Digital Investigation 29 (2019) S124–S142

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

DFRWS 2019 USA — Proceedings of the Nineteenth Annual DFRWS USA

A Practitioner Survey Exploring the Value of Forensic Tools, AI, Filtering, & Safer Presentation for Investigating Child Sexual Abuse Material (CSAM)

Laura Sanchez [a], Cinthya Grajeda [a], Ibrahim Baggili [a, *], Cory Hall [b]

[a] Cyber Forensics Research and Education Group (UNHcFREG Tagliatela College of Engineering, ECECS, University of New Haven, 300 Boston Post Rd., West Haven, CT, 06516, USA
[b] The MITRE Corporation, 7515 Colshire Drive McLean, VA, 22102-7539, USA

ARTICLE INFO

Article history:

Keywords:
Digital forensics
Data science
Artificial intelligence
Digital forensic tools
Investigations
Child sexual assault
Child sexual abuse material
Law enforcement
Safer presentation

ABSTRACT

For those investigating cases of Child Sexual Abuse Material (CSAM), there is the potential harm of experiencing trauma after illicit content exposure over a period of time. Research has shown that those working on such cases can experience psychological distress. As a result, there has been a greater effort to create and implement technologies that reduce exposure to CSAM. However, not much work has explored gathering insight regarding the functionality, effectiveness, accuracy, and importance of digital forensic tools and data science technologies from practitioners who use them. This study focused specifically on examining the value practitioners give to the tools and technologies they utilize to investigate CSAM cases. General findings indicated that implementing filtering technologies is more important than safe-viewing technologies; false positives are a greater concern than false negatives; resources such as time, personnel, and money continue to be a concern; and an improved workflow is highly desirable. Results also showed that practitioners are not well-versed in data science and Artificial Intelligence (AI), which is alarming given that tools already implement these techniques and that practitioners face large amounts of data during investigations. Finally, the data exemplified that practitioners are generally not taking advantage of tools that implement data science techniques, and that the biggest need for them is in automated child nudity detection, age estimation and skin tone detection.
© 2019 The Author(s). Published by Elsevier Ltd on behalf of DFRWS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The increased use of mobile technologies and the willingness of general society to capture and share even the most intimate details of their lives has created an abundant source of evidence for investigators. In particular, digital images and videos have become a substantial component in the investigation of child sexual abuse material (CSAM). Whether serving as direct or circumstantial evidence, images and videos can provide significant details, such as the identity of victims and suspects.

While beneficial to an investigation, the sheer amount of digital content encountered and handled by investigators is problematic. Individual investigations may deal with terabytes of data, containing millions of still images and hundreds of hours of video, across multiple devices (Quick and Choo, 2014). Examining such large volumes for relevant evidence may be a long, unmanageable, and unproductive process. This is true even with the use of automated digital forensic tools, which struggle to keep up with the constant evolution of devices and to rapidly process large datasets. Consequently, this can lead to lost time during urgent cases and an increase in backlogs (Vidas et al., 2014).

In addition to putting a strain on organizational resources, the processing and examination of CSAM can be traumatic for investigators. Exposure to such imagery may elicit symptoms similar to those of post-traumatic stress disorder. Investigators may experience physical and emotional impact on self and home, intrusive images and thoughts about the viewed content, and a heightened protectiveness or paranoia regarding the safety of children (Burns et al., 2008).

* Corresponding author.
E-mail addresses: lsanc3@unh.newhaven.edu (L. Sanchez), Cgraj1@unh.newhaven.edu (C. Grajeda), IBaggili@newhaven.edu (I. Baggili), clhall@mitre.org (C. Hall).
URL: http://www.unhcfreg.com, http://www.unhcfreg.com, http://www.Baggili.com, http://www.mitre.org

https://doi.org/10.1016/j.diin.2019.04.005
1742-2876/© 2019 The Author(s). Published by Elsevier Ltd on behalf of DFRWS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

A growing amount of research has been conducted on the impact of working on CSAM cases (Burns et al., 2008; Krause, 2009; Perez et al., 2010; Bourke and Craun, 2013; Seigfried-Spellar, 2017) and specific work-related challenges (Franqueira et al., 2018). However, research has not concentrated on the specific role of the technologies and tools utilized by investigators in these cases. In particular, there has not been a comprehensive study examining the value investigators give to these tools and technologies. While usability studies focusing on the appearance, ease of use, and data presentation of forensic tools have been conducted (Hibshi et al., 2011; Bennett and Stephens, 2008), there are a lack of studies focusing specifically on the role and value of tools and technologies currently available to those investigating CSAM cases.

Our study presents survey findings aimed at gathering the insight of child exploitation investigators regarding the functionality, effectiveness, accuracy, and importance of the tools and technologies they use. The motivation is to gain an understanding of how investigators utilize and view tools and technologies in order to improve workflow, shorten the amount of time of an investigation, and limit investigative exposure to CSAM content. Results will inform current research and development projects to improve or establish new products or processes for investigating CSAM. Our contributions are as follows:

- This is the first comprehensive survey study to explore the value assigned by practitioners to current tools utilized in the investigation of CSAM.
- This is the first comprehensive survey to study the current use of Data Science techniques and technologies in the investigation of CSAM.
- This study explores and identifies what investigators deem as acceptable false positive and false negative rates of CSAM investigative technologies.

The paper is organized as follows. Section 2 presents background information and related work. Next, Section 3 details the methodology employed in the study. The survey design is discussed in Section 4, followed by a presentation of the results in Section 5. Limitations of the study are discussed in Section 6, and Section 7 recaps the main findings and areas of future work. Finally, recommendations are made in Section 8 while Section 9 concludes the paper by acknowledging those that have contributed to the study.

2. Related work

2.1. Psychological aspect/trauma

Investigating the potential emotional impact of exposure to CSAM, Burns et al. (2008) points out that those working on Internet Child Exploitation (ICE) teams are at risk for developing secondary traumatic stress, caused by having knowledge of or helping those in traumatic situations. Powell et al. (2015) further investigated the impact of CSAM on ICE investigators via in-depth interviews asking participants to discuss their subjective experience of viewing material that evoked a negative reaction. Their results indicated that perceptions vary, with some participants considering themselves "secondary victims" while most feel that their work was not any riskier than other types of policing. Overall, participants indicated having short-term emotional and physical reactions.

Secondary traumatic stress and coping mechanisms are also explored by Bourke and Craun (2013) in their large-scale study of Internet Crimes Against Children (ICAC) personnel. However, in their study, for the ($n = 600$) participants who completed survey responses, the severity of secondary traumatic stress was calculated. The results indicated that about one quarter of the respondents faced significant secondary traumatic stress, while more than one half scored in the lower to mild range (Bourke and Craun, 2013). Those with higher levels of secondary traumatic stress reported more difficulty and engagement with explicit content, indicating a correlation between difficulty and stress.

Subsequently, Seigfried-Spellar (2017) surveyed ($n = 129$) law enforcement officers from ICAC's listserv. Results showed that personnel who had to perform both the duties of an investigator and digital forensic examiner in CSAM cases were the most affected in experiencing secondary traumatic stress, low self-worth, and lack of concentration when compared to personnel who only worked as a digital forensic examiner.

2.2. Tools, techniques, and automation

Recent studies indicate that supportive work environments play an important role in coping strategies for CSAM investigators (Burns et al., 2008; Bourke and Craun, 2013). Further analysis concentrating on investigators' perceived challenges in the field demonstrate that technology is an essential component in developing and maintaining a supportive environment, particularly in the form of increasing efficiency and reducing workloads. Participants in a study by Powell et al. (2014), for example, discussed the strain of large workloads and long hours and indicated that computer technology is an asset in this regard.

Recognizing the importance of technology in such investigations, a few researchers have proposed or introduced new technologies and tools for detecting and processing CSAM. Among them is Sae-Bae et al. (2014), who proposed a method to automatically detect CSAM. Improving on skin tone and face detection, the proposed system intended to identify explicit images and detect children's faces in such images.

Ulges and Stahl (2011) introduced a system of CSAM detection based on color visual words. The system analyzes small color patches of an image and provides classifications (CSAM v. non-CSAM) with the assistance of a Support Vector Machine (SVM).

To assist investigators at a crime scene, de Castro Polastro and da Silva Eleuterio (2010) presented the NuDetective Forensic Tool. The tool was intended to automatically filter files and quickly detect CSAM so investigators may analyze files on the spot. To accomplish this, the tool detects nudity via image pixel analysis.

2.3. Triaging

Regardless of the lack of official definition of triaging, much work has been done in an effort to improve the process. For instance, Rogers et al. (2006) introduced The Cyber Forensic Field Triage Process Model (CFFTPM), where the goal is to rapidly process forensic evidence on the field using a six-phase forensic processing model. The phases include "planning, triage, usage/user profiles, chronology/timeline, Internet activity, and case specific evidence."

Another approach was developed by Shaw and Browne (2013), called "enhanced previewing." As an automated process, all located devices are processed, but only those with evidence are forensically examined (Shaw and Browne, 2013). Although the approach enables practitioners to process more evidence, it has its limitations, including not being suitable for field work and complicating case management.

Utilizing Machine Learning (ML), Marturana and Tacconi (2013) proposed a triaging methodology automating the categorization of digital media. To be used in both "live" and "dead" forensic investigations, the process analyzes devices and, based on a series of quantitative measures, or evidence, assigns a class and associates the device with a type of crime.

Finally, Baggili et al. (2014) developed a triage tool called

Forensics2020. This tool implemented a five phase, multi-threaded bootable technique. Investigators do not have to wait until all phases are finalized; rather, they are able to examine any evidence that has been processed in a completed phase.

2.4. Artificial Intelligence (AI)

As the capabilities of AI continue to expand, researchers have begun utilizing ML to develop tools for quickly detecting and identifying CSAM. Among these tools is the iCOP toolkit, introduced by Peersman et al. (2016). The toolkit, intended for peer-to-peer networks (P2P), utilizes two modules, filename categorization and media classification, to identify and flag new CSAM (Peersman et al., 2016). Furthermore, the tool also allows investigators to see who has shared known CSAM (based on generated hash values) and what other content has been shared by the client (Peersman et al., 2016).

Utilizing Caffe,[1] Yahoo has also developed a model called Not Safe for Work (NSFW) to detect pornographic images. The tool works by taking an image as input and providing a probability score, "which can be used to detect and filter NSFW images" (Mahadeokar and Pesavento, 2016). Similarly, Microsoft offers a technology called PhotoDNA[2] to detect and report CSAM. This tool, however, is typical in that it detects CSAM based on hash values, comparing the value to that of known child explicit content (PhotoDNA Cloud Service, n.d.).

Focusing on shaping Deep Convolutional Neural Networks (CNNs), Vitorino et al. (2018) used data-driven concepts to effectively automate detection of CSAM and other related content using a training dataset composed of thousands of images gathered from the Brazilian Federal Police. Anda et al. (2018), on the other hand, concentrated their efforts on evaluating established age prediction services, such as Amazon Rekognition,[3] Deep Expectation (DEX),[4] Kairos,[5] and Microsoft Azure Cognitive.[6] The goal was to identify any performance related issues and trends among different services to improve overall results. One major issue hindering the advancement of these types of technologies is the lack of properly labeled age range databases. Thus, the researchers created a dataset generator in order to efficiently evaluate these services.

3. Methodology

The following methodology was employed to conduct this research:

1. Performed a literature review, obtaining and examining previous, related work. The findings are discussed in Section 2.
2. Designed and piloted the survey utilizing Qualtrics[7] survey software.
3. Obtained approval from the Institutional Review Board (IRB) at the University of New Haven (UNH)[8] to collect data.
4. Distributed the survey to several agencies and organizations via the MITRE Corporation.[9]
5. Obtained data by exporting recorded responses as PDF and CSV files and raw data from Qualtrics.
6. Data was analyzed by the researchers, each tackling a different focal point or section of the survey.

4. Survey design

The questions were designed to address a particular need in the field and in research: understanding what tools and technologies are utilized by CSAM investigators and how they feel about these tools and technologies. Design of the survey commenced January 2018 and, following three drafts and two brief testing phases, the survey was distributed April 2018. Any revisions made to the survey were to address issues of wording and formatting.

The survey consisted of 49 questions of the following question types: 7 Likert Scale, 11 Multiple Choice, 7 Multiple Selection, 6 Free Response, 2 Rank, 1 Drop Down List, and 15 Numerical Slider. Furthermore, the survey contained a page of definitions explaining the technologies that would be explored in many of the questions.

As the survey was intended to gauge an understanding of the tool and technology usage experiences of CSAM investigators, it was distributed to various types of agencies that have concentrated their efforts on this type of work. Agencies ranged from state to local, public to private, and national to international to provide a diverse and global perspective.

5. Results

The survey was distributed through Qualtrics and remained open for about two months. The ideal population sample size was calculated to be $(n = 97)$ participants[10]. One hundred and nineteen participants were recorded to have accessed the survey, and out of those, 118 consented to take the survey, one did not consent, and 12 did not submit any responses at all. Not all $(n = 106)$ participants answered every question in the survey. This discrepancy is addressed in every section and discussed in the Limitations section (Section 6) of the paper.

This section presents the following results: Section 5.1 Demographics, 5.2 Tools, 5.3 Technology, 5.4 Workflow, and 5.5 Acquisition, Processing, Analysis, & Reporting Time.

5.1. Demographics

Results relating to demographics (Appendix A; Table A.1), show that the majority of the sample population were white (93.40%) males ranging from ages 35–54 (65.10%) with at least a high school diploma. The majority (41.51%) of respondents indicated a Bachelor's degree as their highest level of education and most degrees were related to the fields of technology and law (Appendix A; Table A.2). In fact, more respondents have a degree in Crime, Law, and Justice (28.41%) than in Digital Forensics (17.05%). It should be kept in mind however, that only 83.02% (88 of 106) of the total number of participants responded to this question.

When asked about competency (Appendix A; Fig. 4) in areas such as Computer Science and Data Science, for example, 94.34% of respondents (100 of 106) provided an answer. Results indicate that respondents *agree* and *strongly agree* in being most competent in Digital Forensics (99%), Internet and Information technology (79%) and Computer Science (75%), while the majority (41%) *neither agrees nor disagrees* on being competent in Data Science. Furthermore, participants were the least competent in Software Engineering and Software Design.

[1] https://github.com/yahoo/open/nsfw.
[2] https://www.microsoft.com/en-us/photodna.
[3] https://aws.amazon.com/rekognition/.
[4] https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/.
[5] http://kairos.com/.
[6] https://azure.microsoft.com/en-us/services/cognitive-services.
[7] https://www.qualtrics.com.
[8] https://www.newhaven.edu.
[9] https://www.mitre.org.

[10] Calculations were made using a 95% confidence level, 0.50 standard deviation and a margin of error (confidence interval) of ± 10%

When asked about their employment category (Appendix A; Table A.1), 94.34% (100 of 106) of participants answered the question. A majority currently work at a local (46%) and state government level (32%), with their major occupation being that of a digital forensic examiner (69.92%) and Investigator (26.32%). It is important to note, that based on the high number of responses to this question (133), some respondents can be identified as having both occupations or more. Additionally, when asked about years of experience (Appendix A; Table A.1), 94.34% (100 of 106) of participants responded. Results show that 5% have less than one year experience working in CSAM investigations, while the majority have more than six years of experience (63%) in the field.

Lastly, 94.34% (100 of 106) of participants answered the question whether they received formal training to investigate CSAM cases. Most respondents (69%) indicated having received formal training while the remaining respondents indicated not having received any training at all. Appendix A, Tables A.3 and A.4 present the training locations provided by respondents. The results are widely spread out, totaling 176 responses, with respondents indicating receiving training from multiple sources. Most of the respondents received training through government funded programs (28.41%), such as Internet Crimes Against Children (ICAC) and National White Collar Crime Center (NW3C), and commercial forensic tool companies (14.20%), such as Guidance Software/EnCase and AccessData.

5.2. Tools

5.2.1. Processing

Participants were asked a series of questions regarding the tools they currently utilize when investigating CSAM cases. Two questions, presented in a multiple-answer format, asked participants to identify all the tools they use to process CSAM images and videos. The questions were posed separately to garner an understanding of whether practitioners are utilizing similar or distinct tools to process both images and videos. Both questions were answered by ($n = 99$) of the 106 survey participants (93%), however, when asked about image processing tools, a total of 536 responses were provided, whereas 441 were provided for video imaging tools.

Only one respondent indicated not using tools to process CSAM images, while the rest indicated using at least one or more tools. As seen in Appendix B, Table B.5, the top three tools selected were Cellebrite UFED/PA (17%), Magnet Forensics IEF/Axiom (16%), and Forensic Toolkit (12%). In regards to video processing, two respondents indicated not using tools, while the rest indicated using at least one tool or more. As shown in Appendix B, Table B.6, the top three tools selected were Magnet Forensics/Axiom (18%), Cellebrite UFED/PA (17%), and Forensic Toolkit (13%). Unsurprisingly, for both questions the results indicated that commercial tools appear to be utilized more than free or open-source tools.

Additionally, participants were asked to provide feedback regarding limitations of the tools they currently use to investigate CSAM cases. The results from this free response question are displayed in Table B.7 of Appendix B. Of the ($n = 106$) participants, 41 (39%) responded to the question. Of those that responded, three (7%) indicated that there were no limitations, while six (15%) did not provide details of the limitations. As Table B.7 illustrates, most respondents indicated having encountered feature/capability related limitations (62%), such as their tool lacking certain filtering or safe-viewing technologies and the ability to carve images, enhance poor quality photos, or automatically group together several images of the same victim, among other things. Time and speed were identified as the top limitation for the features/capabilities category, as well as overall. The provided feedback indicates that respondents are also encountering limitations related to accuracy and user-friendliness.

5.2.2. Detection

To understand if practitioners are taking advantage of currently available tools implementing AI and ML to automatically detect pornographic content, specifically Yahoo NSFW and iCOP/ICAC COP, participants were asked about their usage of such tools. Of survey participants, 73.58% responded to the question. As seen in Appendix B, Table B.8, 50% indicated that they have used or currently use iCOP/iCAC COP, 2.56% indicated they have used or currently use Yahoo NSFW and 1.28% indicated they have used or currently use both. The remaining respondents (46.15%) indicated that they have not used either of the tools.

Participants were also asked about the benefits and limitations of such tools, the results of which are displayed in Appendix B, Table B.8. Regarding the benefits of using such tools, 39.62% of the participants provided feedback. Quickness was identified as the top benefit (22.92%), with many explaining that these tools cut down on processing time, help prioritize sooner, and encourage faster resolution of investigations. This contrasts the feedback provided about the image and video processing tools, where speed was seen as a limitation (e.g. the tools process too slowly). As for limitations, 36.79% of participants provided feedback. The ability of the tool to identify only known or hashed content was listed as the top limitation (25%). Interestingly, this ability was also recognized by several as being the reason why the tools work quickly. This means that investigators need technologies that are capable of filtering known media without relying on their hash values.

5.3. Technology

5.3.1. Implementation and usage

Participants were provided multiple-answer questions to identify filtering and safe-viewing technologies employed by the tools they use to process CSAM images and videos. The questions were asked separately to determine if image and video processing tools are implementing the same technologies. There was not much variance among the number of respondents or the responses themselves. While 94 (87%) participants provided a response regarding the technologies found in their image processing tools, 97 (92%) provided a response regarding the technologies found in their video processing tools. For both tool types, 13 respondents indicated that none of the listed technologies were implemented by their tool. Results are displayed in Table C.9, Appendix C.

Skin tone detection, a filtering technology, was found to comprise 56% of the responses for both tool types. Both tool types were indicated to be lacking pose estimation, least explicit frame, selective body part viewing, and neural net detection tag presentation technologies. Among the other types of technologies, many had identical results while a small difference existed between a few. These distinct similarities and small differences indicate that, among respondents, image and video processing tools are implementing the same technologies.

Table C.9 also shows the results of two multiple-answer questions asking participants to specify which of the available filtering and safe-viewing technologies they actually use to process CSAM. Ninety-four (90%) participants provided a response indicating which technologies they use to process images and 93 (89%) indicated which technologies they use to process videos. Unsurprisingly, skin tone detection was found to be the most utilized technology for processing both images (52%) and videos (46%). Age and gender estimation, both filtering technologies, were found to be the least utilized, while pose estimation, least explicit frame, selective body part viewing, and neural net detection tag presentation were not utilized at all. As anticipated, safe-viewing technologies appeared to be the least utilized, particularly as they also

appeared to be implemented by tools much less than filtering technologies.

5.3.2. Value, ranking, and preference

A series of value questions regarding filtering and safe-viewing technologies was also given to participants. Two questions specifically asked how valuable the implementation of a particular type of technology would be in an image or video processing tool. The results are shown in Figs. 1 and 2.

Close to 70% of participants (68.87%) provided feedback regarding the value of certain technologies for image processing tools. Of those that responded, over half (54.79%) indicated that implementation of child nudity detection would be *very valuable*, followed by age estimation (46.58%) and skin tone detection (36.99%). Overall, for most of the filtering technologies, such as child nudity detection, implementation was seen as being *valuable* or *very valuable*, with the exception of three (gender estimation, pose estimation, and object detection), which were deemed *slightly* or *moderately valuable*.

The implementation of safe-viewing technologies, such as selective body part viewing, on the other hand, was viewed as predominately *slightly* to *moderately valuable*. Of the safe-viewing technologies, implementation of selective body part viewing was chosen by the highest percentage of respondents (17.81%) as being *very valuable*, followed by neural net detection tag presentation (16.44%) and nudity blocking (13.70%).

When asked about the value of implementing certain technologies in video processing tools, 67.92% of participants responded. The implementation of both child nudity detection and age estimation were rated as *very valuable* by over half of the respondents (51.39% each), followed by skin tone detection (43.06%). Similar to filtering technologies for image processing tools, the implementation of filtering technologies for video processing tools was generally viewed as being *valuable* or *very valuable*. As for the safe-viewing technologies, implementation of each was chiefly viewed as being *moderately valuable*, with implementation of neural net detection tag presentation being viewed as *very valuable* by the highest percentage of respondents (20.83%).

In general, the results indicate that respondents assigned a higher value to the implementation of filtering technologies than safe-viewing technologies for both image and video processing tools. These results may be due in part to participants not being aware of these technologies as their tools do not implement them.

While the value of implementing safe-viewing technologies was generally viewed as *slightly* to *moderately valuable*, when posed with a value question focusing on the usage of these technologies, respondents indicated that usage was, overall, *moderately valuable* to *valuable*. This may signify that although the implementation of these technologies by processing tools may not be as important, if available, their usage can be beneficial in an investigation. Receiving the highest percentage of *very valuable* scores were the face presentation and nudity blocker technologies (18.31% each). Incidentally, the nudity blocker technology also received among the highest *not valuable* scores (15.49%), after selective body part viewing (16.90%). Results are displayed below in Fig. 3.

In addition to determining the value of filtering technologies, participants were asked to rank each technology in order of importance, with one being the most important and eight being the least. The previous value questions turned out to be indicative of how each filtering technology would be ranked. Overall, each of the filtering technologies was ranked according to the *very valuable* score it received, with those focusing on video processing tools being the most accurate. As expected, child nudity detection was ranked as most important by respondents (47.89%) and object detection as least important (64.79%). Figure C.5 in Appendix C illustrates the results.

Surprisingly, the opposite was found to be true when ranking safe-viewing technologies. In the image and video processing tools value questions, selective body part viewing and neural net detection tag presentation received the highest *very valuable* scores, but ended up being ranked among the least important, whereas least explicit frame and face presentation received the lowest *very valuable* scores and ranked among the most important. In comparison, however, the ranking results were closer to the results of the value question pertaining to the usage of safe-viewing technology. The results can be seen in Figure C.6 in Appendix C.

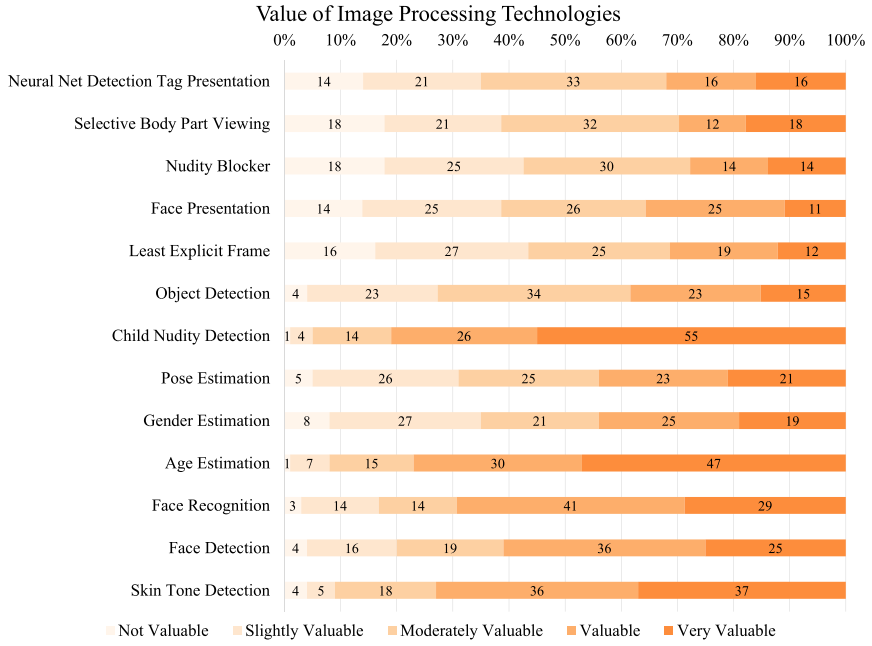


Fig. 1. Each bar represents one Likert scale question. Approximate percentages are displayed for each answer selection.

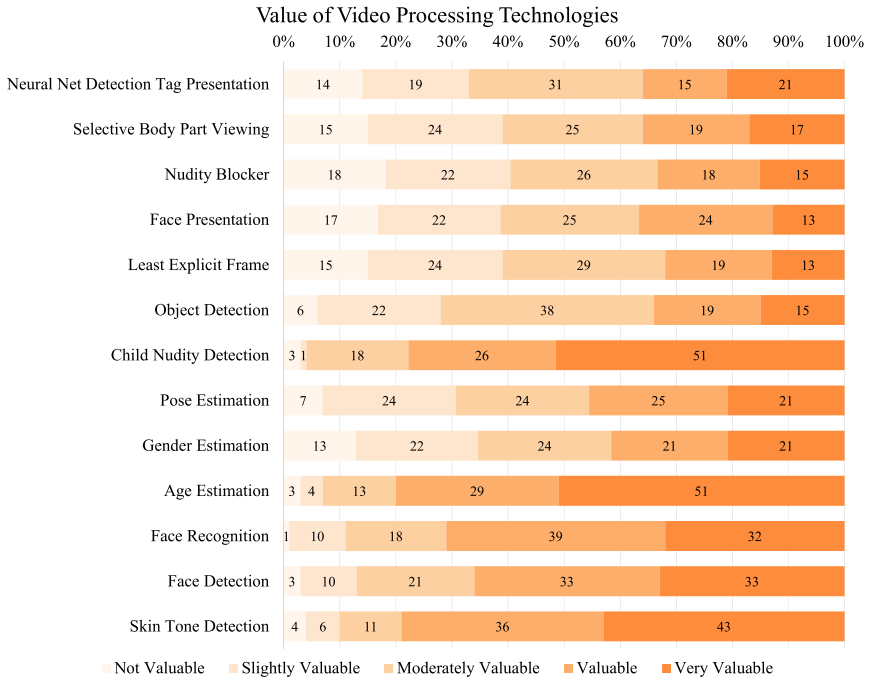


Fig. 2. Each bar represents one Likert scale question. Approximate percentages are displayed for each answer section.

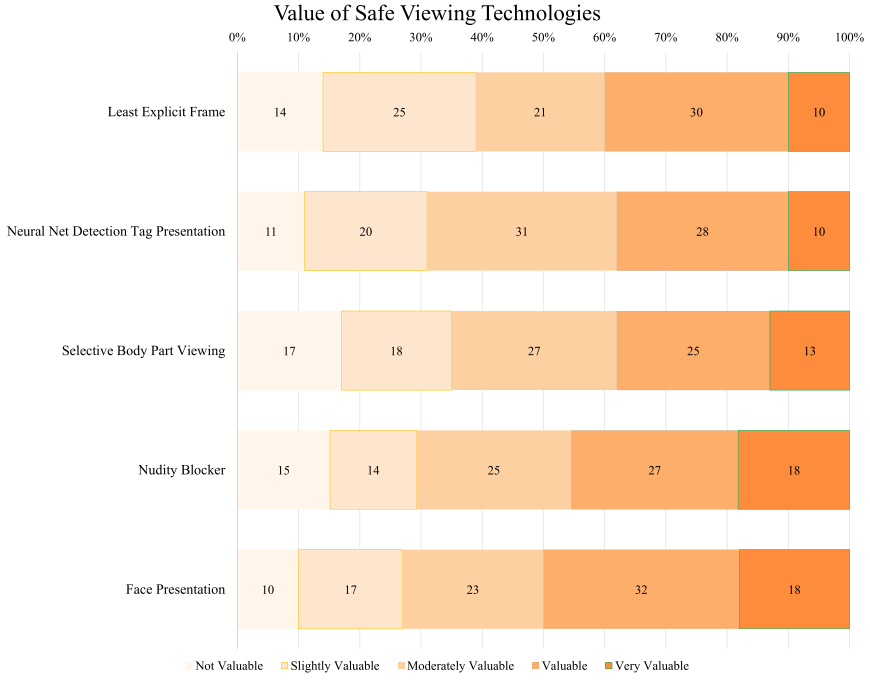


Fig. 3. Each bar represents one Likert scale question. Approximate percentages are displayed for each answer selection.

Participants were also asked a value question regarding technology that could identify certain age groups. The results are shown in C.7 in Appendix C. The age group 0 months to 12 years received the highest percentage of *very valuable* ratings (71.62%), followed by three to five years and five to twelve years (68.92% each).

As it had been anticipated that the age range of 0 months–12 years would likely receive the highest percentage of *very valuable* ratings, a follow-up value question for this age range was asked. The question asked how valuable it would be to have a technology that could identify children between the ages of 0 months and 12 years. As seen in C.8, Appendix C, the majority of respondents (67.57%) selected the rating *very valuable*. As expected, this percentage is very close to that of the previous question, confirming that a technology that can identify a broader range of ages is more desirable.

One-way ANOVA tests were conducted to determine if mean

differences existed between certain groups when assigning value to video and image processing technology, safe-viewing filtering technology, and technology for identifying specific age groups. Here, groups were categorized according to employment type and time on job. Results indicated that a statistically significant difference exists among group means when analyzing time on job against value of video and image processing technology ($F = 4.349$, $p = .003$) and safe-viewing filtering technology ($F = 4.510$, $p = .003$). When measured against technology for identifying specific age groups, only employment type showed a mean difference between groups ($F = 2.527$, $p = .037$). This does not indicate a relationship or correlation between particular value questions and employment type or time on job but, simply, that a mean difference exists and it may be interesting to explore in future work.

5.3.3. False positives/negatives

Due in part to the concern practitioners have expressed regarding false positives and negatives, participants were asked a series of questions regarding this topic. The questions focused on the expected rate and optimal rate of false positives and negatives. This was conducted to determine what rate of false results participants expect to get (based on their previous experience) when utilizing a filtering or safe-viewing technology and what rate of false results they would realistically like to see.

Four questions were posed, with each asking participants to provide a percentage rating to 13 filtering and safe-viewing technologies, indicating what rate of false positives and negatives they would expect and desire for each. The questions were answered by 50%–65% of the participants, each question being answered by a different number of people. It should be noted that in order to obtain a more accurate picture of the results, calculations have been achieved with the exclusion of outliers, given that some participants simply chose zero as a number for answers when using a slider.

Table C.10 in Appendix C shows the results of the expected and optimal false positive rates, respectively. Of the number of participants, 65.10% provided a response regarding the expected rate of false positives and 53.77% provided a response regarding the optimal rate of false positives. As anticipated, the optimal rates were lower than those of the expected rates, indicating that, ideally, respondents would like to see a lower number of false positives. For example, age estimation, which saw the biggest change, had a mean of 27.84 and a median of 25.75 for expected false positives *(SD = 14.74)*. When the optimal time for false positives of age estimation was considered, the mean dropped to 14.47, a 13.37 difference, and the median dropped to 10.45, a 15.3 difference. Additionally, the standard deviation dropped to 13.20, 1.27 points from the mean.

Overall, this type of change was across the board, with the exception of face detection, face recognition, and face presentation, all of which experienced an increase in standard deviation. However, although the standard deviation increased among these technologies, the number of points between the mean and standard deviation dropped. When grouping together the technologies, the mean of expected false positives fell from 268.80 to 182.60 for optimal false positives.

The results for the expected and optimal rate of false negatives are shown in Table C.11 in Appendix C. Similar to the questions about false positives, a larger amount of participants answered the question regarding the expected rate of false negatives (61.32%) than the question regarding the optimal rate of false negatives (50%). As expected, the optimal rates were also lower than the expected rates. For example, age estimation, which, again, experienced the greatest amount of change, had a mean of 23.00 and a median of 20.5 for expected false negatives *SD = 14.86*. For optimal rate of false negatives these numbers dropped. The optimal rate saw a mean of 11.17, an 11.83 difference, and a median of 6.7, a 13.8 difference. Additionally, the standard deviation dropped to 11.12, putting it at .05 points from the mean.

All of the technologies listed experienced this trend, except for pose estimation and object detection, whose standard deviation increased. However, while their standard deviation increased, the number of points between the mean and standard deviation dropped. When grouped together, the mean of expected false negatives for these technologies was 251.04 and fell to 145.10 when considering optimal false negatives.

5.4. Workflow

This section discusses results pertaining to the limitations, and possible improvement, of workflows. Table D.12 in Appendix D presents results on the limitations participants have encountered with their current workflow when investigating CSAM. Out of 106 survey participants, 66.04% provided input for this question. Of that percentage, only 65.71% provided useful feedback on their limitations and 25.71% claimed they did not have any limitations, while 8.58% submitted invalid data or the question was not applicable to them.

Results indicate that participants have limitations in their workflow across the board, with the majority pertaining to current tools and technology (28.77%). For instance, respondents are limited by the lack of filtering mechanisms to identify images, videos, and unrelated artifacts with their current tools. Hash filtering technologies, such as one from the National Institute of Standards and Technology (NIST),[11] are another problem because they are limited to filtering through known CSAM.

Participants also encountered limitations in their workload (20.55%), time (17.81%), and resources (10.96%). For example, the vast amount of data and devices to process, along with large caseloads, adds to the workload. Additionally, the lack of investigators, time, and advanced hardware and tools, to process and analyze data continuously increases backlogs.

The next question focused on the improvements participants would like to see in their current workflow. The same number of people that answered the previous question, provided a response (66.04%). Of this percentage, 84.28% suggested improvements to their current workflow, 4.29% did not have any suggestions, while 11.43% were not sure, submitted invalid data or found the question inapplicable to them.

Table D.13 in Appendix D illustrates results gathered from 83 suggestions. Most participants indicated needing more resources (36.14%) and better tools and technology (33.73%). For instance, employing more analysts/investigators would help ease the workload.Respondents also indicated that tools and technologies could be improved by adding efficient filtering mechanisms for identifying images, videos, and unrelated artifacts, along with more filtering options and task automation. Other suggestions include adding more CSAM hash databases and making it feasible for investigators nationwide to share hash values of such content; providing more training for management; and implementation of more realistic policies and standards that reflect the current workflow.

On the next question, participants were given the opportunity to rate the value of and suggest any enhancement to a proposed workflow design provided by BLINDED FOR REVIEW. The design was composed of rapid acquisition, multimedia file extraction and analysis, automated leads presentation, and safer presentation of

[11] https://www.nist.gov/.

explicit material. Over sixty-five percent (69 of 106) of all surveyed participants evaluated the suggested workflow, with most respondents finding some value in it. Specifically, most people (44.93%) found the design *valuable*, 34.78% found it *very valuable*, 18.84% found it *slightly to moderately valuable*, while only one respondent found no worth in it.

Only 42.45% (45 of 106) of those surveyed provided feedback about the suggested design. Of that percentage, 35.56% suggested improvements to the workflow, which are highlighted in Appendix D, Table D.14. The majority of suggestions (61.11%) indicated tools and technology as the main area for improvement. This included providing encryption and file hiding analysis to discover steganography attempts after rapid analysis, adding an explanation of how the automated process generated its results to test for false positives, and implementation of safer presentation at an earlier stage.

Some responses (13.33%) involved criticism of the safety presentation feature, suggesting that safer presentation of CSAM was not necessarily a concern. Some explained that CSAM still had to be found by the examiner and verified by several people in the chain of command. The prosecutor and jury, for example, are required to view such material. Two other participants also claimed that viewing CSAM was no longer an issue to them.

Additionally, 4.44% suggested that the automated leads presentation of the design is limited. This is because automation is not plausible when physical evidence is encountered, thus, manual analysis is still necessary. Additionally, automation can cause mistakes and miss evidence.

5.5. Tool processing times

5.5.1. Acquisition, processing, analysis, & reporting time

In this section, results from ten survey questions pertaining to approximate tool processing times are discussed. The questions focused on the estimated number of hours respondents' current tools may take when processing a complete case and as individual phases of a forensic investigation, such as acquisition, analysis, and report generation in a laboratory or on-scene setting. The case provided consisted of two phones with 60,000 images and three hours of video. Not all participants from the sample population ($n = 106$) answered all questions, thus, results do not have the same number of responses. Nevertheless, to provide a more accurate comparison, outliers were removed from the overall results and One-Sample T and Mann–Whitney tests were conducted.

Respondents were asked to approximate the typical amount of time it would take their current tool to process a CSAM case (acquiring, analyzing and generating a report) in a laboratory setting. They were also asked to suggest the preferred optimal processing times for such tools in the same setting. Only 83.96% of participants provided typical processing times, while 76.42% provided optimal times. Unsurprisingly, results indicate respondents prefer a tool that is faster than their current one. The average optimal time is 14.64 h ($SD = 12.30$) which is slightly over six hours less than the average typical time ($M = 21.05$, $SD = 15.01$). A One-Sample T test was performed, assuming that the respondents shared an accurate average time representative of the real world. This reinforced the finding that the optimal time was statistically significantly lower than the typical time, ($t(73) = 4.485$, $p = .000$). An interesting observation worth mentioning is that there is an enormous difference of 58.6 processing hours between the typical minimum (2.5 h) and maximum (61.1 h) processing times. Therefore, these results demonstrate how different investigator experiences are when using tools.

Continuing with typical times for processing in a lab setting, respondents were asked to estimate the number of hours their current tool may take when individually acquiring, analyzing, and creating an evidentiary report. Out of 106 participants, only 83.96% provided an estimate number of hours their tool takes to accomplish acquisition, while 83.02% approximated analysis times and 80.19% estimated reporting times. Results demonstrate that the average tool takes approximately over six more hours to analyze ($M = 17.80$, $SD = 15.26$) the evidence than acquiring ($M = 11.52$, $SD = 10.94$) the data, and over 11 h more than creating a report ($M = 5.94$, $SD = 5.40$).

Respondents were also asked similar questions about processing times within an on-scene or tactical setting. Approximately 69.81% of the sample population answered the question pertaining to typical processing time, while (66.04%) answered the question pertaining to optimal processing time. Results suggest respondents prefer an optimal average processing time of 5.54 h ($SD = 4.86$) when in the field. This ideal time is over six hours less than the normal average time ($M = 11.68$, $SD = 10.62$). A One-Sample T test ($t(58) = 9.703$, $p = .000$), suggested the optimal average time is statistically significantly lower than the typical average time. Lastly, while recognizing other factors involved in processing evidence in the field, it is intriguing that the submitted typical minimum amount of time was one hour, while the maximum (40.7 h) is 39.7 h longer.

Finally, respondents were asked to approximate the time their current tool takes when individually acquiring, analyzing, and creating an evidentiary report while working in the field. Fewer participants answered these questions when compared to the lab setting's feedback. In fact, 68.87% provided acquisition times, 62.26% presented reporting times, and 66.98% submitted analysis times. Results demonstrate that on average, analyzing ($M = 7.74$, $SD = 7.59$) data on the field takes almost two hours longer than acquiring ($M = 6.01$, $SD = 4.72$) data and over two hours longer than creating a report ($M = 5.13$, $SD = 5.63$).

While recognizing discrepancies in sample size feedback, all lab and on-scene processing times were compared to approximate which location may take the longest in completing investigative tasks. A Mann–Whitney test suggests that processing times in the laboratory were statistically significantly higher than the on-scene group ($U = 1509$, $p = .000$). It would take approximately over nine hours more to process a complete case in the lab than on the field.

When comparing lab and on-scene processing times for individual phases (acquiring, analyzing, and generating a report), conducting them in the lab setting also takes the longest. This conclusion was reached by conducting a Mann–Whitney test and comparing their averages. For instance, for the acquisition phase, a Mann–Whitney test suggests that the groups (lab and on-scene) were significantly statistically different ($U = 1783$, $p = .001$), taking over five hours more to acquire data in the lab than on the field. Moreover, the analysis phase takes the longest time to accomplish in both settings out of all phases. A Mann–Whitney test ($U = 1367$, $p = .000$), shows that the lab analysis times were statistically significantly higher than that of the field, taking over ten hours more to analyze the data in the lab than on the field.

On the other hand, creating a report on the field and in a lab setting takes almost the same average time. In fact, the difference is less than an hour. A Mann–Whitney test with a result of $U = 1937$, $p = .321$, suggests that distribution of times across both groups is similar. It is important to add that the lab setting group contained fourteen more responses than the on-scene setting, suggesting that if both groups had the same number of responses, then creating a report on the field may actually take longer than in the lab.

5.5.2. Acquisition processing times for android phones

Respondents were asked a series of six questions to estimate tool acquisition processing times for Android phones with different

storage capacities. Not all participants answered all questions, thus, outliers were removed from the final calculations and a Mann–Whitney test was conducted to provide comparison of statistics. Out of 106 respondents, 61.23% submitted acquisition times pertaining to the 8 GB phone, 62.26% for the 16 GB and 64 GB phones, 63.21% for the 32 GB phone, 60.38% for the 128 GB phone, and 55.66% for 256 GB phone.

As expected, the higher the storage capacity of a phone, the longer it would normally take for a tool to acquire an image. Therefore, tests were conducted to compare each capacity against the highest capacity (256 GB) to evaluate their differences. Calculations confirmed expectations that the averages and standard deviations increased the higher the storage capacity. For example, on average, it takes a 256 GB phone ($M = 8.39$, $SD = 6.71$), over six more hours to acquire an image than an 8 GB phone ($M = 2.17$, $SD = 2.04$). A Mann–Whitney test ($U = 394$, $p = .000$), shows that both groups are statistically significantly different, with the 256 GB phone producing higher acquisition times.

Calculations performed on the rest of the storage capacities resulted in the following: 16 GB ($M = 3.24$, $SD = 2.87$), ($U = 654$, $p = .000$); 32 GB ($M = 4.33$, $SD = 3.63$), ($U = 867$, $p = .000$); and 64 GB ($M = 5.62$, $SD = 5.14$), ($U = 1072$, $p = .002$). This shows that all groups are statistically significantly different when compared against the 256 GB phone. Comparison of the 128 GB phone ($M = 6.70$, $SD = 5.65$), ($U = 1265$, $p = .108$) resulted in a similar distribution group, suggesting that there is not much of a difference in acquisition times between the 128 GB and 256 GB, even though one has twice the capacity.

6. Limitations

The number of participants responding to questions was not the same across the board. This was due in part to intentional skipping of questions and early drop-out rates. This resulted in questions having varying response counts. Additionally, wording may have caused some questions to be misinterpreted by participants. As a result, several non-related answers were provided by respondents. Furthermore, an observation was made that some respondents were not consistent in their answers. For example, there is a difference in the number of people who selected “I do not use tools” when asked about the tools they use to process images and videos and the number of people who selected “I do not use tools” when asked what technologies were implemented by their tools. To address these concerns data cleansing was performed.

7. Discussion/conclusion

The results of this study demonstrated a few things, many of which were already known. Among them is the issue of demographic diversity. The majority of participants were older white males, a population that dominates the field. Additionally, participants generally had six years of experience and, aside from a few, an educational background not related to law, forensics, or security. Regardless of a relating educational background, most participants work at the local or state level as a digital forensic examiner or investigator.

The minimal level or lack of knowledge in Data Science is interesting given the amount of data encountered by practitioners. Practitioners should feel comfortable in their ability to obtain, process, and extract value from data acquired in an investigation. This may not only help in the investigation itself, but in resolving issues faced by practitioners, such as the inefficiency of some tools or processes. Skill development in this area would allow practitioners to contribute more to improving areas of concern.

On the job, examiners and investigators use a variety of tools, predominately commercial, to process images and videos. Most utilize more than one tool, each to target a specific need. Of the technology implemented by tools, filtering technologies appear to be the most useful and important to participants. While participants acknowledged that, if available, they would use safe-viewing technologies, these were not as important to them as filtering technologies. As many participants explained, they feel it is essential to the job to look at and confirm the results of processing; they feel they need to look at the explicit content. However, for some it may be beneficial in terms of their concerns regarding endurance and how it may limit their work. As one participant said, “an examiner can only look at so much child pornography at a time.”

In terms of limitations, common trends were found among responses. For tools, the main concerns are speed, accuracy, and reporting, which ultimately affect workflow and the speed of an investigation. Many respondents provided their own insight for an improved workflow to increase efficiency in a investigation. Ultimately, the sentiment across the field was that an improved workflow would require adequate funding, personnel, and time.

Finally, in hindsight, the researchers acknowledge that certain follow-up questions would have been beneficial. For example, those that indicated that they did not use any type of tools or technology could have been asked why tools or technology are not being used and what their work process entails. Such a question would have provided additional, and differing, insight regarding the topic at hand.

8. Recommendations

Our interaction with practitioners yielded actionable recommendations. Based on the analysis of the survey results and augmented by our interviews with practitioners, the following recommendations should be considered in the area of CSAM investigations:

- Incorporating courses on AI, software design, engineering, and data science into digital forensic programs to provide understanding of low-level concepts in data science.
- Establishing a continuous funding model to support research in CSAM investigations.
- Encouraging the development and use of open source tools focused on CSAM forensics by increasing “confidence in the tools through publication, review, and formal testing,” (Carrier, 2002).
- Establishing and implementing an up-to-date, standardized workflow, allowing for quicker investigations while minimizing the exposure of practitioners to CSAM.
- Encouraging non-practitioners, such as upper management, to engage in training for developing comprehensive knowledge of the work entailed in such investigations, and of the resources needed.
- Achieving accurate CSAM identification without the use of hash values by implementing state of the art AI techniques.
- Focusing research on the age estimation problem, especially for adolescents. This will help develop more accurate AI models for CSAM identification.
- Developing technology that can detect and cluster victim faces and apply age estimation techniques. The current state of the art open source models are not effective in this domain.
- Employing novel filtering techniques beyond the widely adopted skin tone detection. Practitioners noted that skin tone detection was not effective when examining large collections of CSAM.

- Leveraging novel techniques, such as object detection, to provide leads, allowing investigators to quickly identify locations, suspects, and victims in CSAM.
- Developing technologies that would allow tools to automatically upload newly identified CSAM to a centrally shared repository between authorities whilst adhering to legal standards.

Acknowledgements

We would like to thank all the digital forensic practitioners that took part in this study, sharing their experiences and insight. We would also like to thank Ahmed Alhishwan for his time and help in designing the survey. Additionally, we would like to thank MITRE for providing us the opportunity to work on this endeavor.

Appendix A. Demographics

Table A.2
Professional degrees amongst respondents. *Any percentage disparities due to rounding.

Degree	Count	Percentage
Accounting	1	1.14%
Agricultural Business Management	1	1.14%
Biology	1	1.14%
Business Administration and Management	4	4.54%
Computer Science	12	13.64%
Crime, Law, and Justice	25	28.40%
Cybersecurity	1	1.14%
Digital Forensics	15	17.04%
English	2	2.27%
Food Science	1	1.14%
Forensic Science	3	3.41%
General Science	1	1.14%
Geography	1	1.14%
Health and Physical Education	1	1.14%
Information Sciences and Technology	4	4.54%
Law Enforcement and Correction	7	7.95%
Political Science	2	2.27%
Psychology	1	1.14%
Sociology	3	3.41%
Other	2	2.27%
– Automotive Technology		

Table A.1
Demographics. *Any percentage disparities due to rounding.

	Count	Percentage
Race		
Asian	3	2.83%
White	99	93.40%
Other	4	3.77%
– Human		
Sex		
Female	14	13.21%
Male	92	86.79%
Age		
21–34	20	18.87%
35–44	38	35.85%
45–54	31	29.25%
55–64	17	16.03%
Level of Education		
High school diploma or equivalent	1	0.94%
Some college	17	16.04%
Associate's degree	22	20.75%
Bachelor's degree	44	41.51%
Master's degree	21	19.81%
Doctorate	1	0.94%
Employment Category		
Federal government employee	12	12.00%
State government employee	32	32.00%
Local government employee (city, county, etc.)	46	46.00%
Private, for profit	5	5.00%
Private, non-profit	1	1.00%
Contractor	2	2.00%
Other	2	2.00%
Occupation		
Digital Forensic Examiner	93	69.92%
Investigator	35	26.32%
Prosecutor	1	0.75%
Researcher	2	1.50%
Other	2	1.50%
– Cyber Security Professional		
– Manager - was investigator and digital forensics examiner (DFE)		
Time of experience working on CSAM cases		
Less than one year	5	5.00%
1–2 years	11	11.00%
3–4 years	13	13.00%
5–6 years	8	8.00%
More than 6 years	63	63.00%

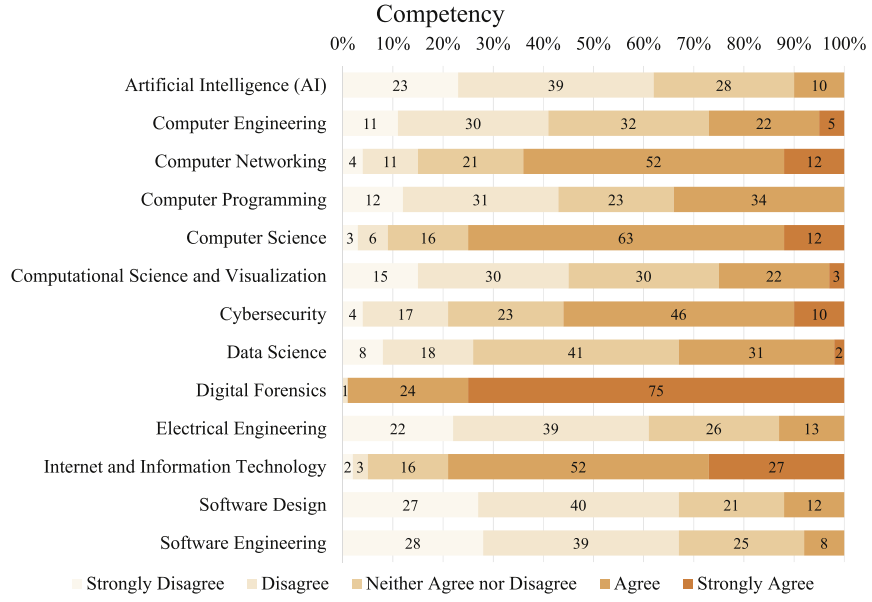


Fig. A.4. Technical fields respondents are competent at. Each bar represents one Likert scale question. Approximate percentages are displayed for each answer selection.

Table A.3
Locations where respondents received training in investigating CSAM.

Type	Location	Count	Total Count	Percentage
Academic Institutions	Dixie State University	1	14	7.95%
	Fox Valley Technical College (FVTC)	6		
	Purdue University	2		
	SANS Institute	2		
	University College Dublin	2		
	University of New Hampshire	1		
Children's Advocacy Centers (CAC)	Alabama (did not specify exact center)	1	5	2.84%
	Center for Child Protection, Austin	1		
	Childrens Hospital of Wisconsin	1		
	Dallas Children's Advocacy Center	2		
Forensic tool training	Access Data	4	25	14.20%
	Berla	1		
	BlackLight	1		
	Cellebrite	4		
	Forensic Toolkit	1		
	Guidance Software/EnCase	7		
	ICACCOPS	1		
	Magnet Forensics	2		
	GridCop	1		
	WetStone Technologies	1		
	Did not specify	2		
Government Funded Programs	Internet Crimes Against Children (ICAC)	22	50	28.41%
	National White Collar Crime Center (NW3C)	23		
	The National Center for Missing and Exploited Children (NCMEC)	5		
In office training or on-the-job experience	In-office	2	7	3.98%
	On-the-job experience	5		
Law enforcement/Police academy	Indiana Law Enforcement Academy	1	7	3.98%
	Maui Police Department Training Academy	1		
	The Police University College, Finland	1		
	Politieacademie, Netherlands	1		
	South Carolina Criminal Justice Academy	1		
	U.S. Air Force Special Investigations Academy (USAFSIA)	1		
	Did not specify	1		
Local Law Enforcement	Harris County Sheriff's Office, Child Abuse Unit	1	2	1.14%
	Montgomery County Maryland Police Department	1		
Non-Profit Organizations	ChildFirst Forensic Interviewing Protocol	1	16	9.09%
	European Cybercrime Training and Education Group (ECTEG)	1		
	Florida Childrens Services Council (CSC)	1		
	The Innocent Justice Foundation (TIJF)	1		
	The International Association of Computer Investigative Specialists (IACIS)	12		

Table A.4
Locations where respondents received training in investigating CSAM - continued.

Type	Location	Count	Total Count	Percentage
Private Organizations/Companies	Backbone Security	1	5	2.84%
	National Coalition of Advanced Technology Centers (NCATC)	1		
	Society for Worldwide Interbank Financial Telecommunication (SWIFT)	1		
	Teel Technologies	2		
State Bureau of Investigation	Oklahoma State Bureau of Investigation	1	2	1.14%
	Tennessee Bureau of Investigation	1		
State Level Department of Justice (DOJ)	Wisconsin	3	4	2.27%
	California	1		
State Police	Arkansas	1	3	1.70%
	Indiana	1		
	West Virginia	1		
State Programs	Child Abuse Training and Coordination Program (CATC), Oklahoma	1	1	0.57%
U.S. Department of Defense	Defense Cyber Investigation Training Academy (DCITA)	2	2	1.14%
U.S. Department of Homeland Security	Federal Law Enforcement Training Centers (FLETC)	4	12	6.82%
	Homeland Security Investigations (HSI)	1		
	Secret Service (including NCFI- National Computer Forensics Institute)	7		
U.S. Department of Justice (DOJ)	Federal Bureau of Investigation (FBI)	4	10	5.68%
	National Domestic Communications Assistance Center (NDCAC)	1		
	The Office of Juvenile Justice and Delinquency Prevention (OJJDP)	2		
	Did not specify	3		
Other	Attorney General's Office, South Carolina	1	4	2.27%
	Child Symposium, San Diego (did not specify conference or workshop)	1		
	Prosecutor's Office	2		
Unknown	Did not specify training	7	7	3.98%

Appendix B. Tools

Table B.5
Image Processing Tools.

	Count	Percentage
Image Processing Tools Utilized by Practitioners		
Analyze 16.1/Griffeye	54	10.07%
Autopsy	21	3.92%
Cellebrite Analytics	23	4.29%
Cellebrite UFED/PA	88	16.42%
Computer Aided Investigative Environment (CAINE)	6	1.12%
Digital Forensics Framework (DFF)	1	0.19%
EnCase Forensic	59	11.01%
EnCase Mobile Investigator	8	1.49%
Forensic Toolkit (FTK)	64	11.94%
Magnet Forensics IEF/Axiom	87	16.23%
NuDetective	1	0.19%
Oxygen Forensics Analyst	10	1.87%
Oxygen Forensics Detective	10	1.87%
Paraben E3: DS	1	0.19%
Paraben E3: Universal	1	0.19%
PhotoDNA	21	3.92%
PlainSight	1	0.19%
The Sleuth Kit	11	2.05%
VizX2/ZiuZ	1	0.19%
X-Ways Forensics	45	8.40%
Other	22	4.10%
– AccessData Lab		
– ADF Examiner		
– Blacklight		
– FastScan		
– FieldView		
– Forensic Explorer		
– ForensicScan		
– GrayKey		
– Lantern		
– Macquisition		
– Mobilyze		
– NetAnalysis		
– Nuix		
– Octoplus		
– Paladin		
– Paladin Recon		
– Secure View		
– Spada		
– XRY		
– Did not specify		
I do not use tools to process images	1	0.19%

Table B.6
Video Processing Tools.

	Count	Percentage
Video Processing Tools Utilized by Practitioners		
Analyze 16.1/Griffeye	52	11.79%
Autopsy	11	2.49%
Cellebrite Analytics	16	3.63%
Cellebrite UFED/PA	74	16.78%
Computer Aided Investigative Environment (CAINE)	3	0.68%
Digital Forensics Framework (DFF)	2	0.45%
EnCase Forensic	50	11.34%
EnCase Mobile Investigator	6	1.36%
Forensic Toolkit (FTK)	56	12.70%
Magnet Forensics IEF/Axiom	78	17.69%
Oxygen Forensics Analyst	8	1.81%
Oxygen Forensics Detective	8	1.81%
Paraben E3: DS	1	0.23%
Paraben E3: Universal	1	0.23%
PhotoDNA	13	2.95%
PlainSight	1	0.23%
The Sleuth Kit	7	1.59%
VizX2/ZiuZ	1	0.23%
X-Ways Forensics	35	7.94%
Other	16	3.63%
– AccessData Lab		
– ADF Examiner		
– Blacklight		
– Forensic Explorer		
– GrayKey		
– Paladin Recon		
– PhotoRec		
– VLC		
– XRY		
– Did not specify		
I do not use tools to process videos	2	0.45%

Table B.7
Tool Limitations.

Type	Limitation	Count	Total Count	Percentage
Features/Capabilities	Time-consuming	6	29	61.70%
	Cannot automatically group together several images of the same victim	1		
	Cannot carve images	1		
	Lacks obfuscation detection	1		
	Lacks ability to enhance poor quality photos	1		
	Lacks built-in definitions and clarity	1		
	Lacks customizable report format	1		
	Does not create reports that a court and jury could easily understand	1		
	Tools are not comprehensive	1		
	Compatibility and integration with other tools	1		
	Missing software components	1		
	Outdated software and features	1		
Accuracy	Accuracy of filtering features	4	11	23.40%
	Receiving false positives	4		
	Manual validation of results	2		
	Tool functioning	1		
User-Friendliness	Graphical user interface	2	3	6.40%
	Overly complicated to use	1		
Other	No limitations	3	4	8.50%
	Cost of tools	1		

Table B.8
Yahoo NSFW and iCOP/iCAC COP Results.

	Count	Percentage
Use of Yahoo NSFW and iCOP/iCAC COP		
Those that have used or currently use the Yahoo NSFW Image Classification Model	2	2.56%
Those that have used or currently use iCOP/iCAC COP	39	50.00%
Those that have used or currently use both Yahoo NSFW and iCOP/iCAC COP	1	1.28%
Those that have not used either Yahoo NSFW or iCOP/iCAC COP	36	46.15%
Benefits of Using Yahoo NSFW and iCOP/iCAC COP		
Quickness	11	22.92%
Identifying the presence of explicit content featuring children at a location or with a subject and monitoring P2P networks	10	20.83%
Saving time	7	14.58%
Filtering/filtering options to narrow down data	6	12.50%
Searching for known content (via hash values)	5	10.42%
Exam thoroughness	2	4.17%
Eliminating false positives/accuracy	2	4.17%
Receiving tips/leads	1	2.08%
Undercover case deconfliction	1	2.08%
Target investigative efforts	1	2.08%
Improving efficiency	1	2.08%
Automated reporting	1	2.08%
Limitations of Using Yahoo NSFW and iCOP/iCAC COP		
Only known files can be identified/detected	7	25.00%
Can throw off hash-value comparison with file alteration	3	10.71%
False positives	2	7.14%
Validation of data is still necessary	2	7.14%
Indiscriminate use of and trust in tools	2	7.14%
Cumbersome when dealing with large amounts of data	1	3.57%
Poor image quality	1	3.57%
Available geolocation tools	1	3.57%
Subject can hide their location	1	3.57%
Ability to obtain downloads from suspect IP can be affected by different variables	1	3.57%
Limited resources and training	1	3.57%
Large case loads	1	3.57%
Variation in child pornography laws	1	3.57%
No limitations/limitations are not the result of tools	3	10.71%
Not sure	1	3.57%

Appendix C. Technology

Table C.9
Technologies.

	Count	Percentage
Technology Implemented by Image Processing Tools		
Skin Tone Detection	76	56.30%
Face Recognition	6	4.44%
Face Detection	16	11.85%
Age Estimation	1	0.74%
Gender Estimation	1	0.74%
Child Nudity Detection	8	5.93%
Object Detection	5	3.70%
Face Presentation	1	0.74%
Nudity Blocker	3	2.22%
None of the above	13	963%
I do not use tools	5	3.70%
Technology Implemented by Video Processing Tools		
Skin Tone Detection	80	56.34%
Face Recognition	6	4.23%
Face Detection	16	11.27%
Gender Estimation	1	0.70%
Child Nudity Detection	10	7.04%
Object Detection	8	5.63%
Face Presentation	1	0.70%
Nudity Blocker	3	2.11%
None of the above	13	9.15%
I do not use tools	4	2.82%
Technology Utilized by Respondents to Process Images		
Skin Tone Detection	67	52.34%
Face Recognition	7	5.47%
Face Detection	10	7.81%
Age Estimation	1	0.78%
Gender Estimation	1	0.78%
Child Nudity Detection	5	3.91%
Object Detection	6	4.69%
Face Presentation	1	0.78%
Nudity Blocker	3	2.34%
None of the options provided	16	12.50%
I do not use any technologies	11	8.59%
Technology Utilized by Respondents to Process Videos		
Skin Tone Detection	57	46.34%
Face Recognition	5	4.07%
Face Detection	9	7.32%
Age Estimation	1	0.81%
Gender Estimation	1	0.81%
Child Nudity Detection	6	4.88%
Object Detection	6	4.88%
Nudity Blocker	2	1.63%
None of the options provided	25	20.33%
I do not use any technologies	11	8.94%

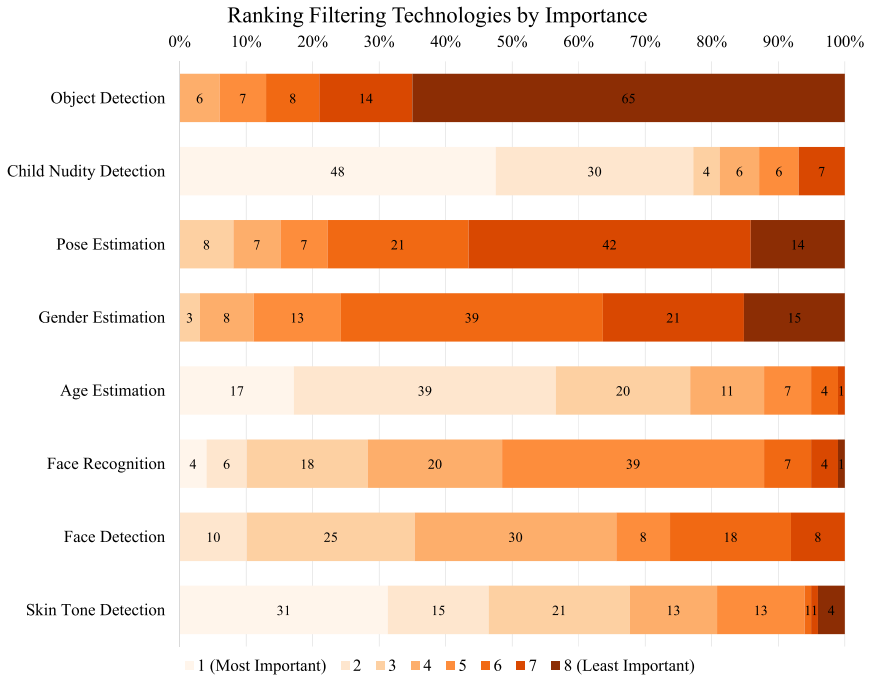


Fig. C.5. Each bar represents one Likert scale question. Approximate percentages are displayed for each answer selection.

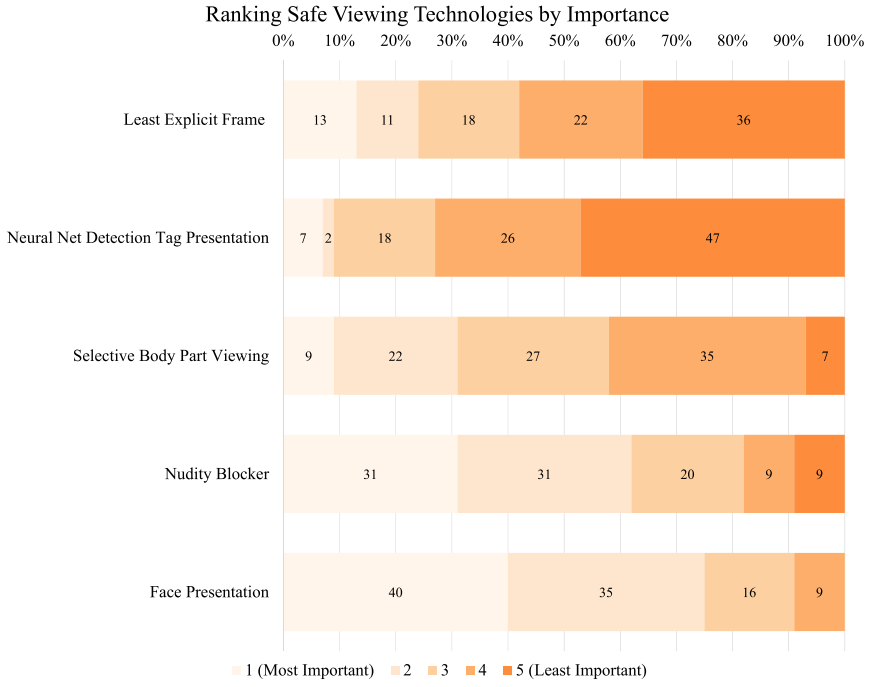


Fig. C.6. Each bar represents one Likert scale question. Approximate percentages are displayed for each answer selection.

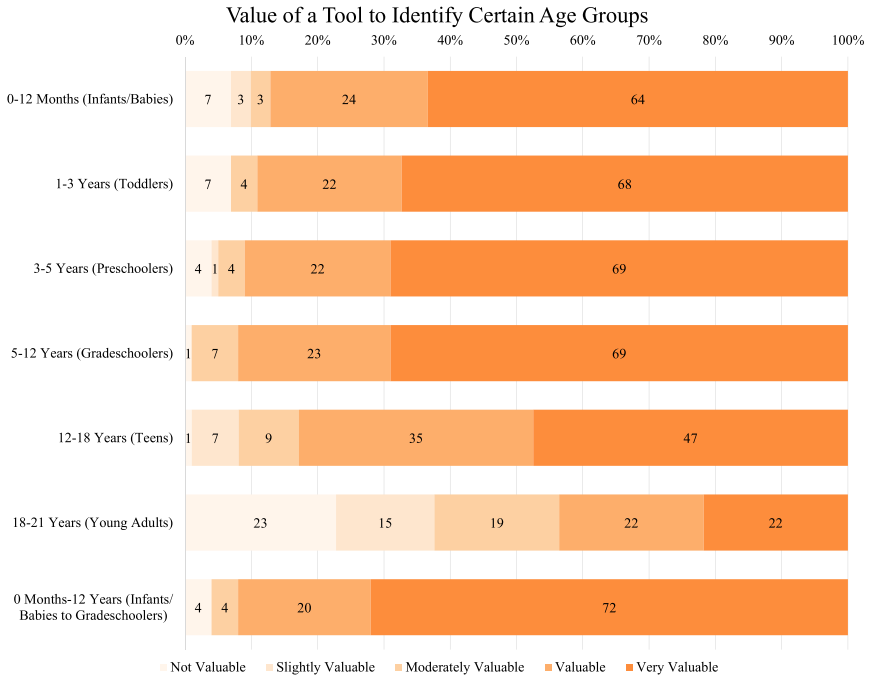


Fig. C.7. Each bar represents one Likert scale question. Approximate percentages are displayed for each answer selection.

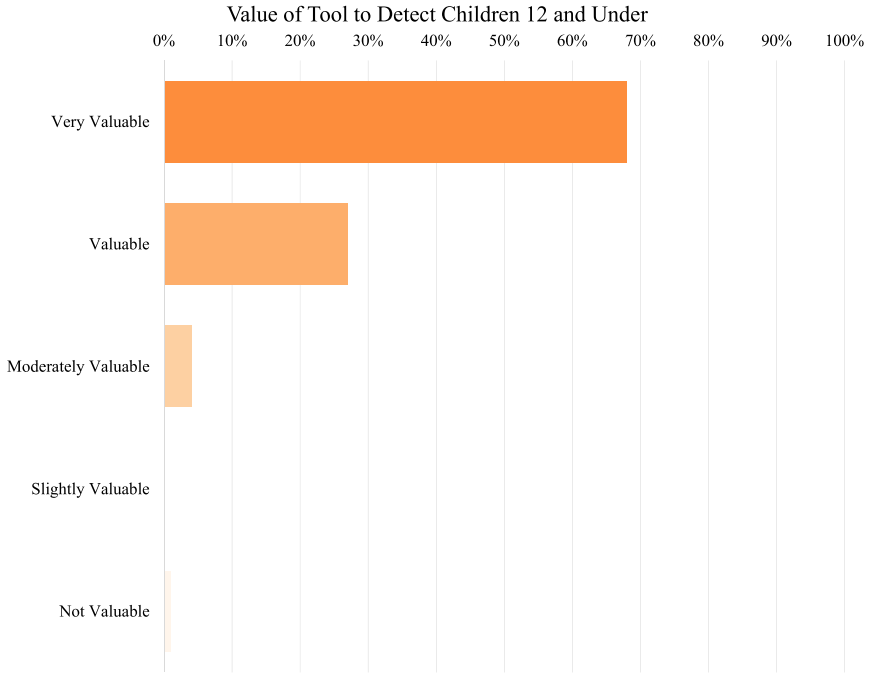


Fig. C.8. Each bar represents Likert scale options for the question. Approximate percentages are displayed for each option.

Table C.10
Expected v. Optimal Rates of False Positives.

	Expected Rate of False Positives				Optimal Rate of False Positives			
	N	Mean	Median	SD	N	Mean	Median	SD
Skin Tone Detection	60	25.01	20.35	15.62	52	15.68	12.55	13.23
Face Recognition	46	18.58	19.20	11.79	52	13.50	10.00	12.65
Face Detection	48	14.47	10.15	11.44	52	13.00	10.15	11.69
Age Estimation	44	27.84	25.75	14.74	52	14.47	10.45	13.20
Gender Estimation	42	18.96	16.20	14.22	52	14.57	9.95	13.91
Pose Estimation	42	19.60	17.90	13.98	52	15.08	11.60	13.40
Child Nudity Detection	45	23.82	20.40	14.85	52	13.63	9.80	13.33
Object Detection	41	19.62	19.00	13.80	52	13.23	9.95	12.45
Least Explicit Frame	41	18.39	17.70	13.65	52	13.06	10.00	12.42
Face Presentation	41	13.71	10.00	12.67	52	13.29	9.85	12.74
Nudity Blocker	41	17.61	15.00	13.68	52	13.16	9.85	12.39
Selective Body Part Viewing	40	19.11	18.4	14.52	52	14.23	10.00	13.55
Neural Net Detection Tag Presentation	40	19.91	18.4	14.27	52	14.35	10.00	13.52

Table C.11
Expected v. Optimal Rates of False Negatives.

	Expected Rate of False Negatives				Optimal Rate of False Negatives			
	N	Mean	Median	SD	N	Mean	Median	SD
Skin Tone Detection	58	22.09	20.15	13.51	47	11.21	9.90	10.65
Face Recognition	58	18.88	19.40	12.73	47	10.89	9.10	10.40
Face Detection	58	16.74	14.65	11.49	47	10.52	5.70	10.29
Age Estimation	58	22.99	20.50	14.86	47	11.17	6.70	11.12
Gender Estimation	58	18.36	16.85	13.35	47	11.88	9.20	11.93
Pose Estimation	58	17.41	15.75	12.03	47	12.58	9.40	12.21
Child Nudity Detection	58	19.94	18.70	12.78	47	10.40	6.10	11.02
Object Detection	58	14.67	14.20	10.11	47	11.06	8.40	10.44
Least Explicit Frame	58	16.45	15.20	12.43	47	10.53	5.10	11.40
Face Presentation	58	14.76	10.80	11.99	47	10.75	6.40	11.79
Nudity Blocker	58	17.35	14.25	13.46	47	11.02	9.20	10.42
Selective Body Part Viewing	58	18.08	18.55	12.71	47	9.95	5.50	10.60
Neural Net Detection Tag Presentation	58	18.39	18.75	12.82	47	10.98	7.40	10.95

Appendix D. Workflow

Table D.12
Current workflow limitations for investigating CSAM cases.

Type	Limitation	Count	Total Count	Percentage
Anti-Forensics	Encryption	1	2	2.73%
	File concealing applications	1		
Hardware	Compatibility	1	3	4.11%
	Computer speed	1		
	Storage	1		
Mental Health	Endurance	2	3	4.11%
	Stability	1		
Resources	Agency support	1	8	10.96%
	Funds for modern software/hardware	2		
	Personnel/examiners/investigators	4		
	Resources in general	1		
Time	Processing evidence	4	13	17.81%
	Return time of subpoenas, search warrants, and court orders/results	2		
	Time in general	7		
Tools andTechnology	Categorization & presentation of images in final report	1	21	28.77%
	Cloud storage	1		
	Detecting images	1		
	Efficient sorting through thousands of images	1		
	Hash filtering (NIST, etc)	4		
	Identifying ages of children	1		
	Not enough filtering mechanisms (images/videos/unrelated artifacts)	3		
	Outdated software/tools	1		
	Preview of videos	1		
	Processing images/videos	1		
	Restricted to validated tools only	1		

Table D.12 (*continued*)

Type	Limitation	Count	Total Count	Percentage
	Skin tone detection	1		
	Speed (acquisition/analysis)	2		
	Tool Integration	2		
Workload	Amount of data to review	6	15	20.55%
	Backlogs	2		
	Case Prioritization (Triaging)	1		
	Large caseload	4		
	Number of devices	2		
Other	Communicating with other investigators	1	8	10.96%
	Inaccurate leads from cyber tips	1		
	Lack of guidance and training on Project VIC	1		
	Presentation to court	1		
	Proper search authority	1		
	Shipping case materials	1		
	Speed	1		
	Travel	1		

Table D.13
Improvements to current workflow for investigating CSAM cases. *Any percentage disparities due to rounding.

Type	Suggestion	Count	Total Count	Percentage
Management/Standards	Administrative stress reduction	1	8	9.64%
	Better information from case agents	1		
	Better polices and procedures regarding prioritization	1		
	Digital forensics education/training for management	1		
	Fewer interruptions	1		
	Have supervisor verify evidence as CSAM	1		
	Meet and communicate with prosecutor/lawyer to determine counts to prosecute	2		
Reports	Better method to formulate report and definitions for court and jury	1	2	2.41%
	Replace general standardized reports with content-specific reports	1		
Resources	Analysts/examiners/investigators	13	30	36.14%
	More and better hardware/equipment	6		
	More and better software/tools	6		
	More funds	3		
	More resources in general	1		
	Training	1		
Time	More time in general	1	5	6.02%
	Speed processing times (tools/equipment)	4		
Tools & Technology	Ability to eliminate redundant images/videos found across multiple devices during analysis	1	28	33.73%
	Accessibility to and ease of use of hash value repositories via forensic tools	1		
	Allow for customizable and user friendly report generation	1		
	Allow seamless reporting between investigative applications	1		
	Automatic age recognition	2		
	Create standard format for cross-platform use	1		
	Improve accuracy of tools depicting online peer to peer activity	1		
	Improve filtering mechanisms (images/videos/unrelated artifacts)	9		
	Improving filtering options	3		
	Improve speed of media imaging	1		
	Improve speeds of recovering deleted content	1		
	Incorporate a tool for bypassing pass codes on Android mobile devices	1		
	Incorporate and utilize TensorFlow as a plugin for digital forensic tools	1		
	Provide link analysis (e.g., metadata report for all images)	1		
	Task automatization	2		
	Tools to sanitize reports/images for review by non-forensic personnel	1		
Other	Better training and distribution of Project VIC	1	10	12.05%
	Better training for cloud based investigations	1		
	Compel passwords for all devices and encryption	1		
	Improve inter-organizational collaboration in regards to timelines and expectations	1		
	Improve use of hash sets	1		
	Make it feasible for investigators nationwide to share hash values of CSAM	1		
	More hash databases of explicit content featuring children	1		
	Speed (processing, etc.)	2		
	Stay up to date with industry standards	1		

Table D.14
Improvements to given suggested workflow for investigating CSAM cases.

Type	Suggestion	Count	Total Count	Percentage
Resources	More analysts/examiners/investigators	2	3	16.67%
	More computers	1		
Time	Limit work hours of exposure to CSAM	1	1	5.56%
Tools & Technology	Ability to process large datasets faster	1	11	61.11%
	Add explanation of how automated process generated its result to check for false positives	1		
	Add report options	1		
	After rapid analysis, provide encryption and file hiding analysis to discover steganography attempts	1		
	Assign priority levels to cases based on known presence of CSAM	1		
	Automatically categorize and extract the content of any device (hard drive, usb key, etc.)	1		
	Faster acquisition and analysis	1		
	Implement safer presentation at an earlier stage	1		
	Quickly generate a report	1		
	Retain explicit format and safe presentation	1		
	Utilize reporting software with efficient filters to manually input the data necessary for a case	1		
Other	Availability of an expert witness to testify on behalf of the tool	1	3	16.67%
	System/workflow should allow for division of work into manageable tasks and collaborative/multi-user effort	1		
	Workflow should address data storage, network security, data retention and disposal, granular access and audit trail	1		

References

Anda, F., Lillis, D., Le-Khac, N.-A., Scanlon, M., 2018. Evaluating automated facial age estimation techniques for digital forensics. In: 12th International Workshop on Systematic Approaches to Digital Forensics Engineering (SADFE), IEEE Security & Privacy Workshops. IEEE.

Baggili, I., Marrington, A., Jafar, Y., 2014. Performance of a logical, five-phase, multithreaded, bootable triage tool. In: IFIP International Conference on Digital Forensics. Springer, pp. 279–295.

Bennett, D.J., Stephens, P., 2008. A usability analysis of the autopsy forensic browser. In: HAISA.

Bourke, M.L., Craun, S.W., 2013. Secondary traumatic stress among internet crimes against children task force personnel: impact, risk factors, and coping strategies. Sexual Abuse J. Res. Treat. 26, 586–609.

Burns, C.M., Morley, J., Bradshaw, R., Domene, J., 2008. The emotional impact on and coping strategies employed by police teams investigating internet child exploitation. Traumatology 14 (2), 20–31.

Carrier, B., 2002. Open Source Digital Forensics Tools: the Legal Argument. Technical report, stake.

de Castro Polastro, M., da Silva Eleuterio, P.M., 2010. Nudetective: a forensic tool to help combat child pornography through automatic nudity detection. In: 2010 Workshops on Database and Expert Systems Applications (DEXA).

Franqueira, V.N., Bryce, J., Al Mutawa, N., Marrington, A., 2018. Investigation of indecent images of children cases: challenges and suggestions collected from the trenches. Digit. Invest. 24, 95–105.

Hibshi, H., Vidas, T., Cranor, L.F., 2011. Usability of forensics tools: a user study. In: 2011 Sixth International Conference on IT Security Incident Management and IT Forensics, pp. 81–91.

Krause, M., 2009. Identifying and managing stress in child pornography and child exploitation investigators. J. Police Crim. Psychol. 24, 22–29.

Mahadeokar, J., Pesavento, G., 2016. Open sourcing a deep learning solution for detecting nsfw images. https://yahooeng.tumblr.com/post/151148689421/open-sourcing-a-deep-learning-solution-for.

Marturana, F., Tacconi, S., 2013. A machine learning-based triage methodology for automated categorization of digital media. Digit. Invest. 10, 193–204.

Peersman, C., Schulze, C., Rashid, A., Brennan, M., Fischer, C., 2016. icop: live forensics to reveal previously unknown criminal media on p2p networks. Digit. Invest. 18, 50–64.

Perez, L.M., Jones, J., Englert, D.R., Sachau, D., 2010. Secondary traumatic stress and burnout among law enforcement investigators exposed to disturbing media images. J. Police Crim. Psychol. 25 (2), 113–124.

PhotoDNA Cloud Service (n.d.). URL: https://www.microsoft.com/en-us/PhotoDNA.

Powell, M.B., Cassematis, P., Benson, M.S.B., Smallbone, S., Wortley, R., 2014. 'Police officers' perceptions of the challenges involved in internet child exploitation investigation. Policing: An International Journal of Police Strategies and Management 37 (3), 543–557.

Powell, M., Cassematis, P., Benson, M., Smallbone, S., Wortley, R., 2015. 'Police officers' perceptions of their reactions to viewing internet child exploitation material'. J. Police Crim. Psychol. 30, 103–111.

Quick, D., Choo, K.-K.R., 2014. Impacts of increasing volume of digital forensic data: a survey and future research challenges. Digit. Invest. 11, 273–294.

Rogers, M.K., Goldman, J., Mislan, R., Wedge, T., Debrota, S., 2006. Computer forensics field triage process model. In: Proceedings of the Conference on Digital Forensics, Security and Law. Association of Digital Forensics, Security and Law, p. 27.

Sae-Bae, N., Sun, X., Sencar, H.T., Memon, N.D., 2014. Towards automatic detection of child pornography. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 5332–5336.

Seigfried-Spellar, K.C., 2017. Assessing the psychological well-being and coping mechanisms of law enforcement investigators vs. digital forensic examiners of child pornography investigations. J. Police Crim. Psychol. 1–12.

Shaw, A., Browne, A., 2013. A practical and robust approach to coping with large volumes of data submitted for digital forensic examination. Digit. Invest. 10, 116–128.

Ulges, A., Stahl, A., 2011. Automatic detection of child pornography using color visual words. In: IEEE International Conference Multimedia and Expo (ICME).

Vidas, T., Kaplan, B., Geiger, M., 2014. Openlv:empowering investigators and first-responders in the digital forensics process. Digit. Invest. 11, S45–S53.

Vitorino, P., Avila, S., Perez, M., Rocha, A., 2018. Leveraging deep neural networks to fight child pornography in the age of social media. J. Vis. Commun. Image Represent. 50, 303–313.