Ethics of Artificial Intelligence in Autonomous Vehicles

Sindre Bergsvik Øvstegård
Oslo Metropolitan University
Electrical Eningeering
Oslo, Oslo
Email: s318337@oslomet.no

Kenneth Ramos Eikrehagen
Oslo Metropolitan University
Electrical Eningeering
Oslo, Oslo
Email: s331475@oslomet.no

***Abstract*—**

I. Introduction

The technology in automobile industry has made some remarkable progress the last century. They have managed to make safe, reliable and affordable vehicles. The last couple of decades have seen significant advancements in computation and communication technologies, and consequentially autonomous vehicles (AV) are now becoming a reality [1]. Several prototypes exists today. Among others, Volvo have started implementing autonomous trucks that deliver goods [2].

A. Autonomous vehicles

1) What is it?: A definition of autonomous vehicle (AV) can be a vehicle which senses it's environment with little or no human input or interaction This means that the vehicle is capable of gathering information about the environment around it such as objects, temperature, position, velocity etc. And execute tasks or functions according to certain directives.

2) Why?: Humans are biological lifeforms which depends and reacts on, among other things but not limited to, sleep, food, stress, isolation, social interactions and other similar processes. Besides being complex lifeforms humans react differently to these processes choices taken, in certain situations, differ greatly.

Machines today are not dependent on such stimuli and therefor they can, theoretically, continue with tasks to an infinite amount of time. At mundane and repeated tasks such as arithmetic calculations or retaining huge amounts of data, machines and computers excel with unrivalled precision and accuracy. This makes them good candidates for making pure rational and logical decisions based on optimal outcomes of certain criteria.

Based on this assessment machines would take more optimal choices over time in comparison with the average human. And their choices would be easier to determine based on simulations and testing scenarios. This would make outcomes and outputs deterministic and easy to predict and consequentially present a good basis for development and optimization.

II. Theory

A. Technology

A brief presentation on certain technologies implemented in todays land based vehicles. Both autonomous and nonautonomous.

1) Real Time Operating System: Cars today rely on numbers of sensors to acquire information about their surroundings. The velocity, GPS, proximity to obstacles, temperature etc. This sensors communicate to certain modules that is connected to a operating system (OS). This is often a specialized form of OS. A so called Real Time Operating System (RTOS) makes up the central core and processor of the car. Within, all decisions based on data form the sensors are made and functions executed. Some aspects of RTOS are presented here. [3]

a) Hardt time RTOS: A deterministic RTOS. Meaning that the output of the RTOS should happen within a certain time or deadline. Used in mission critical systems like medical devices, fight control space shuttles etcetera.

b) Soft time RTOS: Not as deterministic as Hard Time, but deadline of tasks and processes should be met within certain limits. The RTOS performance will degrade if tasks are never done within deadline and would be useless. Examples here can be audio-visual-systems for entertainment, games and general purpose OS's like Windows, Ubuntu and iOS.

2) Camera: Cameras are indispensable for providing visual information and feedback to any system, in regards to its environment. In comparison to radar, LIDAR or similar sensors, cameras are affected by weather, dust and pollution.

3) RADAR: Radio Detection and Ranging (RADAR) can complement the camera and is resistant to both pollution, weather and similar. It emits electromagnetic waves which reflects on obstacles and back to the RADAR. Distance, velocity and position (relative to the RADAR) can then be calculated with great accuracy. There are different types of RADAR and the most used is Frequency-modulated continuous Wave (FMCW). [4]

a) FMWC: Allows the detection of small objects with the use of the Doppler effect. Since the known transmitted frequency is known, stable and slow-moving objects can be filtered out of the receiving signal. This reduces interference and increases the signal to noise ratio.

4) LIDAR: Goes under many names. One of them is ”Laser Imaging and Ranging”. This method utilizes laser beams and their reflection of of objects to measure distance. Also called ranging. The lidar’s output would be point cloud data, which provides all the necessary information, to a software for detecting and determining where potential obstacles and objects are located in the environment. [5]

5) GPS: A property of the United States Government. GPS is acronym for ”Global Positioning System”. It utilizes satellites for positioning. The system is freely available for anyone with a GPS receiver and has global coverage. Several other systems from other countries exists. As long as a GPS receiver has line of sight to four or more satellites, a pinpoint of position can happen with an accuracy of two meters or less.

B. Autonomy

The definition for autonomy would be the capacity to make informed and uncoerced decisions, based on some set of information. From a more metaphoric perspective, this would be that an entity, person or organization would be given self-governing capacity. This to make decisions not forced on it by a higher ranking entity, regime, person etc.

C. Machine learning

Machine learning (ML) is a subset of artificial intelligence (AI) and by definition is ”the study of computer algorithms that improves automaticcaly through experience” [6]. Said in simpler terms it makes the machine/system learn from experience without beeing explicitly programmed, but to do so it need a lot of data. This data set is divided in to ”training data” and ”testing data”, where the ”training data” obviously is used to train the model and the ”testing data” is used to check the accuracy of the model.

ML is often devided into three categories:

1) Supervised learning:: This type of algorithms is designed to learn by example, and the idea behind is that training this type of algorithms is like having a ”supervisor” watching over the whole prosess. The training data consists of input that is paired with the correct output and during training the algorithm will search for patterns that correlates to the desired output. After training you get a model that takes new unseen input data (of the same kind) and determine which label this new inputs classifies as based on the training data. The whole essence of this kind of algorithms is to predict the correct label for newly presented inputs based on past data. On it’s simplest form it can be written as

$$Y = f(x)$$

where Y is the desired output from the mapping function f with value x [7]. Supervised learning can be devided in to two categories

a) Classification: This algorithm is typically used to put the new data into a category or a class based on it’s training data. Under training this algorithm will be given data points with an assigned category. The most common example is filtering out spam emails.

b) Regression: This algorithm is a predictive statistical process where the model tries to find relationship between dependent and independent variables in the data set. The goal of a regression algorithm is to predict a continuous number such as sales, income, and test scores.

2) Unsupervised learning:: In contrary to supervised learning that needs to train it’s algorithms to recognize patterns in dataset with known labels, unsupervised learning is used on big dataset to recognize patterns in unlabeled dataset. In simpler terms this type of algorithm learns useful patterns and properties in the data structure in big datasets with no labels. The most used algorithm used in unsupervised learing is cluster analysis [8] [9].

a) Cluster analysis: Cluster analysis, also called clustering, is used to group datasets with shared properties in order to find algorithmic relationships. Each group made is called a cluster hence the name cluster analysis or clustering, but the name ”cluster” can not be precisely defined. Since the notion cluster isn’t well defined there are many different clusering algorithms out there, but they all have one thing in common they group data [10].

3) Reinforcement learning:: Reinforcement learning (RL) agent is learning by interacting with an environment, in other words it learns from the consequences of its actions. The RL agent can do this because of a numerical reward system that encodes the sucsess of an action’s outcome. The RL agent seeks to learn how to find the actions that accumulate the most reward over time. The RL agent learns from exploring it’s environment and making decision upon how to act in every singel scenario, instead of explicitly getting tought how to respond in advance. Thus this kind of algorithms is very complicated compared to the other two other ML categories algorithms [11].

D. Artificial neural network

Artificial neural networks (NN) are interconnected sets of model neurons that simulate the function of biological neural networks. The neurons in a neural network is a mathematic function that collects and classifies information. The network is much like the statistical methods such as curve fitting and regression analysis. NN contains layers of interconnected nodes and each node is a binary linear classifier called perceptron and is simular to multiple linear regressions. [12]

E. Semantic image segmentation

Partitioning a digital image into multiple segments (sets of pixels) based on some charistics is image segmentation. Semantic image segmentation is when we assign a class to each pixel in the image [13]. This technique uses the unsupervised learning category in ML along with NN (this is also called deep learing) to partitioning the image and classifying the pixels.

III. Implementation of AI in AV’s and it’s consequences

RTOS are essential software’s used in our daily lifes. They are implemented in between the hardware abstraction layer and the user application interfaces. They handle all the tasks, semaphores and how these are prioritesed. This is why they are implemented in AV's, and every other system which needs handling of time- and safetycritical objectives.

A. *Why do we use AI in vehicles?*

We need to go 30-40 years back to find a vehicle that did not use any kind of AI, we can say that the vehicle was without a "brain". The vehicles where purely mechanic and the human behind the "steeringwheel" was the one in control, the "brain" of the vehicle. Take cars as an example, there where no anti-lock breaking system (ABS), electronical stability control (ESC), cruise control, detection for tire pressure the list goes on. The reason we use AI in vehicles is to make the vehicle safer, not only for the driver but for everybody.

As a thought experiment, AI could be implemented with the intention of detecting all possible obstacles and objects in an environment. Ranging from traffic lights and road signs, to humans, houses, animals and so on. The insentive would be the safety aspect of a traffic environment. When it comes to simple arithmetic operations, computers are far superior than humans when it comes to calculation and the precissions of these calculations. With clear directives, protocols and instructions AI, would react and take action much faster than a humanbeeing ever could. This would make AV's far more optimal for collective transport, and also a tremendous supplement when it comes to assistive functions in human operated vehicles.

If we take a look at modern cars, this kind of AI is allready in use. There is for example a city safe systems that helps the driver avoid collision in small velocities. This system uses lidar and cameras along with object detection technology to detect if there is a pedestrian or even an animal entering it's lane, or just to avoid collision in slow going queues. We are actually not so far away from beeing able to have vehicles that can be fully autonomous as explained in the thought experiment.

Allready this year (2020) 60 countries, including Norway, have reached a milestone in mobility with the adoption of a United Nations Economic Commission for Europe (UNECE) regulation that will allow for safer introduction for AV's in certain traffic situastions. In this regulation there is given strict requirements for Automated Lane Keeping Systems (ALKS). This new regulation therfore marks an important step towards the further deployment of AV's in the future, and helps realize the vision in the thought experiment for safer, more sustainable mobility for everyone. [14]

B. *Implementation*

1) Which algorithms are used?: What kind of algorithm used is dependent on computational ability and determined on what kind of technology implemented within the AV. Be it GPS, LIDAR or similar. There is no problem implementing all sorts of algorithms, but a factor that limits range of viable candidates are, the beforementioned demands and also demands in regards to safety and security.

2) Semantic image segmentation: An essential part of a modern AV's driving system, for navigating through it's surroundings and action handling, is semantic image segmentation. Their excellent performance in recognition makes this technique optimal for various AV implementations [15]. Unfortunately the the technique per now, is computationally quite costly and therefor not implemented in practice. [16]

Instead predefined features from images are used in simple driverassisted functions like recognition of road signs, road markings and traffic lights. So there is no real-time training happening in the vehicles system as it drives.

C. *Ethics*

1) Evaluation: We will briefly discuss how semantic image segmentation is implemented in AV's and some of it's effects and consequences. Certain thought experiments and future conjectures will be utilized to reveal certain ethical and safety aspects which mus be critically evaluated when considering deploying AI into dailiy life AV.

2) Decisions: How would AI decide which action to take when it comes to a life and death situation. Certain cornerstone examples could be to hit a child or an elderly person. Or the vehicle has 5 people in it and is on the collision course with one person, should it hit an kill the one person or drive off the road potentially killing every one in the vehicle? This kind of edge problem all though important to think about is in an engineering stand point unnecassery. Why not try avoid a situation like this all together? This kind of dilemmas has as the assumption that the car only have two option when in reality it would have infinite. That the vehicle has so high velocity it could'nt brake down to a speed that would'nt kill anyone is also unprobable in this scenario [17]. The easy solution to this problem is to have the vehicle have a safe velocity in areas that these scenarios are probable, and therfore a city safe system that is in modern vehicles today would take care of that problem.

We should'nt discard these kinds of edge problems all togehter, it is very important to have in mind when developing AV technology. We should'nt limit the vehicle to only two choices but rather have many decision to choose from and then look for the answer in the statistics. With this statistical outcome we as a society have to consider how we want to distribuate the fatalities between the passangers and the pedestrians.

IV. Conclusion

The end goal of implementing AV's should always be to minimize the risk and fatalities of travelig along the roads or streets in the cities. That is also why edge problem, trolley dilemmas is important to think about and should be talked about to make the society aware of the risks. But it shoud'nt be suffocating all the good that comes out of implementing AV's. City safe system that is used in vehicles today along with brake technology makes it alot safer for pedestrians to travel along the road side where the speedlimit is below 40 km/h. I am posetive that this technology will only get better and makes the future for AV's brighter. We have this year already come to level 3 AV's with the UNECE regualtion for ALKS, we still have to pay atention on the roads and be ready to interfiere if something goes wrong.

REFERENCES

[1] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun, "Towards fully autonomous driving: Systems and algorithms," 2011.

[2] P. Sawers, "Volvo's vera autonomous trucks will transport dfds goods on public roads." https://venturebeat.com/2019/06/13/volvos-vera-autonomous-trucks-will-transport-dfds-goods-on-public-roads/, 2019. Accessed 27-11-2020.

[3] Wikipedia, "Rtos." https://en.wikipedia.org/wiki/Real-time_operating_system, Dec. 2020. Accessed 18-11-2020.

[4] Wikipedia, "Radar." https://en.wikipedia.org/wiki/Radar, Dec. 2020. Accessed 18-11-2020.

[5] Wikipedia, "Lidar." https://en.wikipedia.org/wiki/Lidar, Dec. 2020. Accessed 18-11-2020.

[6] Wikipedia, "Machine learning." https://en.wikipedia.org/wiki/Machine_learning, Dec. 2020. Accessed 27-11-2020.

[7] A. Wilson, "Supervised learning." https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590, Dec. 2020. Accessed 27-11-2020.

[8] T. Wood, "Unsupervised learning." https://deepai.org/machine-learning-glossary-and-terms/unsupervised-learning, Dec. 2020. Accessed 27-11-2020.

[9] Wikipedia, "Unsupervised learning." https://en.wikipedia.org/wiki/Unsupervised_learning, Dec. 2020. Accessed 27-11-2020.

[10] Wikipedia, "Cluster analysis." https://en.wikipedia.org/wiki/Cluster_analysis, Dec. 2020. Accessed 27-11-2020.

[11] Scholarpedia, "Reinforcement learning." http://www.scholarpedia.org/article/Reinforcement_learning, Dec. 2020. Accessed 27-11-2020.

[12] Scholarpedia, "Reinforcement learning." https://www.investopedia.com/terms/n/neuralnetwork.asp, Dec. 2020. Accessed 28-11-2020.

[13] A. Kumar, "Semantic image segmentation." https://towardsdatascience.com/semantic-image-segmentation-using-fully-convolutional-networks-bf0189fa3eb8, Dec. 2020. Accessed 28-11-2020.

[14] UNECE, "Regulation on automated lane keeping systems." https://www.unece.org/info/media/presscurrent-press-h/transport/2020/un-regulation-on-automated-lane-keeping-systems-is-milestone-for-safe-introduction-of-automated-vehicles-in-traffic/doc.html, Dec. 2020. Accessed 28-11-2020.

[15] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[16] M. Bugała, "Algorithms applied in autonomous vehicle systems," *ResearchGate*, p. 21, Dec. 2018.

[17] J. Bonnefon, A. Shariff, and I. Rahwan, "The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars [point of view]," *Proceedings of the IEEE*, vol. 107, no. 3, pp. 502–504, 2019.