A Framework for Human-Algorithm Teaming in Biometric Identity Workflows

John J. Howard[1], Laura R. Rabbitt[1], Isabelle M. Shuggi[1,*], and Yevgeniy B. Sirotin[1]
[1]The Identity and Data Sciences Laboratory at the Maryland Test Facility
(Authors listed alphabetically)
[*]Corresponding author (email: ishuggi@idslabs.org)

Biometric identity workflows are made up of multiple subsystems which can be automated to varying degrees using computer algorithms. Setting the appropriate level of automation for each subsystem is crucial for optimal system performance, which relies on synergistic human-algorithm teams. In this work, we leverage an automation design framework from prior research to define levels of automation for each biometric subsystem. This framework aligns a four-stage model of human processing with equivalent system functions. We propose applying this framework as a method to determine the appropriate allocation of tasks between humans and algorithms within a biometric identity workflow. While previous work has focused on the role of the human in the comparison and decision subsystems, we emphasize the need to consider the full biometric system in determining the role of humans in biometric identity workflows.

As technology continues to improve, human interactions with automation are constantly increasing. While human centered design principles aim to optimize technology for human users, these principles may not be implemented due to organizational constraints and/or technology being adapted for different use cases. Furthermore, as humans are required to increase their interactions with automation to complete complex tasks, the notion of a human-autonomy team should be considered (Lyons et al., 2021). As described by Lyons et al. (2021), human-autonomy teams are separate from human-automation interactions because automation is simply a tool to facilitate a workflow. When an automated system includes decision making and requires interdependence with a human, there is a shift towards considering the automated system as a teammate. In the context of biometric identity workflows, we specifically consider humans teaming with biometric algorithms (human-algorithm teams).

Human-algorithm teams are of interest because many technologies are deployed with a human operator to oversee or supervise the automation. Research focused on face recognition has suggested that humans and algorithms can be combined in ways that result in improved performance compared to each entity working alone (Phillips et al., 2018). However, other research has found that humans may limit the performance of these teams due to unconscious cognitive biases (Carragher & Hancock, 2022; Howard et al., 2020). While there are some disagreements in the literature as to whether humans and algorithms can obtain synergy in face recognition systems, it ultimately depends on the use case and what level of automation (LOA) should be implemented.

While many frameworks for human-autonomy teaming exist (e.g., Dekker & Woods, 2002; Donahue et al., 2022; Lyons et al., 2021; O'Neill et al., 2022), Parasuraman et al. (2000) present an adaptable framework which can be applied in various contexts. Parasuraman et al.'s (2000) work developed a model to define system functions and appropriate levels of automation (LOAs) for those functions based on human interactions with the system. This framework highlights the importance of designing for the human and not designing an "optimal" system with the human operator as an afterthought. Identifying the appropriate LOA allows for proper placement of the human within the workflow (i.e., the human's skills can be utilized to add value). Prior work on automation related to biometric systems has only focused on comparisons and decisions (e.g., Howard et al., 2020, Phillips et al., 2018). Here we apply Parasuraman et al.'s (2000) framework to the entire general biometric system with the aim of determining optimal task allocation between a human operator and each automated subsystem while simultaneously considering the capabilities of both.

BIOMETRIC SYSTEM OVERVIEW

Biometric characteristics are physiological measures that differentiate one person from another. There are many biometric modalities that can be used to differentiate people, such as fingerprints, face, and iris. When discussing biometric systems, international standard ISO/IEC 19795-1 defines seven subsystems comprising a generic biometric system: data capture, transmission, signal processing, data storage, comparison, decision, and administration. To maintain a generalized model similar to ISO/IEC 19795-1, this work excludes the transmission and administration subsystems (Figure 1). The following terms will be used in describing the relevant subsystems (ISO/IEC 2382-37):

- Biometric characteristic: a characteristic from which distinct, repeatable features can be extracted from individuals
- Biometric sample: holistic representation of biometric characteristics (e.g., image of entire face, fingerprints, or iris)
- Biometric feature set: mathematical representation of biometric characteristics (e.g., distances between points on face, fingerprints, or iris)
- Reference: one or more stored biometric samples or feature sets attributed to an individual and used as a baseline sample during comparison

- Probe: a biometric sample or feature set captured at another point in time and compared to a reference
- Identification: process of comparing a probe against a database of references to return a single biometric reference (i.e., who is this individual?)
- Verification: process of confirming an identity claim through a comparison to a reference (i.e., is this individual who they claim to be?)

Verification and identification are two separate processes (comparison of two images and comparison of one image to many images, respectively), which occur through the same general biometric identity workflow. As previously mentioned, to maintain a general biometric workflow this analysis focuses on five subsystems: data capture, signal processing, data storage, comparison, and decision (Figure 1).

Data capture subsystem. A biometric device that captures an image of an individual's biometric characteristics (e.g., face, iris, fingerprint). The captured output is then sent to the signal processing subsystem.

Signal processing subsystem. The image from the data capture system undergoes a quality assessment to determine if a request to re-capture an image should be sent back to the data capture subsystem user and/or operator. Once the image quality requirement is satisfied, the signal processing system creates a biometric feature set from the captured image. During enrollment, this feature set is saved in the data storage subsystem and becomes a biometric reference. During verification and identification, this feature set is sent to the comparison subsystem.

Data storage subsystem. References are stored in the data storage subsystem and sent directly to the comparison subsystem when enrolled biometric characteristics are presented.

Comparison subsystem. The probe is compared to the reference which results in a comparison score. This score is sent to the decision subsystem. For verification one score is generated and for identification a score is generated for every image in the reference database.

Decision subsystem. Verification and identification outcomes are generated by the decision subsystem based on comparison scores from the comparison subsystem. Scores above a pre-defined threshold will determine if the probe is a match (verification) or on a candidate list (identification).

HUMAN PROCESSING AND SYSTEM FUNCTIONS

Parasuraman et al. (2000) present a simplified human information processing model, which defines four stages of processing: sensory processing, perception/working memory, decision making, and response selection. These four stages of human processing are equated to four types of system functions: information acquisition, information analysis, decision and action selection, and action implementation. For each stage of processing and system function we provide examples for face and iris biometric modalities to highlight differences between humans and algorithms.

Sensory processing and information acquisition. Human sensory processing includes acquiring information through auditory, visual, and tactile cues. Similarly, the first phase of information processing for an automated system is information acquisition. At this point both the human and automated system are processing information relevant to the required task in preparation for the next phase of information processing.

- Humans process faces holistically (Tanaka & Farah, 1993), algorithms process faces differently depending on how the algorithm was developed.
- In general, humans are unfamiliar with iris patterns and do not employ this modality to recognize other humans without an algorithm.

Perception/memory and information analysis. Manipulation, integration, and interpretation of any acquired information will occur during the perception stage for humans. At this stage humans also store and retrieve relevant information from working and long-term memory. Automated systems perform information analysis in a similar manner, manipulating and integrating relevant data to either present to the human operator or act on itself.

- Humans are highly capable of comparing familiar faces but struggle when comparing unfamiliar faces (Megreya & Burton, 2006). Algorithms are highly capable of comparing faces when appropriate reference images are available (Cook et al., 2019).
- Algorithms can compare iris images.

Decision making and decision/action selection. At this point the human and automated system have fully processed the acquired information and are ready to select one of the available options. The number of available options can vary depending on the objectives of the system.

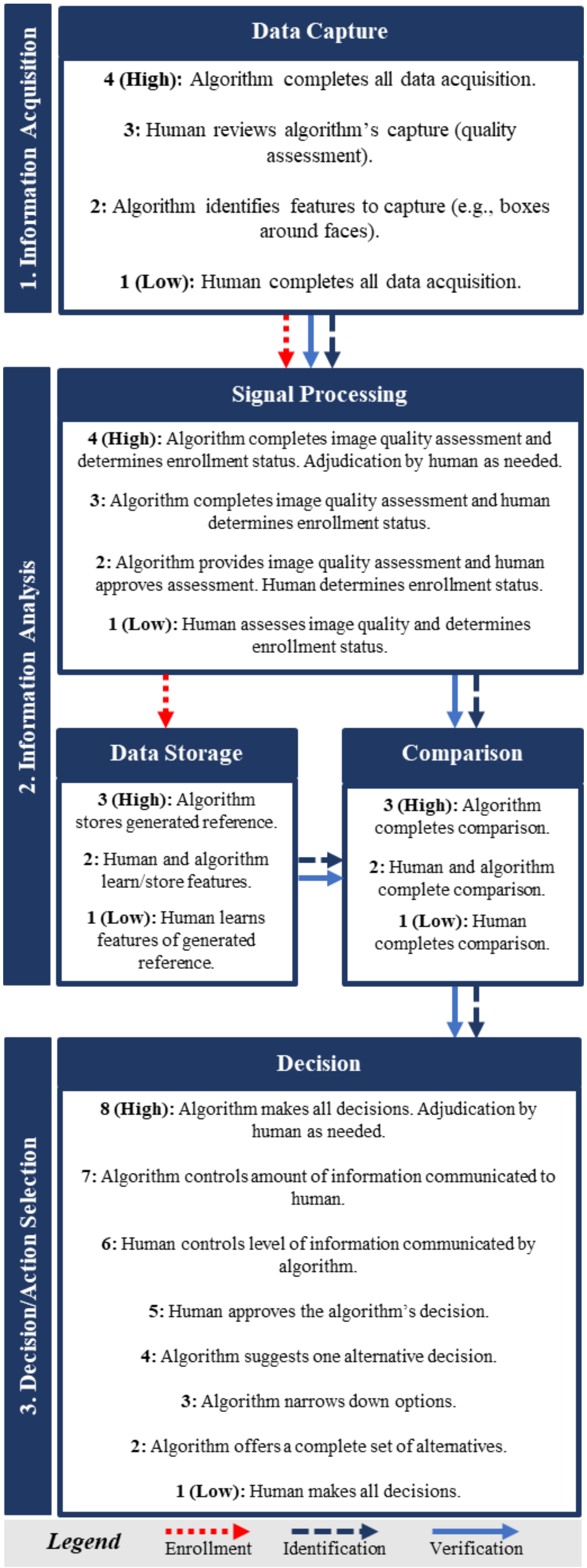
- Humans can decide whether faces match; however, algorithms generally match faces at higher levels of accuracy (Carragher & Hancock, 2022).
- Algorithms can decide if iris images are a match with high levels of accuracy.

Response selection and action implementation. Once a decision has been made by the human or the automated system, the final phase is the execution of that decision. This process varies by task complexity and may require one or more actions from the human or automated system.

- This stage of processing and system function is not currently incorporated into the general biometric system model as defined by ISO/IEC 19795-1.

BIOMETRIC IDENTITY WORKFLOWS AND AUTOMATION DESIGN

Our review of the general biometric system employed the four system functions model to determine appropriate task allocation between humans and algorithms for each of the defined subsystems. We found that the general biometric system incorporates three of the four system functions defined by Parasuraman et al. (2000). As previously stated, action implementation occurs outside of the biometric system once a decision outcome is established.



4 (High): Algorithm completes all data acquisition.

3: Human reviews algorithm's capture (quality assessment).

2: Algorithm identifies features to capture (e.g., boxes around faces).

1 (Low): Human completes all data acquisition.

4 (High): Algorithm completes image quality assessment and determines enrollment status. Adjudication by human as needed.

3: Algorithm completes image quality assessment and human determines enrollment status.

2: Algorithm provides image quality assessment and human approves assessment. Human determines enrollment status.

1 (Low): Human assesses image quality and determines enrollment status.

3 (High): Algorithm stores generated reference.

2: Human and algorithm learn/store features.

1 (Low): Human learns features of generated reference.

3 (High): Algorithm completes comparison.

2: Human and algorithm complete comparison.

1 (Low): Human completes comparison.

8 (High): Algorithm makes all decisions. Adjudication by human as needed.

7: Algorithm controls amount of information communicated to human.

6: Human controls level of information communicated by algorithm.

5: Human approves the algorithm's decision.

4: Algorithm suggests one alternative decision.

3: Algorithm narrows down options.

2: Algorithm offers a complete set of alternatives.

1 (Low): Human makes all decisions.

Figure 1. Application of Parasuraman et al.'s (2000) framework to biometric subsystems as defined by ISO/IEC 19795-1. LOAs were determined based on the current state of biometric technology.

Each system function incorporated into the general biometric system has uniquely defined LOAs and by extension, unique task allocation between teammates depending on the biometric identity application (Parasuraman et al., 2000). Information acquisition and decision/action selection are each mapped to one biometric subsystem leading to a relatively simple breakdown of automation levels (Figure 1). Information analysis is presented with unique LOAs for each subsystem, which need to be considered separately. LOAs are defined based on the current state of the biometric identity field; however, technological advancements will alter LOAs. Selection of automation levels is dependent on the biometric modality and identity workflow. Across all subsystems, the lowest LOA is broadly defined by the human completing the task independent of the algorithm.

We defined four LOAs for data capture with the anchor of high automation defined as the algorithm completing data acquisition alone. Within information analysis, signal processing is defined by four LOAs, notably the highest LOA may still require adjudication by the human as needed. Additionally, determination of whether an individual is enrolled or unenrolled is never divided between the human and algorithm due to differences in processes. The data storage and comparison subsystems each have three LOAs. At the second LOA both the human and algorithm perform recall and complete a comparison (e.g., a security guard remembers a person they frequently see pass through an entrance with a biometric system). Lastly, we defined eight LOAs for the decision subsystem, where the highest LOA may still require adjudication by the human. At the fifth LOA, there is an increased risk of complacency since the human provides approval for a decision made by the algorithm.

Algorithms generally exhibit high accuracy within the comparison subsystem, allowing human operators to add value in a complementary manner (e.g., differentiating twins based on features the algorithm may not consider). Humans can also add value to other subsystems, leading to improvements in overall performance. For example, many operational biometric identity systems currently employ the highest LOA at the data capture stage; however, research has demonstrated that the majority of biometric system errors are due to photo capture (2021 Biometric Rally Results). Biometric systems with operators, should focus human efforts on data capture to decrease these types of errors, which could improve system performance across modalities (e.g., face, iris, fingerprints).

Automation Determination Flowchart

Based on the automation determination flowchart from Parasuraman et al. (2000) we developed a set of criteria specific to the automation of a biometric system in the context of human-algorithm teaming (Figure 2). Our flowchart highlights two sets of criteria to determine an optimal automation level for each of the identified system functions when applied to a specific type of biometric system. The first set of criteria highlights the importance of team performance including the human's cognitive performance, and the second set of criteria focuses on broader impact criteria including organizational and societal impacts. Both sets of criteria are presented as part of an iterative process, where optimization of team performance is reached before introducing broader impact criteria.

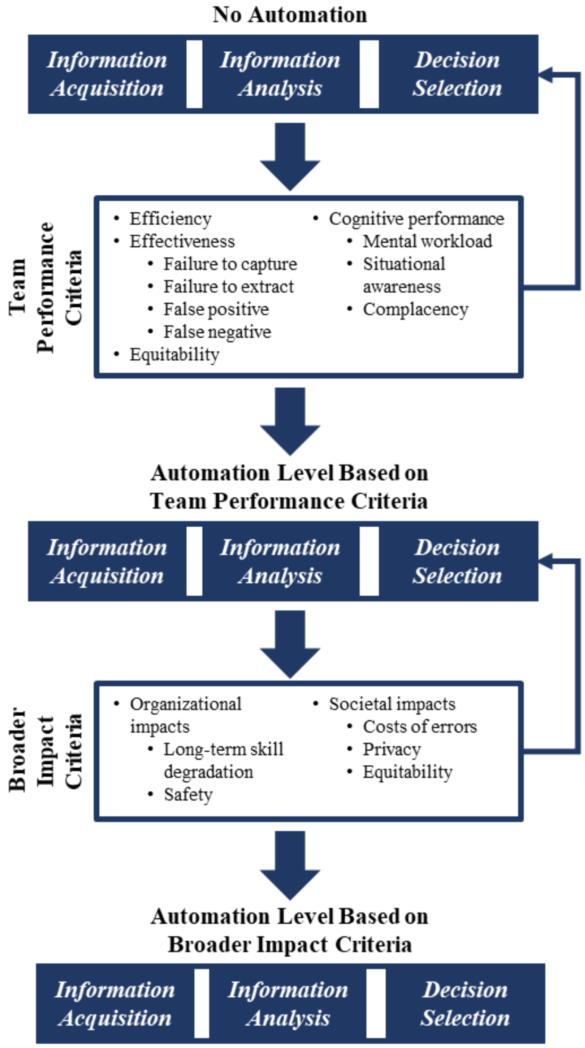


Figure 2. Automation determination flowchart. Shows three relevant system functions and criteria for biometric systems based on a review of Parasuraman et al. (2000).

Team performance criteria. We define four categories to assess team performance: efficiency, effectiveness, equitability, and cognitive performance. These categories characterize the performance of biometric identity workflows and importantly, provide consideration of the human. Assessment of system outcomes (i.e., efficiency, effectiveness, and equitability) and the human's cognitive state allow for optimal placement of the human within the workflow.

Efficiency is defined as the resources (e.g., time and staff) required by the biometric system to determine an individual's identity. Effectiveness is the accuracy and completeness with which the system can determine an individual's identity. This includes metrics such as failure to capture, failure to extract, false positives, and false negatives. Equitability is the extent to which the system performance is invariant across specific demographic categories and biological phenotypes. All these categories are focused on the overall team's performance; however, it is also important to separately assess the human's cognitive performance.

When incorporating the human's cognitive performance, three key factors must be considered: mental workload, situational awareness, and complacency (Parasuraman et al., 2000). Pairing a human with an autonomous teammate does not guarantee that the human's contributions or team performance will be improved. It is possible for the human to experience an increase in mental workload, a decrease in situational awareness, and an increase in complacency, which could all negatively impact the team's performance. However, if each of these factors are properly assessed alongside the system outcomes of efficiency, effectiveness, and equitability then the human's skills can be optimally placed within the workflow.

Broader impact criteria. Two main categories are defined as part of our broader impact criteria: organizational and societal impacts. Organizational impacts are defined as those which could potentially affect employees and the structure of the organization. A key factor of organizational impact is the potential for the human to experience skill degradation over an extended period if certain tasks have been allocated to the algorithm (Parasuraman et al., 2000). Additionally, in the context of a biometric identity system, the safety of employees could be jeopardized if tasks are not correctly allocated between the human and algorithm (e.g., a bad actor is allowed to enter a restricted space).

Societal impacts are defined as those which could have a wider effect on society. The cost of an error from a biometric identity system could range from minimal (e.g., additional time is needed to verify an individual resulting in increased wait times) to quite detrimental (e.g., a bad actor carries out an attack in a densely populated area). Furthermore, as biometric identity systems become more prominent, vast amounts of data will be collected, which could lead to privacy concerns such as individuals being tracked, or their data being used without their consent. Increased use of biometric identity systems could also increase equitability concerns related to how these systems are applied across different groups (Howard et al., 2021).

Both evaluation criteria are equally important but should be implemented in a serial manner. By first implementing criteria to optimize team performance, the number of errors experienced by a larger population can be mitigated.

USE CASE EXAMPLES

To illustrate the proposed human-algorithm teaming framework in Figure 1, we review two use cases where LOAs can be minimized or maximized. The presented use cases focus on face recognition due to its deployment across various applications and humans' unique ability for innate face recognition. The first use case we consider is at an airport security checkpoint where security personnel verify a traveler's identity. In this scenario, a human operator adds value to the data capture (could be set to level 3) and signal processing (could be set to level 4) subsystems by reviewing the algorithm's capture and image quality assessment. The algorithm adds value to the data storage, comparison, and decision subsystems (each subsystem could be set to the highest LOA). Specifically, consider when a probe image is captured in real time and compared to the presented identification card. The human operator adds value through contextual awareness by quickly recognizing and handling exceptions (e.g., confirm the individual is not wearing a mask or sunglasses, which impact biometric matching). However, humans do not perform well with unfamiliar face matching under a time constraint. The algorithm can quickly and accurately complete the comparison to the reference image on the identification card and decide if the probe is a match. In this use case, task allocation employs each teammate's strengths to keep travelers moving through the checkpoint efficiently while maintaining a high level of security.

A contrasting use case to consider is a forensic examiner tasked with identifying a suspect based on multiple pieces of information including images of suspects captured at different angles. The data capture and data storage subsystems are out of scope for this use case. The signal processing subsystem could be set to the highest LOA. The comparison subsystem could be set to an automation level of 2, where both the human and algorithm complete the comparison. The decision subsystem could be set to level 3, the algorithm may return a candidate list of potential matches based on the facial comparison. Examiners will then follow a structured set of guidelines to determine if any of the candidates are a true match (FISWG, 2018; 2022). They may use additional information, such as distinctive scarring, tattoos, marks, or other physical features on a suspect's arm and a timeline of relevant events. Here the human adds significant value to the subsystem's performance due to their ability to incorporate this additional information into the final decision. Furthermore, the ability to explain the decision-making process is crucial and currently an ability unique to humans.

CONCLUSION

We present an automation determination framework based on Parasuraman et al.'s (2000) model and adapted it for biometric systems as defined by ISO/IEC 19795-1. We consider the entire biometric system to determine LOAs on a more granular level. The defined LOAs are based on the current state of technology and aim to optimize placement of the human within the workflow. Appropriate utilization of the human's skills not only improves the cognitive state of the human but can also improve the overall performance of the biometric system. While biometric research has primarily focused on the forensic scenario, there is a need to address use cases which involve a large portion of the general population (e.g., airport security checkpoints, accessing restricted areas, etc.). As advancements in technology across fields continue, it is important to consider human-algorithm teams and the possibility of the human in a non-supervisory role.

ACKNOWLEDGEMENTS

This research was funded by the U.S. Department of Homeland Security, Science and Technology Directorate under contract number W911NF-13-D-0006-0003. The views presented here are those of the authors and do not represent those of the Department of Homeland Security, the U.S. government, or their employers. Author contributions are listed with each author's initials. JJH: developed framework and edited the manuscript. LRR: developed framework, wrote and revised manuscript. IMS: developed and applied framework, wrote and revised manuscript. YBS: conceived the work, directed framework development, and revised the manuscript.

REFERENCES

2021 Biometric Rally Results. (2021). MdTF.org. Retrieved February 10, 2023, from https://mdtf.org/Rally2021/Results2021

Carragher, D. J., & Hancock, P. J. (2022). Simulated automated facial recognition systems as decision-aids in forensic face matching masks. *Journal of Experimental Psychology: General*.

Cook, C. M., Howard, J. J., Sirotin, Y. B., Tipton, J. L., & Vemury, A. R. (2019). Demographic effects in facial recognition and their dependence on image acquisition: An evaluation of eleven commercial systems. IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(1).

Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work*, *4*, 240-244.

Donahue, K., Chouldechova, A., & Kenthapadi, K. (2022, June). Human-algorithm collaboration: Achieving complementarity and avoiding unfairness. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1639-1656).

FISWIG. (2018). *Facial image comparison feature list for morphological analysis.*https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf

FISWG. (2022). *Facial Comparison Overview and Methodology Guidelines.* https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V2.0_2022.11.04.pdf

Howard, J. J., Rabbitt, L. R., & Sirotin, Y. B. (2020). Human-algorithm teaming in face recognition: How algorithm outcomes cognitively bias human decision-making. *Plos one*, *15*(8), e0237855.

Howard, J. J., Sirotin, Y.B., Tipton, J. L., & Vemury, A. R (2021). *Quantifying the extent to which race and gender features determine identity in commercial face recognition algorithms*. DHS Technical Paper Series.

ISO. (2021). *Information technology – Biometric performance testing reporting – Part 1 Principles and framework* (ISO/IEC 19795-1:2021). https://www.iso.org/standard/73515.html

ISO. (2022). *Information technology – Vocabulary – Part 37: Biometrics* (ISO/IEC 2382-37: 2022). https://www.iso.org/standard/73514.html

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human–autonomy teaming: Definitions, debates, and directions. *Frontiers in Psychology*, *12*, 589585.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. Memory & cognition, 34, 865-876.

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors*, *64*(5), 904-938.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, *115*(24), 6171-6176.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Q. J. Exp. Psychol.* 46A, 225–245.