flow between the user participant with the system, and the researcher with the system.
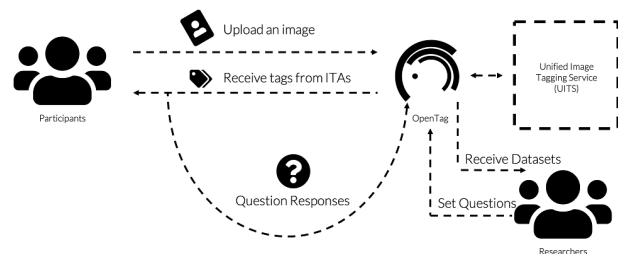

Figure 1: Task Flow.

System Overview

OpenTag is an online platform, developed using Web technologies.[5] The structure and components are shown in Figure 2. For our purposes, we integrated three of the most popular Image Tagging APIs (ITAs) namely, Google Vision, Amazon Rekognition and Clarifai.

Unified Image Tagging Service (UITS). OpenTag provides a unified solution integrated with the selected ITAs using their SDKs. Through UITS, OpenTag communicates with the ITAs, sends the image to the ITAs, receives and

[1]https://www.weforum.org/platforms/shaping-the-future-of-technology-governance-artificial-intelligence-and-machine-learning

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

[3]https://aws.amazon.com/rekognition/

[4]https://www.clarifai.com/

[5]A walk through of the demo can be found at: https://www.youtube.com/watch?v=Yhw1kIfydjo
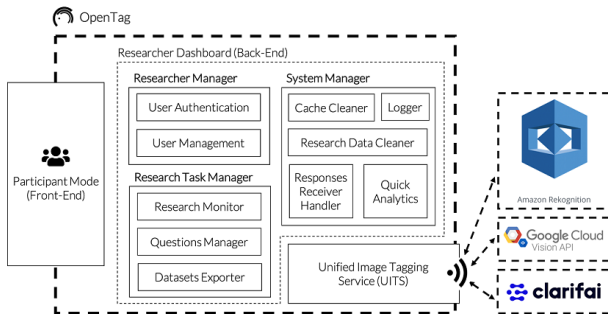
Figure 2: OpenTag Platform Architecture and Components.

pre-processes the corresponding tags, and returns a unified JSON response.

Participant Mode. The user interaction with the system is taking place through this interface. A user can visit the home page of OpenTag, where he or she can *"Upload an image"*. A hint is presented at the bottom of the submission form, stating that *"Images with people get more interesting results"* as a call-to-action element to motivate users to upload images depicting people. When the participant uploads the image, an *Analyze* button is shown to start the execution of the ITAs using the UITS (see Figure 3). All the neces-
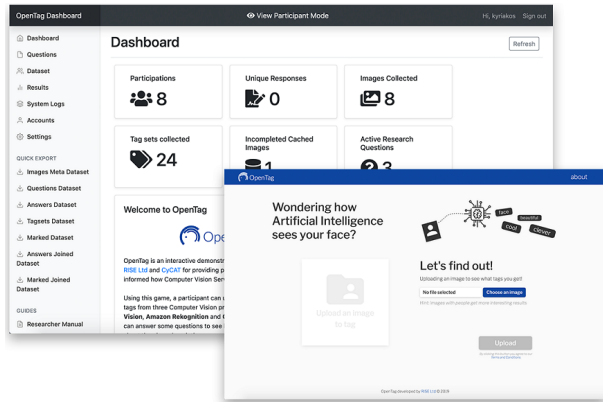

Figure 3: Front- and Back- End of the platform.

sary security measurements were taken into consideration to securely collect, analyze and store people's images. A dedicated statement to that is included in the *About* section, while a short statement is included right under the *Analyze* button, which emphasizes the Terms and Conditions.

After clicking *Analyze*, the participant is redirected to a loading page, in which the photo is shown with a *"We are asking the AIs what they think of this picture. Please wait, it might take a while..."* message. When the UITS returns the unified response, the title of the screen changes to *"This is what AIs discovered about you"*, the tags are presented below the image in three lists (one for each ITA with the ITA commercial name). The user then can observe/compare in three side-by-side columns, how the ITAs perceived the provided photo, through a set of tags. A set of questions (that can be modified by the researcher) follows, for the partici-

pant to answer, focusing mainly of his or her perception of the outputs.

The main research motivation of this tool is to understand how the user perceives the ITA's output (e.g. Do they consider the tags used for describing their photo '‘fair''?) and at the same time intrigue users to question some of the assumptions that ITAs make about them from just one photo. Finally, by clicking *Submit*, the responses to the questions are stored in the system and an option to *"Try another image"* is shown, along with a *"Thank you"* message.

Researcher Dashboard. This mode is available only to the researchers (administrators) of the system. The researcher can log in, monitor and manage the OpenTag platform. From the *Questions* section, he or she can setup the questions that will be shown to the participant during the research task, right after the tags received are presented on screen (see above). There are four types of questions to choose from, e.g., plain *Textbox*, *Single Choice*, *Multiple Choice* or *Tag Selection* questions. Tag Selection is a custom type developed for the current research purposes and consists of a question (e.g. *"Please select the tags which you find unacceptable for the provided image"*) and three multiple-choice checklists of the collected tags as options per each ITA. We made all questions mandatory by design, except the *Tag Selection* types, where a participant can choose zero or more tags from all of the lists. Through managing the questions a researcher can modify the research task based on specific research objectives. At the end of the research study, the image dataset, along with csv files containing the tags and responses collected, can be exported in zip files. The raw Unified JSON files containing the raw tags and confidence scores collected from each ITA per image provided by the users, can also be exported.

Additionally, there are some utility functionalities for maintaining the platform (e.g., cleaning any incomplete cached data, create a new set of questions). Furthermore, through *System Logs* she can monitor the state of the system and the UITS integrated component.

Conclusion & Future Direction

Leveraging OpenTag capabilities, we aim to investigate further people's perception and the effects of the biases and implications of the ITAs by following a crowdsourcing approach. A first pilot study has to be conducted to optimize the system both in terms of functionality and usability. Following an initial usability study, a number of crowdsourcing research tasks will be designed and deployed on the OpenTag platform to study users' perception (e.g., fairness perception, perceived bias) on the set of tags used to describe his or her photo. Most preferably, we will study how different groups of people who share the same characteristics or demographics perceive these judgments.

Acknowledgments

References

[Barlas et al. 2019] Barlas, P.; Kyriakou, K.; Kleanthous, S.; and Otterbacher, J. 2019. Social b (eye) as: Human and machine descriptions of people images. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 583–591.

[Buolamwini and Gebru 2018] Buolamwini, J., and Gebru, T. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, 77–91.

[Hase et al. 2019] Hase, P.; Chen, C.; Li, O.; and Rudin, C. 2019. Interpretable image recognition with hierarchical prototypes. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, 32–40.

[Kyriakou et al. 2019] Kyriakou, K.; Barlas, P.; Kleanthous, S.; and Otterbacher, J. 2019. Fairness in proprietary image tagging algorithms: A cross-platform audit on people images. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 313–322.

[Nushi, Kamar, and Horvitz 2018] Nushi, B.; Kamar, E.; and Horvitz, E. 2018. Towards accountable ai: Hybrid human-machine analyses for characterizing system failure. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.