Responsibility - The bridge between the robot-cars & the criminal justice

Laszlo Pitlik (sen), Laszlo Pitlik (jun), Matyas Pitlik, Marcell Pitlik – MY-X research team

Abstract: Criminal justice has to explore responsibilities for specific anomalies concerning the human behaviours. Responsibility can be interpreted as a kind of causal connection between not expected impacts and human activities before – even if it is a kind of reckless endangerment committed during the line of duty by a human expert. Big-data-oriented approaches have a wider scope for the term of crime: crimes are all anomalies against a valid rule system where rule systems can be declarative or non-declarative ones. Robot-cars have a high-level complexity compared to each other robot-like constructions, therefore the robot-cars can be seen as an ideal object to explore changes concerning responsibility constellations/interpretations of affected human beings (c.f. personal responsibility vs. chained responsibility). In the rel. near future, the anomalies in the human-driven quality management should be interpreted as one of the most important responsibilities and therefore as one of the most relevant crime-sources against the whole society.

Keywords: similarity-driven analyses, norm-like behaviours, anti-discriminative models, robot-judges

Introduction

This paper tries to present an introduction into a world designed through the KNUTH's principle (https://www-cs-faculty.stanford.edu/~knuth/): "Science is what we understand well enough to explain to a computer; art is everything else." [1992] (https://www.goodreads.com/quotes/26035-science-is-what-we-understand-well-enough-to-explain-to). With other words: Knowledge/science is what can be transformed into source code, each other human activity is a kind of art where art und science can not be evaluated compared to each other. They are the two sides of the same medallion.

Within the whole picture, this paper also tries also to present an adaptation of the KNUTH's world concerning the innovation needs of the legal activities (c.f. Global Legal Hackathon – e.g. https://miau.my-x.hu/miau/quilt/reality_driven_education.docx) – especially from the point of view of the AI (artificial intelligence).

This paper has been created by authors being not legal experts, but they have a lot of successful closed "fights" against more or less massive criminal activities on diverse levels. Criminal activities can be seen as kinds of anomalies concerning valid rules systems and/or quality assurance/management expectations. So, experts of the artificial intelligence should be seen as specific legal experts trying to convert the magic of words (e.g. the legal spells) into source codes.

Classic legal objects like laws, regulations, rules, etc. are here and now classic documents written in different human languages. Parallel, there are a lot of other rule systems (e.g. regulation system of a nuclear power station) where quasi each human knowledge unit is already transformed into source codes. Robot-cars are probably the most relevant artificial objects having a specific level of complexity like judging about life and dead (c.f. moral machine: http://moralmachine.mit.edu/).

Robot-cars are participants in the traffic systems therefore traffic-anomalies can be interpreted as relevant case studies for proving responsibilities.

In the strategic level, the book of KAZOHINIA (http://www.kevius.com/kazohinia/) makes possible to understand in an intuitive way, what the ideal society might be in the future step by step – and we are already headed this ideal direction...

In the more operative level, two groups of case studies will be presented here and now about the specific interpretation possibilities of robot-cars concerning the responsibility and therefore concerning the criminal justice too.

Case studies

First of all, we have to introduce the term of the Turing-Test in order to invert it and to approximate a new view of the term of the responsibility. The first group of the cases studies will have three sublayers about turning left, parking, and navigating.

In the second case, we will see a kind of almightiness of robots (including robot-cars) as a kind of challenge approximating the final and general truth based on the principle "the ocean in a drop".

Direct and inverse Turing-Tests

The Turing-Test is a rel. simple and rel. subjective test. "*The Turing test, developed by Alan Turing in 1950, is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human*." (https://en.wikipedia.org/wiki/Turing_test).

The test of the test

In order to be able to see the necessary details of this test, we should create a test situation here and now, and we should make a decision about it.

Please, imagine – that all this text before is a product of a person, who never learned English in a classic way – it means never co-operated with human teachers. The author of this text has an other mother-tongue (not English) and this person also learned in a classic way other languages too (but never English). This person uses a kind of Robot-English what means: these sentences can be translated to an other language and then the translated texts to further languages in form of a chained translation and finally back to English so, that the original meaning will not be lost. The teacher of this person is a Robot being well-known. The teacher is namely: the Google Translator (the GoTt).

Relevant remarks: Contrary to AWS (Academic Writing Skills where the Students always need the presence of a human mentor/teacher – in case of the GoTt-based learning, the Students can have a totally sovereignty because the knowledge is in a form available what being capable of reflecting the Knuth's principle.

The AWS-methodology are never capable of delivering an objectively "good" level of competences because the knowledge (the rule system to evaluate and/or improve texts) is only present in form of written rules where negative and positive impacts of the magic of words can never be influenced/controlled. A negative impact is the possibility of arbitrary distortions of the meanings in the sent messages. A positive impact is the possibility to have a kind of misunderstanding being capable of generating innovative ideas.

The results of a GoTt-based learning process can always be evaluated in an objective way: because the first English sentence and the last English sentence should always be the same – or in case of changes/differences the similarity of the two sentences can be evaluated based on the Knuth's principle (c.f. https://miau.my-x.hu/miau/quilt/reality_driven_education.docx).

The questions are after this short story-telling as follows:

- Is it really possible to co-operate only with GoTt as English teacher based on the strategy of the chained translation leading to this level/kind of success?
- If the answer is a "YES", then:
    - Is the particular Student already a kind of cyborg?
    - Is the Robot-Teacher "good" enough compared to a Human-Teacher? (c.f. Turing Test)
    - Is the methodology of the chained translations a "good" methodology both for Robot-Teachers and Human-Teachers?
- If the answer is a "YES", then: Why not?

The inverse test

The direct Turing-Tests derive evaluations about the quality of robots (direct view). The inverse Turing-Tests derive evaluations about the quality of human actions (indirect view). The chapter "The test of the test" should be able to demonstrate what kind of subjective thinking processes are necessary to make a decision about possibilities without interactive testing of the performances.

The inverse Turing-Test is successful if a human being has the same behaviour pattern (e.g. in a traffic situation) like a robot-car. In such a case the human driver can never be have a kind of responsibility. This person might never be punished – not relevant what happened in the particular situation.

Case study about a manoeuvre of turning left

Please, image a traffic situation where a car will turn left through 2 further lanes. The basic "rule" (in the Hungarian law system) has only one word for this kind of interaction between cars. This particular word is however without any real meaning: priority (giving way). For each human brain seems to be trivial, that the priority as such can not be used for two cars with unlimited high of distance. This is a real kind of the magic of words. There is a "rule", but unlimited interpretation seems to be necessary to transform it into source-code. Robots for development of legal strategies for a process like this, say it is a kind of case-based reasoning process – even in Europe without a clear relationship to the precedence-based thinking (to the generalizability of similarities).

[Remark: The resolutions concerning uniformity are a kind of quasi rational handling strategy of similarities in Hungary/Europe.]

The AI-based thinking needs a data-driven benchmark. It needs the term: norm-like behaviour. It is not relevant here and now, which kind of mathematical approximation leads to a robot-regulation in cases like wanting to turn left. It is relevant, that the norm-like behaviour can be seen as a mathematical problem where the probability of a turning manoeuvre can follow the aggregated view of the next partial rules:

- the more is the distance between the cars the higher is the necessity to act (to turn left) – because the traffic as such can be seen as traffic if the cars are not just thinking on the spot, but they are acting/driving as far as possibly/rational
- the less the speed of the cars having seemingly priority the higher is the necessity to act (to turn left)
- the more the direction of the cars can be estimated the higher is the necessity to act (to turn left) – where the estimation means: is the road free or there are a lot of objects letting cause unexpected reactions
- the less is the number of cars having seemingly priority the higher is the necessity to act (to turn left), …

The legal documents for interpretation of the keyword of priority could also use this kind of rules but in real processes before a judge these rules should be proven and aggregated to a final resolution by the judges – in an intuitive way – although the whole system is pure physics.

A study (http://miau.my-x.hu/miau/239/kvant_behavior_patterns_v1.docx) demonstrates, the norm-like behaviour can be derived:

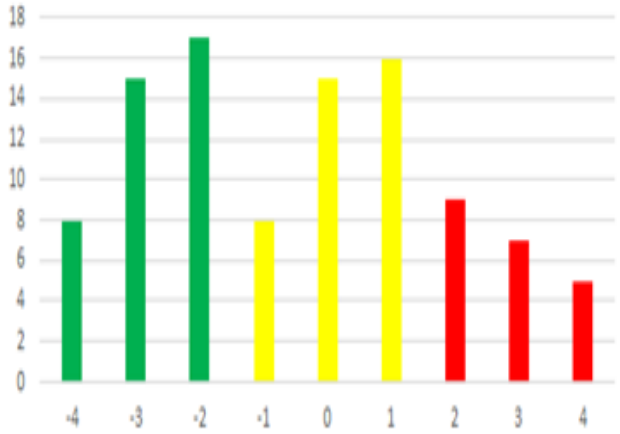

Figure-1: Frequencies of behaviour-types (source: own presentation – where X-axis = differences from norm, Y-axis = amount of cases / Legend about the behaviour types: red = endangerment of the rule of the priority / yellow = norm-like / green = drivers who act rather too slow)

In a test case where

- a driver with theoretically priority has a massive high speed
- the driver's blood contains massive volume of alcohol and/or addictive drugs
- the driver's reactions are irrational (like slaloming), …

this driver will have the whole responsibility for the crashing and/or caused deaths if the other (human-driven and/or robot-driven) car (having to give way theoretically) belongs to the green or yellow groups.

Case study about a parking manoeuvre

Parking as such can not be seen as a typical crime-oriented situation. Yet, if somebody (a human driver) seems to be a kind of racist (e.g. in case of invalid people) and this driver uses reserved parking slots frequently, then a robot-car can be involved into the legal process as a kind of agent/expert responsible for quality assurance in the traffic system. The robot-car is namely capable of deciding without any subjectivity: whether a parking manoeuvre coming from a particular direction can be seen as rational or not (for more details see the photos in the background study: https://miau.my-x.hu/miau/242/onvezeto_parkolas_v1.pdf). It might be possible, that the suspected driver was always confronted with situation where the traffic signs (protecting the reserved parking slots) are placed not correct enough. It means: coming from a given direction, the robot-car should have an other conclusion/decision as coming from an other direction concerning the reservation status of the particular parking slot.

Therefore, robot-cars can be used for checking the equilibrium of the interpretability of traffic rules. Each traffic experts in duty who nowadays can still be an arrogant "expert" confronted with the lack of the quality of the own decisions, in case of the robot-cars will become a little "mouse" in the shadow of the objectivity. This kind of a new force field let ask the question: Can be seen the role of an arrogant

traffic expert (and even the whole chain of jobs/positions for creating and evaluating traffic rules/signs, etc.) as a kind of reckless endangerment committed during the line of duty or not?

Case study about a preferred navigation route

Robot-cars can also be used for testing navigation services compared to the preferred routes enforced through traffic signs (for more details see the paper: http://miau.gau.hu/miau/205/robotizalt_kockazatelemzes.doc, https://miau.my-x.hu/miau2009/index.php3?x=e0&string=robot-driver)

Figure-2: Specific traffic situation without matured regulations (source: own presentation)

If the navigation services will not use additional human declarations (rules) just the official traffic signs, then it will be possible to explore situations where the official logic and the solution of a navigation service are not the same. Additional human rules are such kind of presumptions what can not always be derived from the traffic signs. Figure-2 demonstrates a situation where road having the upper end of the red line is a specific road: nobody knows (for 4 years), whether this road is a trivial part of the T-form-intersection (highlighted with the green curve) or it is not?

If navigation services can have information about arbitrary risk layers like amount of pedestrian crossing, schools, stops for busses and other public transport devices, etc., then the navigation can be made safety-oriented based on non-declarative rules. Non-declarative are directions of attributes of objects, where object can be the routes, and their attributes can be the following directions:

- the less is the amount of pedestrian affections the better a route is
- the less is the amount stops for public transport devices the better is a route
- the less is the affected distance concerning schools the better is a rote, …

Finally, a navigation service (a robot for route-planning) could involve each relevant information units and the best route might not be an other route as preferred by the traffic authorities – let alone: the preferred route of the robot might not be permitted by human experts having not the holistic view about the traffic system as such (c.f. Mercury Rising - https://www.imdb.com/title/tt0120749/ - about a kid with autistic performances).

Therefore, this interpretation can also be seen as a kind of reckless endangerment committed during the line of duty because the human experts are responsible for a traffic regulation ensuring the lowest level of risks originating from the traffic signs as such.

Partial conclusions

Robot-cars can be seen as a kind of robot-judges concerning traffic anomalies (partially criminal activities) based on the holistic capability of seeing systems in an optimized way contrary to human eyes with a lot of immoral motivations (like working less earning more, and/or having less changes/stress, and/or protecting corrupt force fields, and/or creating fuzzy situation for police men having a quote for punishments per day, etc.).

- The principle of the priority is operationalized.
- The equilibrium of traffic sings is operationalized.
- The optimum for routes is operationalized.

Traffic experts should search for other activities like communication with the affected persons to derive needs and wishes in a collaborative way (instead of being arrogant).

Legal experts should also search for other activities like detailed regulation of chained job being responsible for evaluation robot-technologies…

Moral Machine

The project “Moral Machine” can be interpreted in unlimited ways. Here and now, we try to demonstrate the almightiness of the robot-cars if the society will have this kind of transparency about life and death:

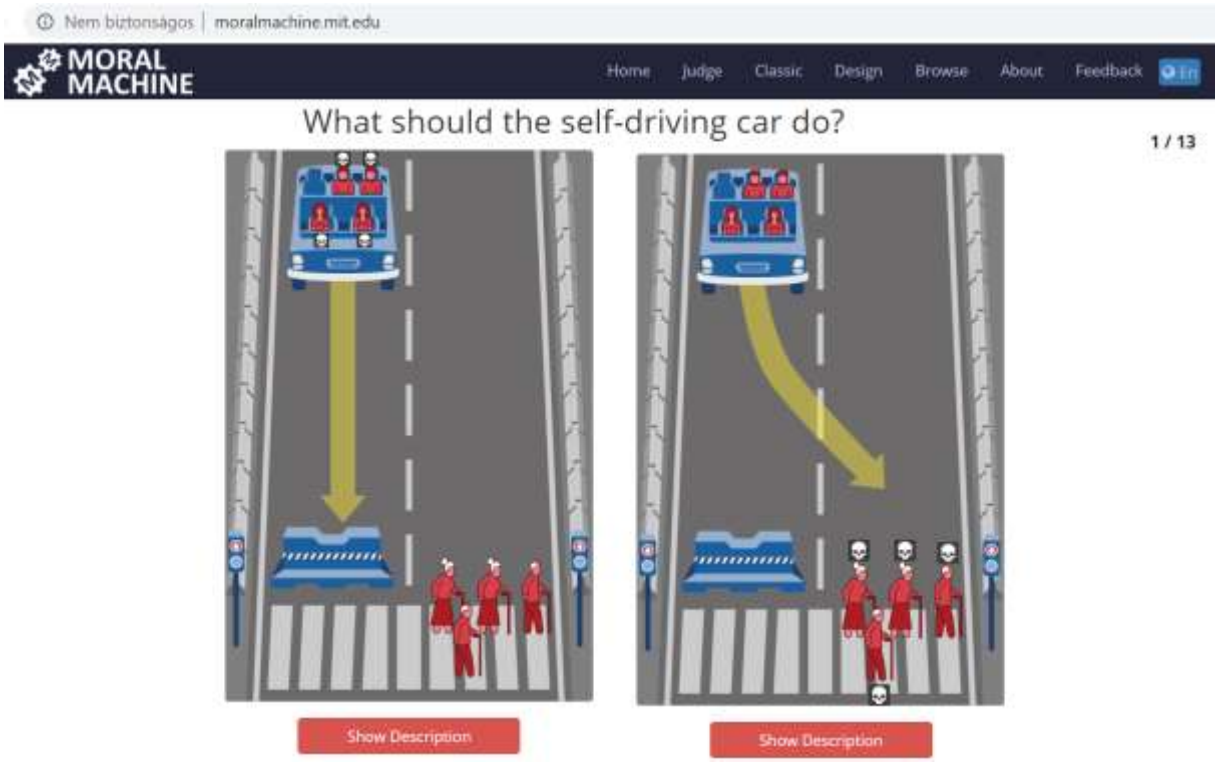

Figure-3: Standard situation of the Moral Machine Project (source: mit.edu)

The moral-machine-project (from now on: MMP) makes possible to collect human decision about a situation where a robot-car can not be stopped without any lethal impacts. The robot-car should have a collision either with a concrete object (wall) or with human individuals. In the first case, the passenger will be “killed”. In the other case, the pedestrians will be “killed”. The question is: who should be “killed”?!

The most simple solutions are:

- deciding based on randomized values or
- not letting a robot-car to have such a situation.

But, if the possibility of the above-outlined situation can not be excluded with a hundred percent probability, then the question is existing: Who should be killed? (c.f. https://en.wikipedia.org/wiki/Ender%27s_Game).

Visibility-based scenario

The MMP delivers a mixed scenario where visible and non-visible attributes about pedestrians and/or passengers are available. In this subchapter, the visible attributes will be interpreted.

[Remarks: the robot-car could have a kind of information asymmetry what means about the own passengers, a robot-car could have more information and these at once than about the pedestrians. More information can lead to more safety and/or to a better approximation of the necessity of real time decisions.]

Visible attributes could be:

- age
- sex
- outfit
- weight, etc.

The question is not the attribute as such, but the direction of an attribute: e.g.

- the younger the better?
- women <?> men
- the more serious/consolidated the better?
- the more proportional the better? etc.

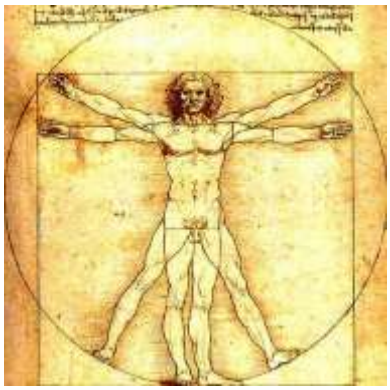

Figure-4: Ideal proportionality? (source: https://m.blog.hu/vi/videotanar/image/aranymetszes.jpg)

[Remark: Directions like before can not handle optima…]

The non-invasive measurement devices can create further attributes like temperature and its pattern concerning the whole body-surface, etc. Face readers can derive basic emotion types (angriness, happiness, sadness, etc.). Eye-tracking mechanisms can follow the targeted objects (c.f. potential sexual crimes?).

Everybody can have the prompt intuition: the single attributes alone can not be used for evaluation of human being where the previous expression “the better” means who is more relevant for the society in general and vice versa: who is the weakest chain-element?

The problem deriving aggregated index-values for evaluating objects – is already solved based on similarity analyses and based on anti-discriminative principles: e.g. each person could have the same evaluation value what means no body can be seen as the best or even as the weakest one – if the weighting (in form of a stair-case function) for each attribute-level will be optimized so that the existence of the anti-discriminative principle can be demonstrated in case of all objects.

The visible attributes could be measured, interpreted and evaluated – maybe in real-time. Therefore, the robot-car would be a kind of judge based on the martial law.

Totality-based scenario

The visible part of the attributes about a human being is a part of each measurable/available attributes. The society can be willing to regulate robots as follows:

- not identifiable objects should not be saved (where being identified means robot-cars can identify each living creatures from an arbitrary distance)
- the higher is the consistence of the data asset of a person the better (where consistence means the attributes of the particular person can be derived from each other own data based on each data unit about each person – c.f. KAZOHINIA) – this is a complex form of transparency where it is expected having data about a person in case of each thinkable attribute but this attribute should reflect a kind of “harmony” (rational ratios vs. suspicion something being covered)
- the quality of data should be as precise as possible
- the more is the lack of data the lower is the surviving level of the given person, …

The criminal activities try to cover information to realize rel. advantages and if it is possible for ever. If the robots can decide about life and death, then criminal activities can only be driven where the safety potential of criminal persons will be (e.g. even radically) decreased. The entire data assets could be cracked through criminal activities – the question is: whether the technologies like blockchains might be stronger than the purposes of criminal force fields?

Partial conclusion

The MMP leads to a kind of martial law independent of the amount and/or characteristics of the variables needed to evaluate. The responsibility to save a common data-assets is a kind of responsibility where each micro anomaly can lead to unexpected (chained) consequences. Therefore the criminal justice can be the most important guard of all these above-outlined processes.

General conclusions

One of the first reaction after reading this paper could be: Is this paper as such already a kind of crime against the canonized views? Or the legal systems and the artificial intelligence should co-operate step by step in future in order to cover new job-requirements (like LAIWYERS:-)

Further literature

gLAWgle: Hackathon-dossier (2019): https://miau.my-x.hu/miau/246/glh2019_gLAWgle/