Wearable Cognitive Assistant Systems for Emergency Response

(Technical Paper)

An Analysis of Morality in Autonomous Vehicles

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Michael Tang
Spring, 2020

Technical Project Team Members
Sarah Preum
M. Arif Rahman
Homa Alemzadeh

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

Introduction

Autonomous vehicles (AVs) are on the verge of completely changing transportation in our society. With companies like Google, Uber, and Tesla making rapid progress every day on AV technologies, it seems as if it is only a matter of time until AVs become widespread. This quick development is not without good reason, there are many benefits to replacing human drivers with autonomous vehicles, one of the biggest being that a computer is constantly aware of its surroundings. AVs becoming widespread could greatly reduce the number of automotive accidents per year. However, an autonomous vehicle would require the use of some kind of computer to replace a human driver. This computer would be responsible for everything a human driver would normally be responsible for, including what happens in an accident, which means certain actions taken by the computer on the road are moral or immoral. A common example many people use when discussing autonomous vehicle morality is a trolley problem situation, where two groups of different people are in the way of an unavoidable crash. This STS research hopes to conclude that there is no technological fix to the social problem of building a moral autonomous vehicle, and that the trolley problem is inherently unfair when discussing the ethics of autonomous vehicles.

A major goal of autonomous vehicles is to reduce accidents caused by mistakes or inattention from human drivers. In the event of a significant crash, first responders are needed at the scene to assist the victims. The scene of an accident can often be noisy and chaotic, which may lead to the first responder missing information or cues from the patient, and thus making an incorrect decision for treatment. The technical project described in this proposal introduces a wearable device that can assist a first responder. This device provides the first responder with treatment suggestions based on the vital signs of the patient as well as the verbal communication

from the first responder. A wearable device has the potential to reduce the mental strain a first responder experiences during an emergency and save lives.

Technical Topic

My technical project is one that has the potential to save lives in the future. I am working with a team to design a wearable device that assists first responders in an emergency situation. First responders are trained based on a set of emergency medical service (EMS) protocols that guide them on what actions they should take in different situations. There are over a hundred of these protocols, each with a different set of actions based on the condition of the patient being treated, which can make it difficult to figure out which protocols to follow when tending to a patient. It is the responsibility of the first responder to check and talk with the patient to figure out as much information as possible, in order to choose the best EMS protocols to execute. The wearable device being designed by this technical project would provide the first responder with reminders and suggestions based on what the first responder is saying while conversing with the patient and observing his/her symptoms, as well as input data from any sensors connected to the wearable device and patient.

The entire project works as follows. In the program, EMS protocols are translated into models that the program can use. When a first responder speaks to a patient, the speech is converted to text using a speech to text library, such as Google Cloud Speech-to-Text. 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 program predicts relevant EMS protocols, and provides reminders and suggestions to the first responder. The program can then also help with filling out forms summarizing what happened during the incident.

For this project, I was brought onto the team to help with the natural language component of the wearable device. Other members of the natural language processing team include Arif Rahman and Sarah Preum, both of whom are graduate students at the University of Virginia. Part of the natural language component includes creating an effective way of detecting negations. Negation detection in a medical context is different from normal negations in English. For example, the sentence "Patient denies vomiting, chest pain, and dizziness" means that the three concepts "vomiting," "chest pain," and "dizziness" are all negated and should be detected as such. In normal English, a negation could be something like "I was not hungry." However, these non-medical negations are not relevant to a patient during an emergency, and therefore they should not be detected. This presents a big challenge: How can desirable medical negations be separated from other undesirable negations when trying to match information to relevant EMS protocols?

My solution to this problem is to use a user-extendable program called pyConTextNLP. This program uses a list of around 400 known negation regular expressions to determine whether a given sentence contains a negation statement or not. If it does, it then looks at the rest of the sentence to determine all the subjects that are related to that statement (Chapman et al., 2011). I am able to provide pyConTextNLP with a list of all the medical subjects of interest, such as "fever," "coughing," or "chest pain." The program uses this list to determine whether a negation pertains to a medical concept or not, thereby eliminating unwanted non-medical negations.

The ground-truth data that will be used to evaluate the performance of pyConTextNLP consists of annotations from EMS transcripts. These transcripts are annotated by members on the team who have formal EMS training. Each annotation created by the team includes the concept being annotated, whether it is negated or not, and what exact phrase in the transcript negates the

concept if it is negated. The program will be run on these transcripts, and its output will be compared to the ground-truth annotations to determine the program's performance.

The metric that will be used for evaluation will be average recall. When pyConTextNLP is run on a transcript, it outputs a list of negated concepts in the transcript. Recall for that transcript can be determined by dividing the number of detected concepts in the ground-truth data by the total number of ground-truth concepts. Average recall can then be calculated by averaging all recall values from each individual transcript.

It is important to show that the negation detection performance of pyConTextNLP for our use case is better than that of other natural language processing tools. These other tools that will be compared against pyConTextNLP include ones that are specifically made to be used for medical texts, such as DEEPEN and NegEx, and ones that are not, such as Stanford CoreNLP and spaCy. A custom model made for spaCy trained on biomedical texts called scispacy will also be evaluated. To compare these tools to each other, they need to output a list of detected concepts so that their average recall can be calculated, similar to how the performance of pyConTextNLP is evaluated. To accomplish this, concepts will be mapped to regular expressions dictating the forms the concepts can appear as in a sentence. The tools will then perform negation detection and output any negated substrings detected. These substrings can then be mapped to any matching concepts, which can then be used to calculate recall.

Presently, the entire program achieves around a 70% accuracy when choosing relevant protocols based on test dialogues spoken by a first responder. As this project is related to a medical context, this result is not great. It is important to minimize error throughout the pipeline of converting spoken words into concepts in the program as much as possible. When the time comes for final production of this product, errors are almost completely unacceptable, because

they can lead to an incorrect reminder or suggestion by the wearable device. However, as this project is still in a very early research state, the current performance is a step in the right direction. The research done for this project will continue until May 2020, and the technical report will be delivered in two parts: one part in December 2019, and the other in May 2019.

STS Topic

In the not-so-distant future, autonomous vehicles will be driving themselves around on our roads. As autonomous vehicles replace human drivers, they will have to take over responsibility for every action a human driver would normally be responsible for when driving: accelerating, braking, steering, and navigation, to name a few. Autonomous vehicles can constantly monitor their surroundings and thus have no problem staying alert, have a faster reaction time than humans, and can connect to a network and interface with navigation data directly while driving. However, one important responsibility is making a decision on what to do in the event of an unavoidable accident. By slamming on the brakes to avoid an impending accident, a human driver shifts the risk from the people outside the car to the passengers inside. Autonomous vehicles will have to consider this, among many other factors, when making a decision in which injury is inevitable.

This research paper hopes to discuss what morality means, as well as what it means for some action to be more moral than another, when it comes to autonomous vehicles. It also hopes to explore whether there are any social or technological solutions to build a moral self-driving car. In order to do this, there will be an exploration of current discussions about autonomous vehicle morality, such as discussions related to the trolley problem. The decision an autonomous vehicle makes in a scenario like the trolley problem could affect whether consumers decide whether or not to adopt these vehicles (Bonnefon et al., 2016). There will also be a discussion on

how the trolley problem could be inherently unfair when considering morality in autonomous vehicles, as the trolley problem should be something autonomous vehicles try avoid in the first place (Mirnig & Meschtscherjakov, 2019). Finally, autonomous vehicle ethics outside of the scope of an accident should also be considered. If autonomous vehicles become widely adopted, the decisions set by an autonomous vehicle become a general policy, in that all vehicles programmed the same way will behave the same way (Himmelreich, 2018).

In order to explore any social solutions to building a moral self-driving car, it is necessary to analyze current legal and liability frameworks, and how prepared they are for the introduction of autonomous vehicles in society. For example, in a fully autonomous vehicle that does not require any input or intervention from the people inside, who or what is considered the "driver" of the car? In addition, if autonomous vehicles are introduced into car sharing companies, who or what is liable in an accident? It is necessary to know this information when determining liability for insurance purposes, for example (Schroll, 2014).

This paper will address the ways in which this scenario exemplifies the STS theories of risk society and technological fixes. The fundamental idea of risk society is that technologies can be developed in a way that attempts to preemptively avoid any potential risks to the technology (Beck, 2000). It would be interesting to see how the AI in an autonomous vehicle would be developed based on the potential risks a self-driving car encounters on the road. In addition, there is the possibility that a technological solution to creating a moral autonomous vehicle would be a technological fix that fail to address fundamental social problems (Rosner, 2013). There are several social factors that interact with autonomous vehicles, such as legal and liability frameworks, or the ability and willingness of people to buy and adopt autonomous vehicles. Can

these factors be addressed solely with better autonomous vehicle software and design? This paper seeks to answer that question.

Research Question and Methods

The research questions that the STS thesis will cover are as follows: What are some possible social or technological solutions to create a "moral" autonomous vehicle? When we discuss autonomous vehicles, how is morality determined? In order to answer these questions, I will make use of documentary research methods and wicked problem framing. Many of the sources I plan to use in this paper have differing and conflicting views on the ethics of autonomous vehicles. For example, Bonnefon et al. (2016) conducted a study revolving around the trolley problem, where participants decided between utilitarian AVs versus AVs that protected their passengers at all costs, while Holstein and Dodig-Crnkovic (2018) reject the notion of the trolley problem altogether. Documentary research methods will help me synthesize the main themes throughout all of them, and help me form main arguments and conclusions. Wicked problem framing will show that the problem of analyzing ethics in autonomous vehicles may not have a technological solution, or a very clear social solution at the moment.

The timeline for researching these topics will extend throughout the year until April 2020. The paper will be completed by May 2020, with several drafts completed throughout the year showcasing the best arguments presented by the research at that time.

Conclusion

The deliverable for the technical topic is a natural language component that is able to successfully separate medical negations from other non-related negations. This natural language component will integrate with the rest of the project, which will be a wearable device for a first responder that can provide reminders and suggestions for the first responder when they are

interacting with a patient. By being able to successfully detect only medical negations, we will be able to determine which medical concepts apply and do not apply to a patient, and choose a correct EMS protocol as necessary. This natural language component will bring this research project one step closer to becoming a fully functioning wearable that first responders can use in the field.

The STS deliverable will be a thesis answering the questions of whether there are any technological or social solutions to creating a moral autonomous vehicle, and what morality means in terms of autonomous vehicles. Answering these questions using STS concepts and frameworks will provide a better understanding of what we can expect in the future with the current rapid development of autonomous vehicles. Both the natural language component and the STS thesis will be completed by May 2020.

References

Beck, U. (2000). Risk Society Revisited: Theory, Politics and Research Programmes. In *The Risk Society and Beyond: Critical Issues for Social Theory* (pp. 211–229). https://doi.org/10.4135/9781446219539

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Brodsky, J. S. (2016). Autonomous Vehicle Regulation: How an Uncertain Legal Landscape May Hit the Brakes on Self-Driving Cars. *Berkeley Technology Law Journal*, 851–878.

Chapman, B., Chapman, W., Dayton, G., & Mowery, D. (2011). *PyConTextNLP*. Retrieved from https://github.com/chapmanbe/pyConTextNLP

Fleetwood, J. (2017). Public Health, Ethics, and Autonomous Vehicles. *American Journal of Public Health*, *107*(4), 532–537. https://doi.org/10.2105/AJPH.2016.303628

Fleischman, W. (2010). Electronic voting systems and the Therac-25: What have we learned? *The "Backwards, Forwards, and Sideways Changes" of ICT, Proceedings of ETHICOMP*, 170–179.

Himmelreich, J. (2018). Never mind the trolley: The ethics of autonomous vehicles in mundane situations. *Ethical Theory and Moral Practice*, *21*(3), 669–684.

Holstein, T., & Dodig-Crnkovic, G. (2018). Avoiding the Intrinsic Unfairness of the Trolley Problem. *Proceedings of the International Workshop on Software Fairness*, 32–37. https://doi.org/10.1145/3194770.3194772

Hughes, T. P. (1994). Technological momentum. *Does technology drive history*, *101*.

Leveson, N. G., & Turner, C. S. (1993). An investigation of the Therac-25 accidents. *Computer*, *26*(7), 18–41. https://doi.org/10.1109/MC.1993.274940

Mirnig, A. G., & Meschtscherjakov, A. (2019). Trolled by the Trolley Problem: On What Matters for Ethical Decision Making in Automated Vehicles. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 509:1–509:10. https://doi.org/10.1145/3290605.3300739

Rosner, L. (2013). *The technological fix: How people use technology to create and solve problems*. Routledge.

Schroll, C. (2014). Splitting the Bill: Creating a National Car Insurance Fund to Pay for Accidents in Autonomous Vehicles Notes and Comments. *Northwestern University Law Review*, (3), 803–834.