Integration of Augmented Reality Vision and Speech Modules into The Cognitive Assistant System
(Technical Paper)

Investigating the effects of biased facial recognition AI algorithms used by law enforcement in criminal investigations on people of color
(STS Paper)

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Sneha Iyer

October 27, 2022

Technical Team Members: Sneha Iyer

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Rider Foley, Department of Engineering and Society

Homa Alemzadeh, Department of Electrical and Computer Engineering and Computer Science

Cognitive Assistive Technology to help EMTs with Patient Care in Emergency Response Situations

In the field of emergency medicine, health specialists utilize a variety of information available in a situation in order to provide the best care possible to a patient in dangerous circumstances. According to Kim et al. (2021), in emergency medical care processes, the responders collect large amounts of data with different levels of importance and confidence, including the patient's past medical history, their present medical conditions, and interventions performed. There are many standards and tools for collecting, storing, and distributing emergency medical services (EMS) data (Becknell and Simon, 2016). However, more attention should be given to reliably translating this variety of information into actionable knowledge for assessing and performing emergency operations. Additionally, recalling this information requires cognitive effort in a crisis situation. If some of the responsibility for aggregating and analyzing information is delegated to assistive technologies, then more cognitive effort can be channeled into improving the speed and precision of pre-hospital care.

Emergency medical technicians (EMTs) collect, filter, and interpret information from real-time sources in order to provide timely and appropriate medical interventions during emergency situations. However, doing so in a high-pressure situation causes a cognitive strain on those performing during a crisis (Lawn et al., 2020). Assistive technologies can help to lessen this pressure on first responders by improving situational awareness and facilitating appropriate decision making (Holthe et al., 2022). Additionally, the integration of machine learning technologies within cognitive assistant systems aims to improve the accuracy and effectiveness of EMTs by utilizing analytical algorithms that collect heterogeneous data streams from the incident scene, aggregate that data with publicly available data, extract valuable information, and provide applicable feedback.

In an emergency, necessary activities in the scene, prehospital, and in hospital setting must be conducted, as precisely and quickly as possible. According to Kalhori (2022), technologies such as AI might be a beneficial support to achieve this crucial aim. Already in emergency departments in some hospitals, AI has been applied for predictive modeling, patient monitoring, and day-to-day running of emergency departments. These intelligent tools support health care providers in reducing waiting times in the emergency department, decreasing errors, and increasing the efficiency of care.

Thus, the technical topic will focus on working on a part of a Cognitive Assistant system that acts as an artificial intelligence (AI) agent that will be designed to assist the first responders observing and processing the data and interacting with responders during response operations to provide them with reminders, feedback, and insights to improve their situational awareness and operations outcome.

Ongoing Research and Integrating Vision and Speech Modules into the Cognitive Assistant

Currently, in an emergency response situation, which is usually initiated by a 911 call, EMS responders follow a certain procedure that produces a flow of information. According to Kim et al. (2021, p. 3), "In each call, responders are dispatched to the incident location and informed of the "Call Type," which is the general reason for the incident. On arrival, responders interact with the patient and others to identify the "Chief Complaints," which are the primary reasons for the EMS call. Then, the responders use the patient's Chief Complaints, Signs and Symptoms, past medical history, history of the present illness (HPI) or injury, and current presentation to form a set of "Impressions." From the Impressions, the responders select and follow the appropriate "EMS Protocol Guidelines" to perform "Interventions" (including

"Procedures" and "Medications"), which are a series of treatments to stabilize the patient before transporting to the hospital. Responders document this information flow from Call Type to Chief Complaints to Impressions and, finally, to Interventions, along with other information described as "Narrative" and/or "Medic Notes", in the EMS incident reports".

The CognitiveEMS system in particular focuses on processing this information collected from the responder, which includes stages for converting speech to text, extracting medical and EMS protocol specific concepts, and modeling and execution of an EMS protocol (See Figure 1 for an architectural overview of this system). Specifically, when a first responder speaks to a patient, the CognitiveEMS system converts speech to text using a speech to text library. Negation detection, value retrieval, and concept mapping to protocols are performed on the converted text for each sentence. Based on the information from the converted text, as well as information from sensors connected to the wearable device and the patient, the system predicts relevant EMS protocols, and provides reminders and suggestions to the first responder. The main technologies used in the system are the Google Speech API for the speech to text conversion, MetaMap and CLAMP software for concept extraction, and a rule engine for enabling the policies and operational decisions to be defined, tested and executed (Preum et al., 2019). The pipeline utilizes a behavior tree framework and machine learning (ML) methods to extract critical information from both audio and visual input and recommend interventions to specialists.

Currently, there are offline, individual ML modules for speech and vision that are separate from the pipeline. My capstone project centers around integrating these modules for speech and vision into the current pipeline, managing the various input and output streams, as well as processing and demoing the real-time results on a graphical user interface (GUI) in the Cognitive assistant. It will also focus on creating a vision module using input from Augmented

Reality (AR) glasses into the pipeline and using ML to analyze and interpret the data to help the assistant recommend protocols and interventions to EMS specialists. Finally, I will also be building an Android application to capture streaming audio data from the AR glasses and integrate that into the system.

Early identification of necessary protocols and/or interventions by a system such as the Cognitive Assistant in an emergency situation may save patients' lives by helping EMT's to make quicker and more accurate decisions regarding the patients' transfer and care. Additionally, systems such as the Cognitive Assistant allow us to analyze the ways that AI and humans interact and collaborate. A better understanding of this will help grow potential for humans to learn from and collaborate with algorithms in an ethical manner.

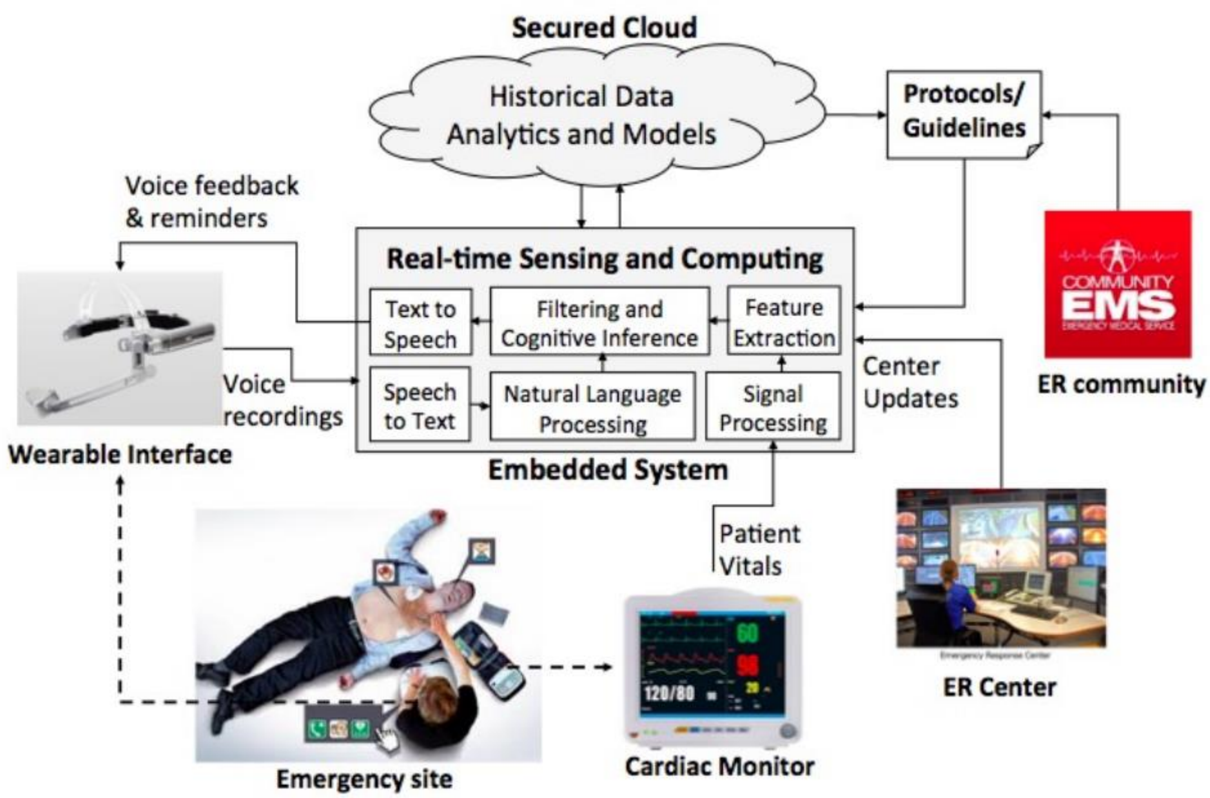

Figure 1. The architecture of the CognitiveEMS system (Preum et al., 2018)

Investigating the Effects of Biased Facial Recognition AI Algorithms Used by Law Enforcement on People of Color

While the emergency care cognitive assistant technical project has significant human and social dimensions, the societal aspect of this prospectus transitions to examining biased facial recognition AI algorithms used by law enforcement in criminal investigations on people of color. Facial recognition is a biometric technology that is used to identify a person's identity by mapping out data points of a person's facial features. It uses AI algorithms to learn how to identify a specific person and verify them against images/videos in a database (Symanovich, 2021). U.S. law enforcement has used facial recognition technology (FRT) since the 2000's and in 2016 one in four US state or local police departments have access to facial recognition technology (Horowitz, 2020). However, the efficacy and accuracy of algorithms used for law enforcement is questionable and there have been several instances of wrongful arrest, false imprisonment, and loss of employment due to inaccuracies of this technology (Hill, 2021). Thus, due to the disparities in accuracy rates and current uses of FRT, there are significant risks of surveillance abuse and disparate impacts on minority groups due to FRT practices by law enforcement.

A survey by the Pew Research Center supports this claim, showing mixed opinions on the effects facial recognition would have on false arrests. Some 53% of U.S. adults say police probably or definitely would make more false arrests if use of facial recognition technology was widespread among police, while 45% say this probably or definitely would not happen. Based on the survey results, it was also clear that there were some notable differences among racial and ethnic groups on these issues: 48% of Black adults think police definitely would use facial recognition technology to monitor Black and Hispanic neighborhoods much more often than other neighborhoods, compared with 37% of Hispanic adults, and about 18% White adults (18%) who say the same (Rainie et al.,2022).

According to a prior study conducted by Computer scientist Joy Buolamwini and Gebru in 2018 at the Massachusetts Institute of Technology (MIT), researchers found that the data sets used to train popular commercially available FRT mostly consisted of light-skinned subjects (79.6% and 89.2%), which led to the model classifying differently based on gender and race with only a 0.8% error rate with white males but up to 34.7% error rate for dark females (Gentzel, 2021). Another study by the National Institute of Standards and Technology (NIST) of 189 commercial facial recognition programs found that algorithms developed in the United States were significantly more likely to return false positives or negatives for Black, Asian, and Native American individuals compared to white individuals (Lee & Chin, 2022). It is also important to consider the consequences stemming from a mistaken arrest have the potential to affect the victim's future freedom, well-being, relationship with family members, finances, and employment status (Jones, 2021).

As AI algorithms become more critical for facial recognition technology, ensuring diversity in learning data, models and in the development teams creating it is essential to avoid "learning bias" which would skew the results of any AI model. Additionally, law enforcement and government agencies should be held responsible for their uses of FRT and the consequences stemming from those actions (Goodwin, 2021). These institutions in particular are the main players involved in regulating FRT as necessary to protect privacy and ensure accuracy. Overall, it is important to investigate the consequences of biased AI in facial recognition technology, think critically about data design and develop more ethical AI to make fairer decisions, and educate/regulate law enforcement on proper ways to use said technology to deliver fair justice.

We can use Latour's actor-network theory (ANT) to investigate this topic. Latour's argument serves as the midpoint between technological determinism and social construction of technology where even the smallest actor in a network is able to affect other actors. The actor-network theory is useful to analyze relationships between actors in a specific network as well as what happens if we add or remove certain actors (Latour, 1992). It also considers both human and nonhuman actors and the characteristics of every actor to determine how systems are built and managed (Ratnayake et al., 2017).

Parts of Latour's arguments especially valuable when considering the effects of FRT are delegation, program of action, and prescription. Delegation is when humans give work to technology. In other words, delegation to nonhuman is when an artifact takes over the manual work of humans. In this case, using FRT delegates the task of identifying suspects from the human to the nonhuman AI. Program of action is about inscribing moral values into technology. Regarding the use of FRT, it will be important to investigate ethics and accuracy, and how developers can fairly develop AI technology as well as how users can fairly use said AI technology. Prescription involves the behavior imposed back on the human by the nonhuman. While humans are the ones who develop AI, it develops us in return. When we evaluate what AI enables us to achieve, what/how we use it, and when certain AIs discriminate for/against certain values, a better understanding of this will help grow potential for humans to learn from and collaborate with algorithms in an ethical manner. Overall, ANT can also help us understand the network by analyzing differing actor perspectives and emerging effects (Cresswell et al., 2010).

We can analyze each actor's role in a network involving AI, law enforcement agents, government, as well as the accused and investigate the consequences of biased algorithms in this industry by looking at different scenarios where inaccurate assumptions lead to wrongful arrests.

We can also use ANT to help evaluate the power dynamics between different actors in this network and that analysis will help us describe responsibilities each actor has regarding fairness in criminal cases. Additionally, we can consider AI technologies as moral agents, meaning that AI can act as agents to which humans delegate different areas of interests and which act on our behalf. They would be regarded as modules that distribute information as well as manage informational relationships between a variety of actors.

Overall, with this framework in mind, we would be able to better comprehend how law enforcement agencies, FRT AI agents, and the government interact in criminal justice cases. Also, we can evaluate the extent to which facial recognition AI algorithms are helpful to law enforcement as well as when they discriminate for/against certain values and the consequences of doing so.

Research Question and Methods

The research question this paper covers is: What are the detrimental effects of biased facial recognition AI algorithms used by law enforcement in criminal investigations on people of color? This question is important to ask because FRT, which has become one of the most critical and commonly used technologies in law enforcement, poses special risks of disparate impact for historically marginalized communities. In order to answer this question, I plan to make use of interviews as well as investigate media accounts and prior research of widely known examples of wrongful arrest due to FRT inaccuracy.

One primary source of data will be interviews, and I plan to interview companies that specialize in facial recognition technology. An example is Clearview AI, a company that specializes in FRT (Klosowski, 2020). Their technology is used by many law enforcement agencies so it would be beneficial to see their opinion on this topic, why they think their technology is useful, and how they seek to improve their technology to prevent inaccuracies. I would also like to interview a big company such as Google, Microsoft, Amazon who have highly skilled engineers that have made their own FRT technology. I would ask them questions about how complicated FRT is and how difficult it might be to make good algorithms that reduce bias in order to get a better understanding of the scope of the challenges that come with creating and using FRT.

I also want to contribute to my analysis by using secondary sources such as media accounts, online reports, and prior research. Since it is difficult to get interviews with law enforcement, I plan to use these types of sources to find out more about law enforcement's uses of FRT, when and how they decide to use it, and how it has impacted their arrest processes. I will research what the process is in a typical criminal investigation, and how FRT has changed that process. I also want to find out if there are specialized training for officers to use FRT and if there is not why is that the case. While conducting research I want to make sure that I get viewpoints from both supporters and opponents of law enforcement's use of FRT to fairly evaluate both the pros and cons of FRT.

I also want to investigate examples where FRT has been inaccurate or has led to wrongful arrests due to a person having darker skin. There are three potential examples that are widely known: Robert Williams, Michael Oliver, and Nijeer Park (Johnson, 2022). Take for example, the case of Nijeer Parks, a thirty-three-year-old Black man from New Jersey, who spent ten days in jail after being falsely accused of theft and attempting to hit a police officer with his car. He was arrested in part due to an image returned as a match in the facial recognition database, and his case demonstrates a fundamental problem with police use of FRT, that repeated assessments have revealed that this technology is much less accurate in identifying people of color, and disproportionately impacts them accordingly. With such examples, I could point to how FRT discriminated against people based on the color of their skin and how that issue can and should be addressed.

Conclusion

In emergency medical care processes, the responders collect large amounts of data. It is important to process this data efficiently and accurately to provide better patient care. Additionally, high-pressure situations cause a cognitive strain on those performing during a crisis. Thus, in order to reduce strain on EMTs and improve the speed, accuracy, and precision of pre-hospital care, a cognitive AI assistant can provide beneficial support to achieve this crucial aim. This capstone project aims to build and integrate modular components into a larger cognitive assistant system that can appropriately analyze relevant data streams at the incident scene in order to infer incident context and subsequently draw conclusions regarding appropriate medical interventions based on standard EMS protocols. Next, research will be conducted to investigate the effects of biased facial recognition AI algorithms used by law enforcement in criminal investigations on people of color. In the form of interviews, media accounts, and prior research, data will be collected to analyze the different actors in this network and the responsibilities and accountability of these actors. This research aims to investigate the role of AI technology in different environments (emergency medicine and law enforcement respectively), how AI agents interact with humans in said environments, as well as the consequences of these interactions.

REFERENCES

Latour, B. (1992) 'Where are the missing masses? The sociology of a few mundane artifacts', in Bijker, W. E. and Law, J. (Eds) Shaping Technology/Building Society: Studies in Sociotechnical Change, Cambridge, MA, MIT Press, pp. 225-258.

Gentzel M. (2021). Biased Face Recognition Technology Used by Government: A Problem for Liberal Democracy. Philosophy & technology, 34(4), 1639–1663. https://doi.org/10.1007/s13347-021-00478-z

Cresswell, K.M., Worth, A. & Sheikh, A. (2010). Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. BMC Med Inform Decis Mak 10, 67. https://doi.org/10.1186/1472-6947-10-67

S. Kim, W. Guo, R. Williams, J. Stankovic and H. Alemzadeh. (2021). Information Extraction from Patient Care Reports for Intelligent Emergency Medical Services. IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 58-69, https://doi.org/10.1109/CHASE52844.2021.00014

Preum, S.M., Shu, S., Hotaki, M., Williams, R.D., Stankovic, J.A., & Alemzadeh, H. (2019). CognitiveEMS: a cognitive assistant system for emergency medical services. SIGBED Rev., 16, 51-60. https://doi.org/10.1145/3357495.3357502

Kalhori S. (2022). Towards the Application of Machine Learning in Emergency Informatics. Stud Health Technol Inform, 291, 3-16. https://doi.org/10.3233/SHTI220003

Rainie L., Funk C., Anderson M., & Tyson A. (2022). Public more likely to see facial recognition use by police as good, rather than bad for society. AI AND HUMAN

ENHANCEMENT: AMERICANS' OPENNESS IS TEMPERED BY A RANGE OF CONCERNS, 29-41. Retrieved from https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS_2022.03.17_AI-HE_REPORT.pdf

Lee N., Chin C. (2022). Police surveillance and facial recognition: Why data privacy is imperative for communities of color. Retrieved from https://www.brookings.edu/research/police-surveillance-and-facial-recognition-why-data-privacy-is-an-imperative-for-communities-of-color/

Jones C. (2021). Law Enforcement Use of Facial Recognition: Bias, Disparate Impacts to People of Color, and the Need for Federal Legislation. North Carolina Journal of Law & Technology, 22, 777-815. Retrieved from https://scholarship.law.unc.edu/ncjolt/vol22/iss4/6

Committee on Guidance for Establishing Crisis Standards of Care for Use in Disaster Situations; Institute of Medicine. Crisis Standards of Care: A Systems Framework for Catastrophic Disaster Response. Washington (DC): National Academies Press (US); 2012 Mar 21. 6, Prehospital Care Emergency Medical Services (EMS) Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK201058/

Becknell, J., Simon, L. (2016). Beyond EMS data collection: Envisioning an information-driven future for Emergency Medical Services (Report No. DOT HS 812 361). Washington, DC: National Highway Traffic Safety Administration. Retrieved from https://www.ems.gov/pdf/Beyond_EMS_Data_Collection.pdf

Lawn, S., Roberts, L., Willis, E., Couzner, L., Mohammadi, L., & Goble, E. (2020). The effects of emergency medical service work on the psychological, physical, and social well-being of ambulance personnel: a systematic review of qualitative research. BMC psychiatry, 20(1), 348. https://doi.org/10.1186/s12888-020-02752-4

Ratnayake, R., Kankanamge, N., Silva, C., Steen, J., & Czarniawska, B. (2017).How can I use actor network theory? Retrieved from https://www.researchgate.net/post/How_can_I_use_Actor_Network_Theory

Hill, K. (2020). Another arrest, and jail time, due to a bad facial recognition match. *The New York Times*. Retrieved from https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

Klosowski, T. (2020). Facial recognition is everywhere. Here's what we can do about it. Wirecutter: Reviews for the Real World; *The New York Times*. Retrieved from https://www.nytimes.com/wirecutter/blog/how-facial-recognition-works/

Johnson, K. (2022). How Wrongful Arrests Based on AI Derailed 3 Men's Lives. *WIRED*. Retrieved from https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/

Horowitz, J. (2020). Tech companies are still helping police scan your face. *CNN*. Retrieved from https://edition.cnn.com/2020/07/03/tech/facial-recognition-police/index.html

Symanovich S. (2021) What is facial recognition? How facial recognition works. *Norton*. Retrieved from https://us.norton.com/blog/iot/how-facial-recognition-software-works#

Holthe, T., Halvorsrud, L., & Lund, A. (2022). Digital Assistive Technology to Support Everyday Living in Community-Dwelling Older Adults with Mild Cognitive Impairment and Dementia. Clinical interventions in aging, 17, 519–544. https://doi.org/10.2147/CIA.S357860

Goodwin G. (2021). Facial Recognition Technology: Federal Law Enforcement Agencies Should Have Better Awareness of Systems Used By Employees. *United States Government Accountability Office GAO-21-105309*. Retrieved from https://www.gao.gov/assets/gao-21-105309.pdf