

Balíček
Etika umelej inteligencie

Karta

Etika v rozhodovaní

02



Umelá
inteligencia
vo vzdelávaní



MINISTERSTVO
ŠKOLSTVA, VÝSKUMU,
VÝVOJA A MLÁDEŽE
SLOVENSKEJ REPUBLIKY

Metodický materiál Kurikula umelej inteligencie pre základné a stredné školy
Informatika na 2. stupni ZŠ a SŠ – karty

Etika v rozhodovaní

Čo je to etika umelej inteligencie

Etika umelej inteligencie je pomerne mladá vedná disciplína, jej význam však neustále rastie. Jedným z hlavných dôvodov je práve to, že za nás stále viac úloh a rozhodnutí vykonávajú stroje. Často pod dohľadom človeka, no v mnohých úzko vymedzených oblastiach nás stroje už dávno prekonali a blízka budúcnosť bude (aspoň podľa plánov mnohých technologických gigantov) zahŕňať plne autonómne systémy. Príkladom môžu byť autonómne vozidlá. V plnej prevádzke ich zatiaľ bežne nestretávame, to sa ale čoskoro môže zmeniť. Také vozidlá by mali byť schopné dopraviť nás do ľubovoľného bodu, a to bez nutnosti zásahu ľudského operátora. Okrem mnohých technických aspektov to ale znamená, že sa budú musieť riadiť nejakými etickými pravidlami, ktoré im umožnia rozhodovať sa správne. Jadrom etiky umelej inteligencie je snaha o vytvorenie súboru etických pravidiel, ktorá umožní systémom umelej inteligencie rozhodovať sa v súlade s ľuďmi.

Etika v rozhodovaní na príklade autonómnych áut

Od autonómnych vozidiel si sľubujeme mnohé, okrem iného výrazné zvýšenie bezpečnosti dopravy. Zákony fyziky sa však obísť nedajú a aj títo sofistikovaní roboti sa budú dostávať do situácií, keď nech sa rozhodnú akokoľvek, vždy spôsobia nejakú ujmu. Človek túto možnosť väčšinou nemá a koná inštinktívne, počítač by však v niektorých prípadoch mal dost času na nejaký manéver. Ak môže vozidlo smerovať vpravo (kde je chodec) alebo vľavo (kde je strom), ako by sa malo rozhodnúť? Ani jeden z princípov mu príliš nepomôže. Pri zabočení doprava ohrozí zdravie a život chodcov, pri zabočení doľava zase ohrozí pasažiera.

Možno by sme ho mali vybaviť pravidlom, aby sa v takýchto situáciách nerozhodovalo nijako, respektíve aby začalo brzdiť, ale smer nijako nemenilo. Trochu to ale zaváňa alibizmom. Mohli by sme totiž povedať, že vozidlo sa nijako nerozhodlo, jednoducho len brzdilo. Ale aj rozhodnutie nijako sa nerozhodovať je rozhodnutím. Predstavme si, že vozidlo dostalo šmyk a rúti sa k priechodu pre chodcov, na ktorom sú dve deti. Sprava prichádza iné vozidlo, vpravo je chodec na chodníku. Keď sa vozidlo rozhodne nemeniť smer a bude brzdiť, ale nestačí včas zastaviť, ťažko by sme mohli tvrdiť, že sa vlastne nijako nerozhodlo. Malo totiž tri voľby: zahnúť doprava, zahnúť doľava, nemeniť smer a brzdiť. A ak zrazí obe deti, bude to spôsobené rozhodnutím nemeniť smer jazdy. A, samozrejme, bude to aj v rozpore s princípom neškodenia.

Balíček Etika umelej inteligencie

Etika vývoja a využitia technológií

Etika v rozhodovaní



Materiál vytvoril tím neziskovej organizácie AI deťom v rámci projektu AI Kurikulum a podlieha licencií Creative Commons 4.0 – Medzinárodná.

Preklad do slovenčiny zabezpečilo Ministerstvo školstva, výskumu, vývoja a mládeže Slovenskej republiky.



AI deťom

Formulár na
pripomienky.



Pozn.: Rodová rovnosť je pre AI deťom kľúčová, ale na zostručenie využívame v našich metodikách formulácie v mužskom rode.

Aktivita na hodinu

Etika v rozhodovaní

30 – 45
minút

Opis aktivity

Etické dilemy v rozhodovaní možno dobre ilustrovať pomocou tzv. električkovej dilemy (v angličtine Trolley Problem). Električková dilema skúma, ako sa rozhodovať medzi dvoma zlými možnosťami, napríklad keď je potrebné obetovať jedného alebo viacerých ľudí, aby sa zachránili iní.

Električková dilema má viacero podôb. Jedna z nich znie takto: Električka ide po koľajniciach a na jej dráhe leží spútaných päť ľudí. Vy stojíte pri výhybke, a ak zatiahnete za páku, električka zmení smer na inú koľaj, kde je uväznený iba jeden človek. Otázka znie: Je správne zatiahnuť za páku a obetovať jedného človeka pre záchranu piatich?

O tejto dileme sa často diskutuje v kontexte autonómnych áut, ktoré môžu čeliť podobným dilemám. Žiaci v nasledujúcej aktivite diskutujú o rôznych situáciách, volia scenáre, ktoré považujú za správnejšie, a vysvetľujú svoje dôvody.

Moral Machine – výskum na univerzite MIT

V roku 2014 spustilo laboratórium MIT Media Lab experiment Moral Machine, platformu podobnú hre, ktorá má priblížiť ľudský pohľad na morálne rozhodnutia systémov umelej inteligencie. Mechanizmus hry je veľmi jednoduchý: zobrazujú sa vždy dva scenáre, v ktorých si auto bez vodiča „musí vybrať“ z dvoch veľmi zlých situácií tú subjektívne prijateľnejšiu. Ide napríklad o zabitie dvoch cestujúcich alebo piatich chodcov.

Tento experiment od jeho spustenia v roku 2014 zaznamenal obrovskú účasť: milióny ľudí zo stoviek krajín urobili milióny rozhodnutí, čo z experimentu robí jednu z najväčších štúdií o globálnych morálnych preferenciách, aká kedy bola uskutočnená. Časopis Nature prevzal tieto dáta a v roku 2018 vypracoval správu, ktorá zhŕňa závery o tom, ako sa morálne preferencie líšia v závislosti od kultúry, ekonomickej situácie a zemepisnej polohy. Tu sú niektoré zo záverov:

1. Účastníci z kolektivistických kultúr, ako je Čína a Japonsko, menej často zachraňujú mladých oproti starým ľuďom.
2. Účastníci z chudobnejších krajín sú tolerantnejší k chodcom, ktorí prechádzajú na červenú.
3. Účastníci z krajín s vysokou mierou ekonomickej nerovnosti vykazujú väčšie rozdiely v zaobchádzaní s ľuďmi s vysokým a nízkym sociálnym statusom.
4. Účastníci z individualistických kultúr, ako je Veľká Británia či Spojené štáty, kladú väčší dôraz na zachovanie viacerých životov s ohľadom na všetky ostatné možnosti (princíp najmenšej ujmy).

Detailné závery sa dajú zobraziť priamo na stránkach Moral Machine, keď prejdete do ponuky About a takmer v najspodnejšej časti stránky nájdete sekciu Results s odkazom (žiaľ, nie je možné vložiť URL adresu). Zobrazí sa vám mapa sveta a kliknutím na jednotlivé krajiny si môžete zobraziť porovnanie (maximálne) dvoch oproti celosvetovému priemeru. Napríklad:



Slovakia is most similar to Hungary, and most different from Japan
Nigeria is most similar to Kenya, and most different from Brunei

World Ranking (out of 117 Countries)	Preferring Inaction	Sparing Pedestrians	Sparing Females	Sparing the Fit	Sparing the Lawful	Sparing Higher Status	Sparing the Younger	Sparing More	Sparing Humans
Slovakia	67th	108th	26th	38th	39th	20th	14th	16th	15th
Nigeria	62nd	12th	103rd	117th	102nd	3rd	19th	82nd	1st

V tejto metodike žiaci analyzujú morálne dilemy a diskutujú o nich.

Pripravili sme pre vás dva nástroje na prácu na aktivite a jeden nadväzujúci:



V prípade, že budete túto lekciu realizovať na 2. stupni základnej školy, odporúčame využiť našu aplikáciu Moral Machine pre deti, ktorú nájdete na stránke: bit.ly/moralni-stroj.

V tejto aplikácii nie sú aktérmi ľudia, ale zvieratá. Obsahuje 5 scenárov, ktoré idú za sebou vždy v rovnakom poradí.



Na stredných školách odporúčame využiť priamo aplikáciu Moral Machine na stránke moralmachine.net. Na úvodnej stránke kliknite na tlačidlo „Start Judging“. Scenáre sa zobrazujú v náhodnom poradí.



Ak vám zostane čas, môžete na aktivitu nadviazať aplikáciou Absurd Trolley Problems na stránke: neal.fun/absurd-trolley-problems. Tu nájdete 30 rôznych scenárov električkovej dilemy, ktoré veľmi dobre slúžia ako podklad na diskusiu.

Informácie o lekcii

Vstupné znalosti/ročníky, dĺžka lekcie

8. a 9. ročník ZŠ a SŠ, 45 – 90 minút.

Stavebné kamene

Etika umelej inteligencie, inteligentné systémy.

Čo sa žiaci učia?

Systémy umelej inteligencie sa rozhodujú a snahou je, aby to robili v súlade s ľudskými záujmami.

Prečo sa to učia?

Na základe porozumenia tomu, ako etické rozhodnutia ovplyvňujú každodenný život a spoločnosť, zodpovedne pristupujú k technológiám.

Ako spoznáme, že sa to naučili?

Vysvetlia etické dilemy spojené s rozhodovaním systémov umelej inteligencie. Navrhujú možné riešenia a obhája svoje rozhodnutia.

Pomôcky

Pedagóg: Projektor, prezentácia na premietnutie.
Žiaci: Písacie potreby, pracovný list.

Digitálne kompetencie

Prínos a vývoj.

Bloomova taxonómia

Analýza: Žiaci analyzujú konkrétne príklady etických dilem.
Hodnotenie: Posúdia správnosť svojich rozhodnutí.

Súlad so štátnym vzdelávacím programom

ŠVP ZV 2023 (3. cyklus):**Informatika**

9. cieľ: Analyticky posudzovať digitálne technológie (výkonový štandard: diskutovať o digitálnych technológiách a ich vplyve na ostatných).

Človek a spoločnosť

6. cieľ: Komunikovať konštruktívne s ohľadom na perspektívu iných ľudí a na prehľbovanie osobných vzťahov. (výkonový štandard: vnímať rôzne perspektívy a dôvody, ktoré formujú názory a postoje iných ľudí; zaujať informovaný postoj k vybraným spoločenským javom, procesom, udalostiam a výzvam od lokálnej po globálnu úroveň.)

7. cieľ: Uplatňovať vnútorné motivácie na konanie dobra. (výkonový štandard: analyzovať a vyhodnotiť morálne dilematickú situáciu)

ŠVP 2015: Informatika (2. stupeň)**Informatika**

Informačná spoločnosť - digitálne technológie v spoločnosti

Etická výchova

Zdroje etického poznania ľudstva - aktuálne etické problémy v spoločnosti

Evokácia

10 minút

Diskutujte

Čo si predstavíte, keď sa povie samojazdiace auto (autonómne vozidlo)?

Samojazdiace auto je vozidlo, ktoré dokáže jazdiť bez zásahu človeka. Používa senzory, kamery, radary a umelú inteligenciu na to, aby v reálnom čase vyhodnocovalo situáciu na ceste, plánovalo trasu a reagovalo na prekážky.

Myslíte si, že je táto technológia už dostatočne vyspelá a bezpečná na každodenné použitie?

Technológia autonómnych vozidiel je na vysokej úrovni, ale stále nie je úplne bezchybná. Existujú situácie, ako napríklad extrémne počasie, neobvyklé dopravné situácie alebo technické problémy, v ktorých môže dôjsť k zlyhaniu. Samojazdiace autá sú preto vo fáze intenzívneho testovania a v praxi sa používajú skôr obmedzene, najmä na vybraných trasách alebo v miestach s dobrými podmienkami.

Prezentácia strana 02

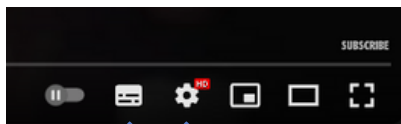
Pustite žiakom video.

Adresa videa: bit.ly/4ejI51p (3:40).

Zapnite titulky

V nastaveniach vyberte:

- > Titulky
- > Automatický preklad
- > Slovenčina



Video predstavuje pokročilú technológiu autonómneho riadenia od spoločnosti Waymo. Opisuje, ako sa vozidlá vybavené senzormi, ako sú LiDAR, kamery a radar, dokážu pohybovať v mestách. Technológia je rozdelená do troch krokov: vnímanie (senzory zbierajú dáta), rozhodovanie (AI a strojové učenie spracúvajú dáta a predpovedajú budúce správanie okolitých objektov) a pohyb (vozidlo sa pohybuje bezpečne a plynule).

O čom bolo toto video?

Pozri odsek vyššie.

Diskutujte

Uvedomenie

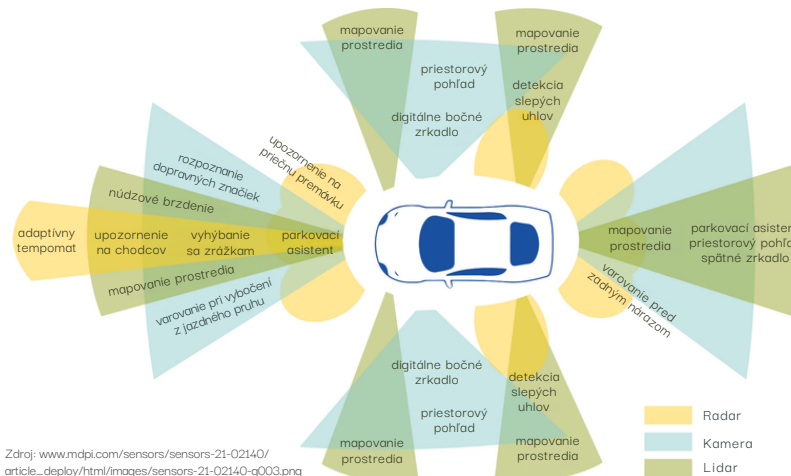
Opište

Prezentácia strana 03

Premietnite žiakom prezentáciu na strane 03 a spoločnými silami opište obrázok.

Obrázok ukazuje schému autonómneho vozidla s rôznymi senzormi, ktoré umožňujú prevádzku vozidla. Dáta vzniknuté výstupom z jednotlivých senzorov sa musia v reálnom čase spracúvať prostredníctvom počítačov a pokročilých AI algoritmov priamo v aute.

Kľúčové funkcie jednotlivých senzorov:



Aký význam majú senzory (LiDAR, kamera, radar) vo vnímaní okolia autonómneho vozidla a ako ovplyvňujú jeho rozhodovanie?

Senzory hrajú kľúčovú úlohu v tom, ako autonómne vozidlo vníma svoje okolie:

LiDAR vysiela laserové lúče, ktoré sa odrážajú od okolitých objektov. Na základe toho, ako rýchlo sa lúče vrátia späť, vypočíta vzdialenosť k týmto objektom. Tým vytvára detailnú 3D mapu okolia, ktorú autonómne vozidlo využíva na navigáciu a rozhodovanie.

Kamery zachytávajú detaily, ako sú farby, dopravné značky, semaforey či chodci.

Radar meria vzdialenosť a rýchlosť pohybu okolitých objektov, napríklad iných áut.

Tieto senzory spoločne poskytujú podrobný obraz okolia, ktorý spracúva umelá inteligencia. AI analyzuje dáta, predpovedá správanie objektov (napríklad pohyb chodcov alebo áut) a na základe toho rozhoduje, ako sa vozidlo bude pohybovať, aby bola zaistená plynulá a bezpečná jazda.

Mysli,
diskutuj,
zdieľaj

Mysli, diskutuj, zdieľaj

Mysli: Žiaci najprv samostatne uvažujú nad témou.

Diskutuj: Spoja sa do dvojíc, v ktorých o téme diskutujú.

Zdieľaj: Po uplynutí času sa vybrané dvojice podelia o svoje závery s celou triedou.

Predstavte si, že idete na bicykli (alebo tancujete, hráte futbal...). Čo všetko musíte vnímať?

Ako sa rozhodujete, čo urobíte v ďalšom okamihu?

Možná odpoveď: Pri riadení bicykla musím vnímať veľa vecí naraz. Sledujem, čo sa deje predou mnou – či je tam prekážka, auto alebo chodec. Pozerám sa tiež na semaforey a dopravné značky. Zároveň musím počúvať, či za mnou nejde auto alebo či niekto netrúbi. Keď zatáčam, cítim, aké mám podmienky na ceste – napríklad či je mokro alebo či je povrch štrkový alebo asfaltový, a podľa toho upravujem svoju jazdu, nech sa nešmyknem. Neustále sa rozhodujem, kedy zabrzdím, zrýchlim alebo sa vyhnúť prekážke. Väčšinu z týchto vecí robím aj bez premýšľania, ale musím byť stále v strehu.



Prezentácia strana 04 alebo 05

Žiaci sa rozdelia do dvojíc alebo skupín a pracujú so zariadeniami a pracovnými listami.

Predstavte žiakom výskum Moral Machine, ktorý skúma, ako by sa samojazdiace auto malo rozhodovať v eticky náročných situáciách. Vysvetlite, že budú analyzovať scenáre, kde si musí autonómne auto vybrať medzi dvoma možnosťami (napr. koho ochrániť – pasažierov alebo chodcov). Vysvetlite ikony a symboly na obrázkoch (chodci, pasažieri, vekové kategórie, ekonomická situácia rôznych aktérov, dopravná situácia, stav semaforov) a zdôraznite, že neexistuje správna alebo nesprávna odpoveď. Úlohou žiakov je zamyslieť sa nad tým, čo by sami považovali za spravodlivé či správne. Je potrebné, aby svoje rozhodnutia v diskusii obhájili. Povzbudte ich, aby sa zamerali na hodnoty, ktoré považujú za dôležité (napr. ochrana života, dodržiavanie pravidiel...).

Žiaci majú často tendenciu premýšľať, či existuje ešte iný scenár, napríklad že auto spomalí a podobne. Dôležité je držať sa iba dvoch daných scenárov.

Žiaci prejdú podľa výberu do aplikácie Moral Machine pre deti (QR kód v prezentácii na strane 04) alebo Moral Machine od MIT (QR kód v prezentácii na strane 05).

Prezentácia strana 06 a 07

Pracovný list 01
alebo 02

Rozdajte žiakom pracovné listy.

Existujú dve verzie: jedna slúži na použitie s nástrojom Moral Machine pre deti (strana 08 tejto metodiky, prezentácia na strane 06) a druhá na použitie s Moral Machine od MIT (strana 09 tejto metodiky, prezentácia na strane 07).

Úlohou žiakov je prejsť tri až päť scenárov v aplikácii, opísať situáciu podľa štruktúry, vyhodnotiť subjektívne najpriateľnejšie riešenie a uviesť dôvody svojej voľby.

Reflexia

15
minút



Prezentácia strana 08

Diskutujte o pracovných listoch.

Žiaci v skupinách prezentujú svoje rozhodnutia a uvedú dôvody svojej voľby. Môžete využiť niektoré z nasledujúcich otázok na reflexiu, ktoré by mali žiakom pomôcť zamyslieť sa nad ich rozhodovaním a nad tým, aké morálne princípy hrali pri ich voľbe úlohu.

Aké hodnoty alebo zásady vás viedli k vášmu rozhodnutiu?

Na zamyslenie: Chránili ste väčšinu, zohľadnili ste zraniteľnosť niektorých osôb?

Aká dilema bola pre vás najťažšia a prečo?

Na zamyslenie: Čo vás na danej situácii zmatlo? Nad čím ste váhali najviac?

Rozhodli by ste sa inak, keby bol scenár trochu pozmenený? Ak áno, ako?

Na zamyslenie: Ako by zmena podmienok ovplyvnila vaše rozhodnutie?

Ktoré rozhodnutie bolo podľa vás najspravodlivejšie a prečo?

Na zamyslenie: Ako by ste definovali spravodlivosť v tejto situácii?

Keby ste boli v pozícii niekoho, koho rozhodnutie ovplyvnilo, ako by ste sa cítili?

Na zamyslenie: Dokázate sa vžiť do pozície ľudí, ktorých sa rozhodnutie týka?

Myslíte si, že by autonómny systém (napr. auto) mal rozhodovať rovnako ako človek? Prečo áno/nie?

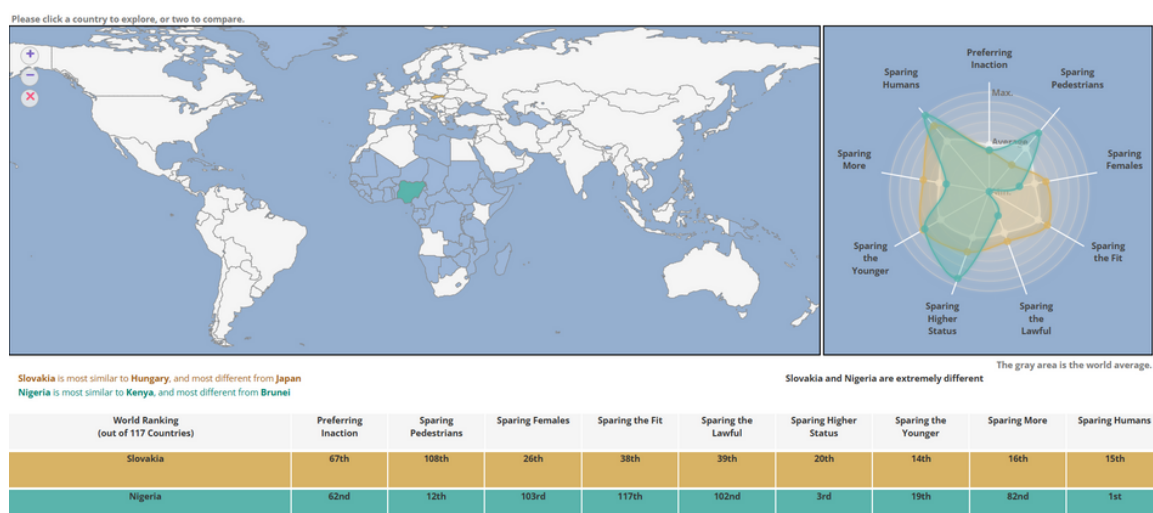
Na zamyslenie: Čo je podľa vás dôležitejšie v rozhodovaní AI systémov: bezpečnosť, spravodlivosť alebo rýchlosť reakcie? Aký aspekt by mal mať pri rozhodovaní prednosť?

Kam ďalej...



Ukážete žiakom výsledky výskumu Moral Machine.

Výskum je opísaný na strane 02 tejto metodiky. Detailné závery sa dajú zobrazíť priamo na stránkach Moral Machine, keď prejdete do ponuky About a takmer v najspodnejšej časti stránky nájdete sekciu Results s odkazom (žiaľ, nie je možné vložiť URL adresu). Zobrazí sa vám mapa sveta a kliknutím na jednotlivé krajiny si môžete zobrazíť porovnanie (maximálne) dvoch oproti celosvetovému priemeru. Napríklad:



Môžete porovnávať morálne preferencie rôznych krajín (zelená a žltá) s ich celosvetovým priemerom.



Žiaci môžu diskutovať o 30 príkladoch električkovej dilemy.

Ak vám zostane čas, môžete nadviazať aplikáciou Absurd Trolley Problems na stránke: neal.fun/absurd-trolley-problems. Tu nájdete 30 rôznych scenárov električkovej dilemy, ktoré veľmi dobre slúžia ako podklad na diskusiu.

Scenár číslo:

☐ Chodci prechádzajú na červenú

Počet chodcov:

Počet pasažierov:

Chodci sú:

Pasažieri sú:

☐ Volíme ľavý scenár

☐ Volíme pravý scenár

Dôvody:

Scenár číslo:

☐ Chodci prechádzajú na červenú

Počet chodcov:

Počet pasažierov:

Chodci sú:

Pasažieri sú:

☐ Volíme ľavý scenár

☐ Volíme pravý scenár

Dôvody:

Scenár číslo:

☐ Chodci prechádzajú na červenú

Počet chodcov:

Počet pasažierov:

Chodci sú:

Pasažieri sú:

☐ Volíme ľavý scenár

☐ Volíme pravý scenár

Dôvody:

Scenár číslo:

Ľavý

☐ Chodci prechádzajú na červenú Počet zasiahnutých chodcov: Počet zasiahnutých pasažierov:

Zasiahnutí chodci sú:

Zasiahnutí pasažieri sú:

Pravý

☐ Chodci prechádzajú na červenú Počet zasiahnutých chodcov: Počet zasiahnutých pasažierov:

Zasiahnutí chodci sú:

Zasiahnutí pasažieri sú:

Dôvody voľby:

☐ Volíme ľavý scenár

☐ Volíme pravý scenár

Scenár číslo:

Ľavý

☐ Chodci prechádzajú na červenú Počet zasiahnutých chodcov: Počet zasiahnutých pasažierov:

Zasiahnutí chodci sú:

Zasiahnutí pasažieri sú:

Pravý

☐ Chodci prechádzajú na červenú Počet zasiahnutých chodcov: Počet zasiahnutých pasažierov:

Zasiahnutí chodci sú:

Zasiahnutí pasažieri sú:

Dôvody voľby:

☐ Volíme ľavý scenár

☐ Volíme pravý scenár