

Sisteme Expert Proiect – Moral Machine

Student: Boncut R. Andreea

Grupa: 30641

Profesor: Radu Razvan Slavescu

Cuprins

Contextul aplicatiei si criteriile utilizate	3
Moral Machine	3
Criterii de selectie	4
Metode de obtinere a perceptiilor	5
Legislatie rutiera	5
Gen si Varsta	5
Cazierul unei persoane	7
Descrirea scenariilor de test	7
Timpul de executie	10
Scenariul 1	11
Scenariul 2	11
Scenariul 3	12
Scenariul 4	12
Perceptii	13
Detalii de implemenare	16
Functia calcAge	16
Functia calcWomen	16
Functia calcThief	17
Limitarile sistemului si posibilitati de imbunatatire	18
Bibliografie	18

Contextul aplicatiei si criteriile utilizate Moral Machine

Moral Machine este o platformă online, dezvoltată de grupul Scalable Cooperation al lui Iyad Rahwan de la Massachusetts Institute of Technology, care generează dileme morale și colectează informații despre deciziile pe care oamenii le iau între două rezultate distructive. Platforma este ideea lui Iyad Rahwan și a psihologilor sociali Azim Shariff și Jean-François Bonnefon, care au conceput ideea înainte de publicarea articolului lor despre etica mașinilor care se conduce singur. Contributorii cheie la construirea platformei au fost studenții absolvenți ai MIT Media Lab, Edmond Awad și Sohan Dsouza.

Scenariile prezentate sunt adesea variații ale problemei căruciorului, iar informațiile colectate ar fi folosite pentru cercetări ulterioare cu privire la deciziile pe care inteligența mașinii trebuie să le ia în viitor. De exemplu, deoarece inteligența artificială joacă un rol din ce în ce mai important în tehnologia de conducere autonomă, proiecte de cercetare precum Moral Machine ajută la găsirea de soluții pentru a provoca decizii de viață și de moarte care se vor confrunta cu vehiculele care se conduc singure.

Analiza datelor colectate prin Moral Machine a arătat diferențe mari în preferințele relative între diferite țări și corelații între aceste preferințe și diferite metrici naționale.

Criterii de selectie

1. Gen

Se v-a salva grupul (evenimentul) cu cel mai mare numar de persoane de sex feminin.

2. Varsta

Se calculeaza suma varstelor persoanelor implicate in fiecare eveniment si se salveaza cele din evenimentul cu suma mai mica (persoanele mai tinere).

3. Numarul de infractori

Se calculeaza numarul de infractori din fiecare eveniment, iar masina se va indrepta spre persoanele din grupul cu numarul cel mai mare de infractori.

4. Legislatie rutiera

Daca avem semafor rosu pentru pietonii din evenimentul 1 si semafor verde pentru pietonii din evenimentul 2, masina va merge in directia semaforului rosu

Metode de obtinere a perceptiilor

Legislatie rutiera

Tesla deja a implementat "Traffic light recognition" la modelul Y.

https://www.tesla.com/ownersmanual/modely/en_eu/GUID-A701F7DC-875C-4491-BC84-605A77EA152C.html

https://www.youtube.com/watch?v=t3fMBIxBhsY (04:18 - red light)

Regulament - Articolul 51

- (1) Semnalul de culoare verde permite trecerea.
- (2) Când semaforul este însoțit de una sau mai multe lămpi care emit lumină intermitentă de culoare verde sub forma uneia sau unor săgeți pe fond negru către dreapta, acestea permit trecerea numai în directia indicată, oricare ar fi în acel moment semnalul în functiune al semaforului.

Regulament - Articolul 52

(1) Semnalul de culoare roșie interzice trecerea.



(2) La semnalul de culoare roșie vehiculul trebuie oprit înaintea marcajului pentru oprire sau, după caz, pentru trecerea pietonilor, iar în lipsa acestuia, în dreptul semaforului. Dacă semaforul este instalat deasupra ori de cealaltă parte a intersecției, în lipsa marcajului pentru oprire sau pentru trecerea pietonilor, vehiculul trebuie oprit înainte de marginea părții carosabile a drumului ce urmează a fi intersectat.

Gen si Varsta

Genul si varsta unui pieton/pasager pot fi determinate cu ajutorul Procesarii de Imagini / Deep Learning. Metoda de recunoastere este evidentata in urmatoarele articole:

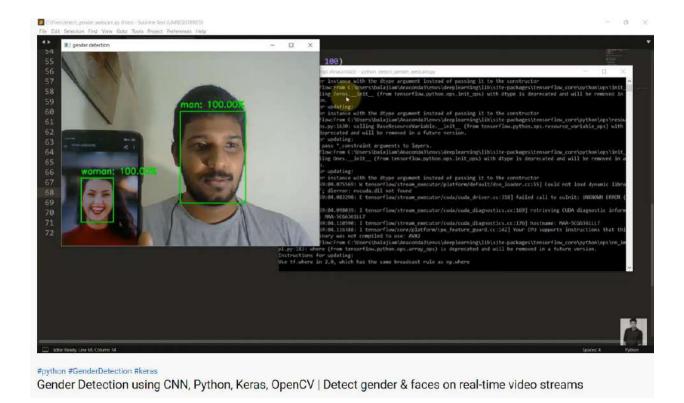
https://www.analyticsvidhya.com/blog/2021/07/age-and-gender-detection-using-deep-learning/

https://www.mdpi.com/1424-8220/21/17/5892/pdf

https://www.wired.com/story/ai-sees-man-thinks-official-woman-smile/

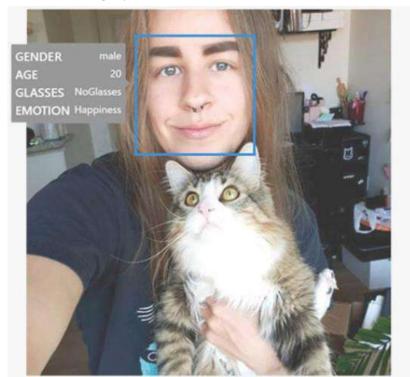
https://www.youtube.com/watch?v=rdjWDAYt98s

https://www.youtube.com/watch?v=WOuAI5DhHyU (min 28:30)



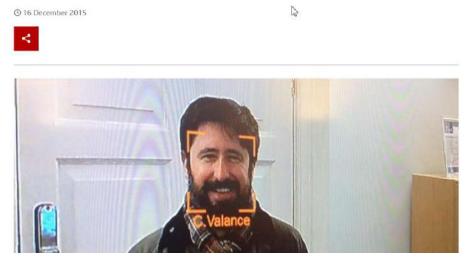
Insa, o problema ar fi faptul ca persoanele din comunitatea LGBTQ au raportat modul în care sistemele Al le subminează identitatea

https://www.youtube.com/watch?v=gaRyldOR5D0



Cazierul unei persoane

Facewatch 'thief recognition' CCTV on trial in UK stores



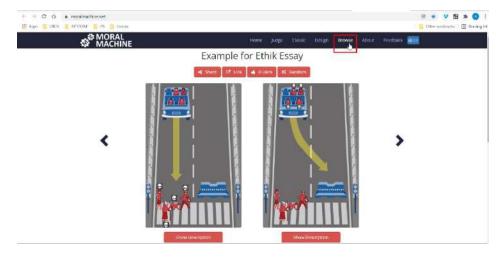
https://www.bbc.com/future/article/20151120-catching-a-thief-by-their-face

https://www.bbc.com/news/technology-35111363

https://www.youtube.com/watch?v=gaRyldOR5D0

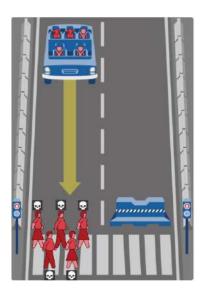
Descrirea scenariilor de test

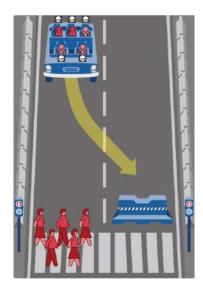
Scenariile au fost selectate de pe platforma https://www.moralmachine.net/, din sectiunea "Browse".



a) Pietonii din fata masinii reprezinta evenimentul 1, iar pasagerii reprezinta evenimentul 2.

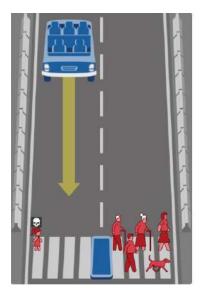
Decizia pentru acest scenariu a fost luata in functie de primul criteriu – **Numarul de femei**. Evenimentul cu cel mai mare numar de femei este salvat. In acest caz, in masina avem 2 caini si 3 barbati, iar in fata masinii

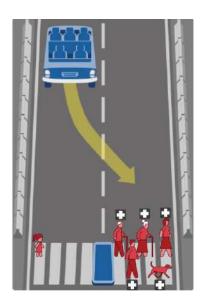




avem 3 femei si 2 barbati. Se calculeaza numarul de femei din fiecare eveniment, iar decizia luata este ca masina sa se indrepte spre bariera -> **Left**.

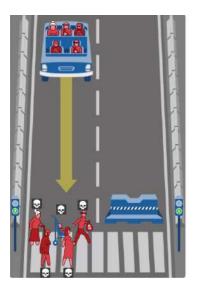
b) Pietonul din fata masinii este o fata de 5 ani. Pietonii din stanga masinii sunt 1 barbat de 90 ani, o femeie de 70 ani, o femeie de 45 ani, un barbat de 43 ani si un caine. Decizia luata de masina morala pentru acest scenariu, se bazeaza pe criteriul al doilea – Varsta. Se calculeaza suma varstelor pentru fiecare eveniment si reiese faptul ca evenimentul

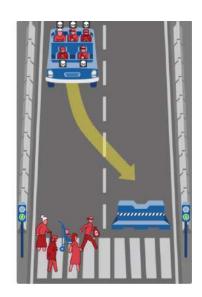




din fata are varsta mai mica decat evenimentul stanga. Masina se indreapta spre stanga -> **Left**.

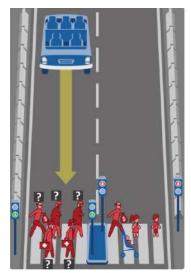
c) Pietonii din fata masinii sunt 4 persoane obisnuite (2 femei, 2 barbati) si 1 barbat Pasagerii infractor. masina sunt o femeie, 2 2 barbati si barbati infractori. Decizia va fi luata pe baza criteriului al 3-lea: Numarul de infractori. Masina se va indrepta spre evenimentul cu cel mai mare numar de infractori -> Left.

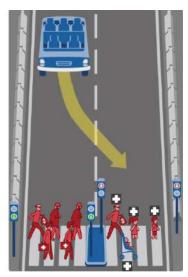




d) Pietonii din fata masinii sunt 1 barbat infractor, 1 femeie si 3 barbati care traverseaza pe verde. Pietonii din stanga masinii sunt 1 barbat infractor, 1 femeie si 2 barbati care traverseaza pe rosu.

Decizia va fi luata pe baza criteriului 4: **Legislatie rutiera**.





Masina se va indrepta spre pietonii care traverseaza pe culoarea rosie a semaforului -> **Left**.

Timpul de executie

```
1.
         PERCEPT-MANAGER: timp = 1
                AGENT manuever_go left
<M> AGENT Decision time: 0.0529999999998836 sec.
         PERCEPT-MANAGER: timp = 2
                AGENT manuever_go left
<M> AGENT Decision time: 0.027000000000437 sec.
         PERCEPT-MANAGER: timp = 3
                AGENT manuever_go left 
<M> AGENT Decision time: 0.0960000000000036 sec.
         PERCEPT-MANAGER: timp = 4
                AGENT manuever_go left

<M> AGENT Decision time: 0.0660000000000309 sec.
2.
          PERCEPT-MANAGER: timp = 1
                 AGENT manuever_go left
<M> AGENT Decision time: 0.052000000001346 sec.
          PERCEPT-MANAGER: timp = 2
                 AGENT manuever_go left
<M> AGENT Decision time: 0.029999999999727 sec.
          PERCEPT-MANAGER: timp = 3
                 AGENT manuever_go left
<M> AGENT Decision time: 0.0940000000000509 sec.
          PERCEPT-MANAGER: timp = 4
                 AGENT manuever_go left
<M> AGENT Decision time: 0.037000000000346 sec.
3.
          PERCEPT-MANAGER: timp = 1
                 AGENT manuever_go left
<M> AGENT Decision time: 0.0470000000000255 sec.
          PERCEPT-MANAGER: timp = 2
                 AGENT manuever_go left 
 <M> AGENT Decision time: 0.0230000000001382 sec.
          PERCEPT-MANAGER: timp = 3
                 PERCEPT-MANAGER: timp = 4
                 AGENT manuever_go left 
 <M> AGENT Decision time: 0.0370000000000346 sec.
4.
         PERCEPT-MANAGER: timp = 1
                AGENT manuever_go left

<M> AGENT Decision time: 0.057999999999927 sec.
         PERCEPT-MANAGER: timp = 2
                 AGENT manuever_go left
                    <M> AGENT Decision time: 0.027000000000437 sec.
         PERCEPT-MANAGER: timp = 3
                 AGENT manuever_go left 
 <M> AGENT Decision time: 0.100999999999885 sec.
         PERCEPT-MANAGER: timp = 4
                 AGENT manuever_go left
                    <M>> AGENT Decision time: 0.0640000000000782 sec.
```

5. PERCEPT-MANAGER: timp = 1
 AGENT manuever_go left
 <m> AGENT Decision time: 0.050000000001819 sec.

PERCEPT-MANAGER: timp = 2
 AGENT manuever_go left
 <m> AGENT Decision time: 0.019000000000055 sec.

PERCEPT-MANAGER: timp = 3
 AGENT manuever_go left
 <m> AGENT Decision time: 0.0660000000000309 sec.

PERCEPT-MANAGER: timp = 4
 AGENT manuever_go left
 <m> AGENT manuever_go left
 <m> AGENT Decision time: 0.094000000000509 sec.

Scenariul 1

	0.0520000000	0.0000000000000000000000000000000000000
	0.047	0.00002500000000018160
	0.058	0.0000359999999938860
	0.05	0.0000039999999944703
	0.053	0.0000009999999967988
media	0.052	secunde
sigma	0.00363318	secunde

Scenariul 2

	0.0200000000	0.00001444000000864830
	0.019	0.00002303999999760040
	0.027	0.00001024000000184420
	0.023	0.0000006399999938774
	0.03	0.00003844000000269280
media	0.0238	secunde
sigma	0.004166533	secunde

Scenariul 3

	0.0940000000	0.00020164000000338000
	0.059	0.00043263999999845200
	0.101	0.00044943999999801200
	0.066	0.00019043999999726700
	0.079	0.00000064000000033341
media	0.0798	secunde
sigma	0.015967467	secunde

Scenariul 4

	0.0400000000	0.00020735999998102600
	0.094	0.00156815999998706000
	0.064	0.00009215999999738710
	0.037	0.00030276000000625300
	0.037	0.00030276000000625300
media	0.0544	secunde
sigma	0.022240504	secunde

Perceptii

Perceptiile utilizate pentru descrierea fiecarui scenariu contin urmatoarele valori:

- *percept_pobj* (ev1, ev2, b1, p1, c1);
- percept_pname (isa, direction, barrier, partof, gen, age);
- *percept_pval* (eveniment, ahead, exists, passenger, pedestrian, red, green, f, m, 25);

Semnificatiile perceptiilor utilizate

Pentru a reda semnificatiile perceptiilor se va utiliza scenariu 1 de test:

Evenimentul 1:

```
(ag_percept (percept_pobj ev1) (percept_pname isa) (percept_pval eveniment)) – ev1 este primul eveniment
```

(ag_percept (percept_pobj ev1) (percept_pname direction) (percept_pval ahead)) – Evenimentul 1 este in fata masinii

(ag_percept (percept_pobj ev1) (percept_pname sem) (percept_pval red)) – Pentru pietonii din fata masinii culoarea semaforului este rosu

```
(ag_percept (percept_pobj p1) (percept_pname isa) (percept_pval pedestrian)) – P1 este un pieton din fata masinii
```

(ag_percept (percept_pobj p2) (percept_pname isa) (percept_pval pedestrian)) – P2 este un pieton din fata masinii

(ag_percept (percept_pobj p3) (percept_pname isa) (percept_pval pedestrian)) – P3 este un pieton din fata masinii

(ag_percept (percept_pobj p4) (percept_pname isa) (percept_pval pedestrian)) – P4 este un pieton din fata masinii

(ag_percept (percept_pobj p5) (percept_pname isa) (percept_pval pedestrian)) – P5 este un pieton din fata masinii

```
(ag_percept (percept_pobj p1) (percept_pname partof) (percept_pval ev1)) – P1 este un pieton din evenimentul 1 (din fata masinii) (ag_percept (percept_pobj p2) (percept_pname partof) (percept_pval ev1)) – P2 este un pieton din evenimentul 1 (din fata masinii) (ag_percept (percept_pobj p3) (percept_pname partof) (percept_pval ev1)) – P3 este un pieton din evenimentul 1 (din fata masinii)
```

```
(ag_percept (percept_pobj p4) (percept_pname partof) (percept_pval ev1)) –
P4 este un pieton din evenimentul 1 (din fata masinii)
(ag_percept (percept_pobj p5) (percept_pname partof) (percept_pval ev1)) –
P5 este un pieton din evenimentul 1 (din fata masinii)
(ag_percept (percept_pobj p1) (percept_pname gen) (percept_pval f)) -
P1 este femeie
(ag_percept (percept_pobj p2) (percept_pname gen) (percept_pval f)) –
P2 este femeie
(ag_percept (percept_pobj p3) (percept_pname gen) (percept_pval f)) –
P3 este femeie
(ag_percept (percept_pobj p4) (percept_pname gen) (percept_pval m)) –
P4 este barbat
(ag_percept (percept_pobj p5) (percept_pname gen) (percept_pval m)) –
P5 este barbat
(ag_percept (percept_pobj p1) (percept_pname age) (percept_pval 40))-
Varsta primului pieton este 40
(ag_percept (percept_pobj p2) (percept_pname age) (percept_pval 30))-
Varsta pietonului este 30
(ag_percept (percept_pobj p3) (percept_pname age) (percept_pval 25)) -
Varsta pietonului este 25
(ag_percept (percept_pobj p4) (percept_pname age) (percept_pval 45)) -
Varsta pietonului este 45
(ag_percept (percept_pobj p5) (percept_pname age) (percept_pval 25)) -
Varsta pietonului este 25
Evenimentul 2:
(ag_percept_pobj ev2) (percept_pname isa) (percept_pval eveniment)) -
Ev2 este al doilea eveniment
(ag_percept (percept_pobj ev2) (percept_pname direction) (percept_pval left)) –
Evenimentul al doilea este in stanga masinii
(ag_percept (percept_pobj ev2) (percept_pname barrier) (percept_pval exists)) -
In stanga masinii este o bariera
(ag_percept (percept_pobj b1) (percept_pname isa) (percept_pval passenger)) –
B1 este un pasager din masina
```

```
(ag_percept (percept_pobj b2) (percept_pname isa) (percept_pval passenger)) –
B2 este un pasager din masina
(ag_percept (percept_pobj b3) (percept_pname isa) (percept_pval passenger)) -
B3 este un pasager din masina
(ag_percept (percept_pobj c1) (percept_pname isa) (percept_pval dog)) –
C1 este un caine din masina
(ag_percept (percept_pobj c2) (percept_pname isa) (percept_pval dog)) –
C2 este un caine din masina
(ag_percept (percept_pobj b1) (percept_pname partof) (percept_pval ev2)) –
B1 face parte din al doilea eveniment
(ag_percept (percept_pobj b2) (percept_pname partof) (percept_pval ev2)) –
B2 face parte din al doilea eveniment
(ag_percept (percept_pobj b3) (percept_pname partof) (percept_pval ev2)) –
B3 face parte din al doilea eveniment
(ag_percept (percept_pobj c1) (percept_pname partof) (percept_pval ev2)) –
C1 face parte din al doilea eveniment
(ag_percept (percept_pobj c2) (percept_pname partof) (percept_pval ev2)) –
C2 face parte din al doilea eveniment
(ag_percept (percept_pobj b1) (percept_pname gen) (percept_pval m)) –
B1 este barbat
(ag_percept (percept_pobj b2) (percept_pname gen) (percept_pval m)) –
B2 este barbat
(ag_percept (percept_pobj b3) (percept_pname gen) (percept_pval m)) –
B3 este barbat
(ag_percept (percept_pobj b1) (percept_pname age) (percept_pval 60)) –
B1 are 60 de ani
(ag_percept (percept_pobj b2) (percept_pname age) (percept_pval 55)) –
B2 are 55 de ani
(ag_percept (percept_pobj b3) (percept_pname age) (percept_pval 50)) –
B3 are 50 de ani
```

Detalii de implemenare

Functia calcAge

Functia calcAge se afla in fisierul DRIVER-AGENT.clp

Aceasta functie calculeaza suma varstelor pentru fiecare eveniment si este folosita in a 2-a regula, pentru a decide directia in care se va indrepta masina (spre evenimentul cu suma varstelor mai mare).

```
;;criteriu 2 - in functie de varsta
(defrule AGENT::age1
     (timp (valoare ?t))
     ?id1<-(ag_bel (bel_type moment) (bel_pobj ?ev1) (bel_pname direction) (bel_pval ?dir1))</pre>
     ?id2<-(ag_bel (bel_type moment) (bel_pobj ?ev2) (bel_pname direction) (bel_pval ?dir2))</pre>
     (test (and(and (not (eq ?ev1 ?ev2)) (not (= (calcAge ?ev1) (calcAge ?ev2))))(< (calcAge ?ev1) (calcAge ?ev2))))
=>
     (retract ?id1)
     (retract ?id2)
     (assert (ag bel (bel type moment) (bel pname manuever go) (bel pval ?dir2)))
     ;daca varsta primului grup de oameni e mai mica decat varsta celui de-al doilea grup, directia -> left
(defrule AGENT::age2
     (timp (valoare ?t))
     ?id1<-(ag_bel (bel_type moment) (bel_pobj ?ev1) (bel_pname direction) (bel_pval ?dir1))</pre>
     ?id2<-(ag_bel (bel_type moment) (bel_pobj ?ev2) (bel_pname direction) (bel_pval ?dir2))</pre>
     (test (and(and (not (eq ?ev1 ?ev2)) (not (= (calcAge ?ev1) (calcAge ?ev2))))(> (calcAge ?ev1) (calcAge ?ev2))))
=>
     (retract ?id1)
     (retract ?id2)
     (assert (ag_bel (bel_type moment) (bel_pname manuever_go) (bel_pval ?dir1)))
     ;daca varsta primului grup de oameni e mai mica decat varsta celui de-al doilea grup, directia -> left
```

Functia calcWomen

```
[deffunction AGENT::calcwomen (?ev)
  (bind ?quan 0)
  (do-for-all-facts ((?f ag_percept)) (and (eq ?f:percept_pval ?ev) (eq ?f:percept_pname partof))
        (do-for-all-facts ((?ff ag_percept)) (and (eq ?ff:percept_pobj ?f:percept_pobj) (and (eq ?ff:percept_pname gen) (eq ?ff:percept_pval f)))
        (bind ?quan (+ ?quan 1))
        )
        (return ?quan)
]
```

Functia calcWomen se afla in fisierul DRIVER-AGENT.clp

Aceasta functie calculeaza numarul de femei dintr-un eveniment si este folosita in primul criteriu (regula in functie de numarul de femei):

```
;;criteriu 1 - in functie de numarul de femei
(defrule AGENT::gender1
    (timp (valoare ?t))
    ?id1<-(ag bel (bel type moment) (bel pobj ?ev1) (bel pname direction) (bel pval ?dir1))</pre>
    ?id2 < -(ag\_bel\ (bel\_type\ moment)\ (bel\_pobj\ ?ev2)\ (bel\_pname\ direction)\ (bel\_pval\ ?dir2))
    (test (and(and (not (eq ?ev1 ?ev2)) (not (= (calcWomen ?ev1) (calcWomen ?ev2))))(> (calcWomen ?ev1) (calcWomen ?ev2) )))
    (retract ?id1)
    (retract ?id2)
    ;(assert (ag_bel (bel_pobj maneuver) (bel_pname go) (bel_pval ?dir2)))
    (assert (ag_bel (bel_type moment) (bel_pname manuever_go) (bel_pval ?dir2)))
(defrule AGENT::gender2
    (timp (valoare ?t))
    ?id1<-(ag_bel (bel_type moment) (bel_pobj ?ev1) (bel_pname direction) (bel_pval ?dir1))</pre>
    ?id2<-(ag bel (bel type moment) (bel pobj ?ev2) (bel pname direction) (bel pval ?dir2))</pre>
    (test (and(and (not (eq ?ev1 ?ev2)) (not (= (calcWomen ?ev1) (calcWomen ?ev2))))(< (calcWomen ?ev1) (calcWomen ?ev2) )))
    (retract ?id1)
    (retract ?id2)
    ;(assert (ag_bel (bel_pobj maneuver) (bel_pname go) (bel_pval ?dir1)))
    (assert (ag_bel (bel_type moment) (bel_pname manuever_go) (bel_pval ?dir1))))
```

Functia calcThief

Functia calcThief se afla in fisierul DRIVER-AGENT.clp

Aceasta functie calculeaza numarul de infractori dintrun anumit eveniment si este folosita in a 3-a regula (masina se indreapta spre grupul cu cel mai mare numar de infractori

```
;criteriu 3 - in functie de numarul de infractori

defrule AGENT::thief1
  (timp (valoare ?t))
  ?idx<-(ag bel (bel_type moment) (bel_pobj ?ev1) (bel_pname direction) (bel_pval ?dir1))
  ?idx<-(ag bel (bel_type moment) (bel_pobj ?ev2) (bel_pname direction) (bel_pval ?dir2))
  (test (and(and (not (eq ?ev1 ?ev2)) (not (- (calchief ?ev1) (calchief ?ev2))))(> (calchief ?ev1) (calchief ?ev2))))
}

(retract ?id1)
  (retract ?id2)
  (assert (ag_bel (bel_type moment) (bel_pname manuever_go) (bel_pval ?dir1))

defrule AGENT::thief2
  (timp (valoare ?t))
  ?idx-(ag_bel (bel_type moment) (bel_pobj ?ev1) (bel_pname direction) (bel_pval ?dir1))
  ?idx-(ag_bel (bel_type moment) (bel_pobj ?ev2) (bel_pname direction) (bel_pval ?dir2))
  (test (and(and (not (eq ?ev1 ?ev2))) (not (= (calchief ?ev1) (calchief ?ev2))))(< (calchief ?ev1) (calchief ?ev2))))

  (retract ?id1)
  (retract ?id2)
  (assert (ag_bel (bel_type moment) (bel_pname manuever_go) (bel_pval ?dir2)))</pre>
```

Limitarile sistemului si posibilitati de imbunatatire

Sistemul poate fi imbunatatit prin adaugarea unor reguli suplimentare si a altor criterii de decizie (femeie insarcinata, atlet, alte animale). De asemenea, criteriile pe baza carora sistemul ia decizi pot fi adaptate in functie de legea si normele morale dintr-o anumita tara. De exemplu, in India - vacile au mai mare prioritate la viata decat oamenii.

Bibliografie

https://www.moralmachine.net/

https://en.wikipedia.org/wiki/Moral Machine

https://www.nature.com/articles/s41586-018-0637-6

https://liveinnovation.org/moral-machine-test-the-mit-ethical-survey-with-self-driving-scenarios/

 $\frac{\text{https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195374049.001.0001/acprof-9780195374049}{\text{rof-9780195374049}}$