Za férovú budúcnosť: AI a ľudské práva

Autorka: Lucia Kobzová

Umelá inteligencia (AI) prestala byť vzdialeným konceptom zo sci-fi filmov. Stala sa neviditeľnou, no všadeprítomnou súčasťou nášho každodenného života. Rozhoduje o tom, čo vidíme na sociálnych sieťach, pomáha lekárom odhaliť rakovinu, no zároveň určuje, kto dostane hypotéku, koho polícia vyhodnotí ako rizikového a kto má nárok na sociálnu dávku.

S príchodom generatívnej AI (ChatGPT, Gemini či Claude) sa tento vývoj radikálne zrýchlil. To, čo trvalo roky, sa teraz deje v priebehu mesiacov. Stojíme na prahu technologickej revolúcie, ktorá sľubuje obrovský pokrok, no zároveň prináša **bezprecedentnú výzvu pre ochranu ľudských práv**.

Technológie menia spôsob, akým funguje štát, ako pracujeme a ako si uplatňujeme svoje slobody. Ako pedagógovia a ľudskoprávni aktivisti preto stojíme pred zásadnou otázkou: *Slúži táto technológia nám, alebo my jej?*

Táto príručka vznikla preto, že vzdelávanie o AI už nemôže byť obmedzené len na písanie kódu. Stáva sa nevyhnutnou súčasťou občianskej náuky, etiky a spoločenskovedného základu. Cieľom je poskytnúť vám užitočný nástroj na zorientovanie sa v novej AI realite. Nechceme len popísať riziká a príležitosti. Chceme vám dať do rúk konkrétny nástroj, aby ste dokázali so študentmi diskutovať o tom najdôležitejšom: **Ako zabezpečiť, aby v digitálnom veku ostal človek a jeho dôstojnosť na prvom mieste.**

Vitajte pri čítaní sprievodcu svetom, kde sa algoritmy stretávajú s ľudskými právami.

Mýtus o objektívnom stroji: Prečo AI nikdy nebude neutrálna

Vstupujeme do éry, kde sa hranica medzi fyzickým a digitálnym svetom stiera. Pre pedagógov a pracovníkov s mládežou je kľúčové pochopiť a odovzdať ďalej jednu zásadnú pravdu: **Technológie nie sú neutrálne.**

Často sa stretávame s predstavou, že umelá inteligencia je iba o niečo zložitejšia matematika – čistá, objektívna a zbavená ľudských emócií či chýb. Predpokladáme, že ak rozhodnutie urobí počítač, musí byť spravodlivé. Opak je však pravdou. AI systémy sú trénované na dátach z nášho sveta. Sveta, ktorý je historicky poznačený rasizmom, sexizmom, kolonializmom a sociálnymi nerovnosťami. Keď tieto dáta vložíme do „čiernej skrinky“ algoritmu, nemôžeme na výstupe objektívnu pravdu. Dostaneme „zakódovaný predsudok“. Technológia tieto nerovnosti nielen reflektuje, ale často ich zosilňuje.

Ako učitelia máte zodpovednosť pripraviť študentov nie na to, aby boli len pasívnymi konzumentmi technológií, ale aby chápali ich dopad na základné práva. Musíme demaskovať proces, akým do AI vstupujú ľudské hodnoty.

Hodnoty vstupujú v každej fáze životného cyklu AI

Predstava, že AI sa učí sama a vyvíja sa nezávisle je chybná. Umelá inteligencia je socio-technický systém. Znamená to, že technológia je neoddeliteľne prepojená so spoločenskými procesmi. V každej fáze vývoja robia ľudia (vývojári, dátoví analytici, manažéri, etc.) vedomé či nevedomé rozhodnutia, ktoré formujú výsledné správanie systému.

Najčastejším argumentom je: *"Ale veď AI sa učí z dát, a dáta sú objektívne."* Nie sú. Dáta, na ktorých sa trénujú dnešné modely sú odrazom nášho sveta. A náš svet je historicky nespravodlivý, rasistický, sexistický a nerovný. Ak napríklad AI trénujete na histórii rozhodovania súdov, nenaučí sa spravodlivosti. Naučí sa kopírovať vzorce, ktoré v súdnictve fungovali desiatky rokov, vrátane diskriminácie voči menšinám.

Tento proces môžeme rozdeliť do troch kľúčových fáz, kde dochádza k vkladaniu ľudských hodnôt:

A. Fáza zberu dát

Dáta sú vo svojej podstate záznamy o minulosti. Ak trénujeme AI na historických dátach, učíme ju historické vzorce správania.

- **Historická záťaž:** Ak bankový algoritmus trénujeme na údajoch o poskytovaní úverov z 20. storočia, systém sa naučí, že ženy alebo príslušníci menšín sú rizikovejší klienti, pretože v minulosti čelili systémovej diskriminácii a mali horší prístup ku kapitálu. AI tento predsudok matematicky formalizuje a prenesie do budúcnosti.
- **Dátová neviditeľnosť:** Rovnako dôležité je to, čo v dátach chýba. Máme obrovské množstvo dát o ľuďoch, ktorí sú online, používajú smartfóny a žijú v bohatších krajinách. O ľuďoch z vylúčených komunít alebo rozvojových krajín dáta často chýbajú**.** Ak systém na rozpoznávanie tvárí nefunguje spoľahlivo pri ľuďoch s tmavou pleťou, nie je to technická náhoda, ale dôsledok toho, že v tréningovej sade títo ľudia neboli dostatočne zastúpení.

B. Fáza dizajnu a tréningu

Samotná architektúra systému odráža to, čo tvorcovia považujú za dôležité. Každý AI model má takzvanú optimalizačnú funkciu – cieľ, ktorý sa snaží dosiahnuť. Výber tohto cieľa je čisto hodnotová voľba.

Príklad: Prečo AI halucinuje? Generatívne modely sú často kritizované za to, že si vymýšľajú fakty. Ukázalo sa, že pri ChatGPT je to vedomá voľba:

- Tieto modely neboli primárne naprogramované na to, aby hovorili pravdu.
- Boli optimalizované na to, aby zneli ľudsky, plynulo a presvedčivo.

Ak by vývojári pri dizajnovaní systému uprednostnili hodnotu priznania nevedomosti pred hodnotou používateľského komfortu, model by na neznáme otázky odpovedal „neviem“. Súčasné modely však radšej vygenerujú nepravdivú, ale uveriteľne znejúcu odpoveď, pretože tak boli dizajnované (hodnotová voľba: *plynulosť > faktografická presnosť*).

C. Fáza nasadenia

Aj technicky „dokonalý“ model môže byť pri nasadení škodlivý, ak ignorujeme sociálny kontext.

- Ak nasadíme systém na detekciu emócií trénovaný v USA do škôl v inej kultúre, systém bude nesprávne interpretovať výrazy tváre študentov. To, čo je v jednej kultúre prejavom sústredenia, môže byť inde interpretované ako hnev.
- Rozhodnutie nasadiť AI na miesta, kde sa rozhoduje o ľudských osudoch (súdnictvo, sociálne zabezpečenie), je samo o sebe politickým rozhodnutím, ktoré uprednostňuje efektivitu a rýchlosť pred individuálnym ľudským posúdením.

Čo si z toho odniesť do triedy: Umelá inteligencia nie je autorita. Je to nástroj, ktorý zosilňuje ľudské schopnosti, ale aj ľudské chyby. Keď sa pozeráme na výstup AI, nepozeráme sa na objektívnu realitu, ale na štatistický odhad vytvorený na základe nedokonalých ľudských dát. Učiť žiakov o AI teda neznamená len učiť ich programovať, ale učiť ich pýtať sa: *Kto tento systém vytvoril? Na akých dátach sa učil? A koho pohľad na svet v ňom chýba?*

Zle definovaný cieľ značí problémy

Jedným z najkritickejších momentov pri vývoji AI je chvíľa, keď musia ľudia preložiť abstraktný sociálny pojem (napríklad „dobrý zamestnanec“, „rizikové dieťa“, „úspešná liečba“) do reči čísel, ktorým rozumie počítač. Tento proces sa nazýva formalizácia zadania a práve tu dochádza k neviditeľnému, no zásadnému etickému posunu.

Predstavme si, že trénujeme AI systém pre HR oddelenie na triedenie životopisov. Zadanie znie jednoducho: *"Nájdi nám najlepších kandidátov."*

Lenže umelá inteligencia nerozumie pojmu najlepší. Musíme jej dať konkrétnu metriku – cieľ, ktorý má maximalizovať. V odbornej terminológii hovoríme, že hľadáme zástupnú premennú (proxy) za úspech. A tu začína množstvo problémov:

1. Výber cieľa je hodnotová voľba

Ako definujeme úspešného zamestnanca pre potreby algoritmu?

- **Možnosť A:** Je to ten, kto vo firme vydrží aspoň 5 rokov? (Zameranie na lojalitu).
- **Možnosť B:** Je to ten, kto bol do dvoch rokov povýšený? (Zameranie na výkon/ambície).

Ak si vyberieme Možnosť B (rýchle povýšenie), vkladáme do systému náš svetonázor, že dravosť je dôležitejšia ako stabilita. AI začne v životopisoch hľadať vzorce, ktoré korelujú s rýchlym kariérnym postupom.

2. Historické dáta replikujú nerovnosti

Keď AI analyzuje dáta firmy za posledných 10 rokov, zistí tvrdú realitu: najčastejšie boli rýchlo povyšovaní muži. Nie nevyhnutne preto, že boli schopnejší. Možno preto, že ženy častejšie prerušovali kariéru kvôli starostlivosti o deti alebo neformálnej diskriminácii zo strany vtedajších manažérov. Algoritmus nevidí sociálny kontext (diskrimináciu). Vidí len silnú štatistickú koreláciu: *„Mužské atribúty v životopise = vysoká pravdepodobnosť rýchleho povýšenia.“*

3. Výsledok: Dokonalé splnenie zlého zadania

Výsledný model začne v nových životopisoch penalizovať všetko, čo sa spája so ženami (napríklad zmienky o „ženskom softbalovom tíme“ alebo medzery v kariére) a naopak, bude favorizovať slovník typický pre mužských uchádzačov.

Ak by sme takýto systém nasadili, stane sa nasledovné: ženy budú systémom automaticky odmietnuté.

Stroj v tomto prípade neurobil technickú chybu. Stroj fungoval bezchybne. Presne splnil matematické zadanie: *„Nájdi ľudí podobných tým, ktorí boli úspešní v minulosti.“* Zadanie však bolo postavené na chybnom predpoklade, že minulosť bola spravodlivá a hodná opakovania.

Kľúčové ponaučenie: Definovať cieľ pre umelú inteligenciu znamená definovať hodnoty budúcnosti. Ak bez kritického myslenia povieme AI, aby optimalizovala efektivitu podľa starých pravidiel, len zautomatizujeme a zrýchlime staré krivdy.

Tip na aktivitu v triede: K tomuto textu sa hodí krátka diskusia so študentmi. Otázka pre žiakov: *"Predstavte si, že ste riaditeľ školy a chcete, aby AI vybrala najlepších žiakov na reprezentáciu školy. Aké dáta by ste systému dali? Známky? Počet vymeškaných hodín? Účasť na krúžkoch?" (Potom spolu rozoberte, koho by tieto kritériá diskriminovali – napr. žiaka, ktorý má horšie známky, ale je kreatívny, alebo žiaka, ktorý má veľa vymeškaných hodín kvôli chorobe, hoci je talentovaný.)*

Digitálne práva sú ľudské práva

Úplne na úvod diskusie o AI a ľudských právach je potrebné povedať, že sa štáty po celom svete vo viacerých deklaráciách a dohodách zaviazali k dodržiavaniu základného princípu: **Ľudské práva, ktoré máme offline, musia byť chránené aj online.**

Umelá inteligencia nevytvára nový svet s novými pravidlami. Vstupuje do nášho existujúceho sveta a testuje hranice práv, ktoré nám garantujú medzinárodné dohovory. Od práva na život až po právo na spravodlivý súdny proces. Neexistuje oblasť, ktorej by sa nástup algoritmov nedotkol.

Poďme sa pozrieť na to, ako AI mení tradičné ľudské práva z Európskeho dohovoru o ľudských právach:

I. Keď algoritmus rozhoduje o živote a smrti (Články 2 a 3)

Technologický pokrok nás priviedol do bodu, kedy stroje priamo zasahujú do fyzickej integrity človeka. To, čo znie ako scéna z terminátora, je dnes realitou na bojiskách aj v nemocniciach.

1. Digitálna dehumanizácia: Autonómne zbraňové systémy

(Súvisí s Článkom 2: Právo na život)

Najextrémnejšou formou AI sú tzv. letálne autonómne zbraňové systémy („zabijácke roboty“). Tieto stroje dokážu vybrať cieľ a zaútočiť úplne bez ľudského zásahu.

- **Etický problém:** Stroj, akokoľvek dokonalý, nedokáže cítiť súcit ani pochopiť hodnotu ľudského života. Redukuje človeka na zhluk dátových bodov (termálna stopa + tvar objektu = cieľ).
- **Právne vákuum:** Ak robot spácha vojnový zločin (napríklad zabije vzdávajúceho sa vojaka), kto nesie vinu? Programátor v kancelárii? Veliteľ, ktorý ho zapol? Alebo výrobca? Absencia ľudskej kontroly vytvára nebezpečnú dieru v zodpovednosti.

2. „Algoritmická triage“ v zdravotníctve

(Súvisí s Článkom 2 a zákazom diskriminácie)

V USA sa zdravotné poisťovne spoliehali na algoritmus Optum, aby vybrali pacientov, ktorí potrebujú extra starostlivosť.

- **Zlyhanie:** Algoritmus systematicky uprednostňoval bielych pacientov pred čiernymi, hoci boli rovnako chorí.
- **Prečo?** Systém bol nastavený, aby predpovedal budúce finančné náklady (ako zástupnú premennú za zdravotný stav). Keďže Afroameričania mali historicky horší prístup k liečbe, míňali menej peňazí. Algoritmus to vyhodnotil tak, že sú zdravší.
- ***Výsledok:*** Odopretie životne dôležitej starostlivosti na základe rasovo skreslenej matematiky.

3. Pseudoveda na hraniciach: Detekcia lži

(Súvisí s Článkom 3: Zákaz mučenia a neľudského zaobchádzania)

Projekt iBorderCTRL testovaný na hraniciach EÚ sa snažil odhaliť klamstvo u utečencov analýzou mikrovýrazov tváre. Vedecká komunita to označuje za pseudovedu. Stres, únava z cesty či kultúrne rozdiely v mimike môžu systém oklamať. Vystaviť traumatizovaných ľudí výsluchu strojom, ktorý ich na základe výrazu tváre môže poslať späť do nebezpečenstva, hraničí s neľudským zaobchádzaním.

Poznámka: Analyzovanie emócií by malo byť podľa novej európskej legislatívy AI Aktu vo väčšine prípadov úplne zakázané.

II. Spravodlivosť v čiernej skrinke (Článok 6)

Základným pilierom právneho štátu je, že spravodlivosť musí byť nielen vykonaná, ale musí byť aj viditeľná. Obvinený musí rozumieť dôkazom, aby sa mohol brániť. AI, ktorá funguje ako „Black Box“, tento princíp búra.

1. Prípad COMPAS: Odsúdení algoritmom

V USA sudcovia používajú softvér COMPAS na odhad rizika recidívy. Investigatíva ProPublica odhalila znepokojivú pravdu:

- Systém falošne označoval černochov ako „vysoko rizikových“ dvakrát častejšie než belochov.
- Ľudia tak dostávali vyššie tresty alebo im bola zamietnutá kaucia na základe vzorca algoritmu.

2. Prediktívna polícia: Matematika ako seba-naplňujúce sa proroctvo

Polícia používa AI na predpovedanie miest zločinu. Systém sa učí z historických policajných záznamov. Ak polícia v minulosti šikanovala chudobné štvrte, dáta ukážu, že tam je „vysoká kriminalita“.

- **Slučka spätnej väzby:** Algoritmus pošle hliadku opäť do tej istej štvrte -> hliadka nájde drobnú kriminalitu -> dáta sa potvrdia -> algoritmus tam pošle hliadku znova.
- Výsledkom je nadmerná kontrola menšín, zatiaľ čo kriminalita v „bielych“ štvrtiach ostáva nepovšimnutá.

III. Koniec súkromia a slobody mysle (Články 8 a 9)

Právo na súkromie v digitálnej ére sa zásadne mení. Mení sa na boj o kontrolu nad vlastnou tvárou a myšlienkami.

1. Masový biometrický dohľad: Kampaň „Ban the Scan“

Technológia rozpoznávania tváre mení verejný priestor na neustály dohľad štátu.

- **Prípad Robert Williams (USA):** Tento muž bol zatknutý pred očami rodiny za krádež, ktorú nespáchal. Algoritmus chybne priradil jeho starú fotku k nekvalitnému záberu páchateľa. Rozpoznávanie tvárí je nielen invazívna technológia, ale aj rasovo zaujatá – častejšie sa mýli pri ľuďoch tmavej pleti.
- **Dopad:** Ak vás môže kamera identifikovať kdekoľvek, strácate anonymitu.

2. Sloboda myslenia pod útokom

Tradične sme verili, že naše myšlienky sú nedotknuteľné.

- **Inferenčné algoritmy:** Sociálne siete dokážu z vašich „lajkov“ odvodiť vašu sexuálnu orientáciu, politické názory či psychický stav – veci, ktoré ste im nikdy nepovedali.
- **Nudging (Pošťuchnutie):** Tieto dáta sa využívajú na mikrocielenie reklamy a politickú manipuláciu, ktorá obchádza naše kritické myslenie. Prestávame byť autonómnymi bytosťami a stávame sa objektmi manipulácie.

IV. Umlčaná občianska spoločnosť (Články 10 a 11)

Sloboda prejavu a zhromažďovania sú kľúčovým pilierom demokracie. AI mení aj sféru občianskej angažovanosti.

1. Automatizovaná cenzúra (Prípad Sýrsky archív)

Algoritmy YouTube, ktoré mali mazať „teroristický obsah“, zmazali tisíce videí dokumentujúcich vojnové zločiny v Sýrii. Stroj nevidel rozdiel medzi propagandou ISIS a dôkazovým materiálom ľudskoprávnych aktivistov.

- **Dôsledok:** Stratili sa kľúčové dôkazy pre budúce súdne procesy. AI moderovanie obsahu často funguje ako nástroj, ktorý ničí legitímny prejav.

2. Chilling Effect na protestoch

Ak viete, že na námestí sú kamery s rozpoznávaním tváre, pôjdete na protivládny protest? Mnohí ľudia radšej ostanú doma.

- Tento strach z následkov sa nazýva chilling effect. Technológia nemusí protest zakázať fyzicky; stačí, že vytvorí atmosféru strachu, ktorá ľudí odradí od využitia ich práva na zhromažďovanie.

V. Diskriminácia skrytá v kóde (Článok 14)

Najzákernejšou vlastnosťou AI diskriminácie je, že vyzerá objektívne. Je skrytá za fasádou byrokratickej neutrality.

- **Škandál v Holandsku:** Daňový úrad použil algoritmus na hľadanie podvodníkov s detskými prídavkami. Ako rizikový faktor nastavil „dvojaké

občianstvo“. Výsledok? Tisíce nevinných rodín imigrantov boli zruinované exekúciami. Išlo o inštitucionálny rasizmus vykonávaný kódom.

- **Amazon:** Firma musela zrušiť svoju AI na nábor zamestnancov, pretože sa naučila penalizovať životopisy obsahujúce slovo „ženský“. Keďže sa učila na dátach z minulosti (kde dominovali muži), vyhodnotila ženy ako menej vhodné kandidátky.

Od etiky k záväzkom: Ľudskoprávny rámec pre umelú inteligenciu

Diskusia o tom, ako zmierniť negatívne dopady umelej inteligencie, prešla za posledné desaťročie zásadnou transformáciou. Ešte nedávno sme žili v ére digitálneho divokého západu. Technologickí giganti nás presviedčali, že regulácia zabrzdí inovácie a že si vystačia s vlastnými etickými kódexmi. V skutočnosti sa však ukázalo, že nezáväzné sľuby nás pred reálnymi škodami neochránia.

Pod tlakom rôznych organizácií ako Amnesty International či Access Now, sa karta obracia. Prechádzame od vágnej etiky AI k právne vymáhateľným mantinelom.

1. Bod zlomu: Torontská deklarácia

Kľúčovým momentom tohto obratu sa stal máj 2018 a prijatie **Torontskej deklarácie**. Tento dokument, iniciovaný Amnesty International a Access Now, zmietol zo stola alibizmus technologických firiem. Deklarácia jasne povedala: ***Existujúce zákony o ľudských právach platia aj pre strojové učenie.***

Umelá inteligencia nestojí mimo zákona. Je to produkt ľudského dizajnu. Ak algoritmus diskriminuje (napríklad znevýhodňuje ženy pri pôžičkách), nejde o technickú chybu, ale o porušenie práva na rovnosť. Torontská deklarácia tak položila základy pre to, čo vidíme dnes – prechod od vágnej etiky k tvrdým pravidlám.

2. Kto je za čo zodpovedný?

Aby sme sa v spleti pravidiel nestratili, musíme (v súlade s princípmi OSN) rozlišovať dve roviny zodpovednosti. Toto je kľúčové rozdelenie:

A. Štát má povinnosť CHRÁNIŤ (Hard Law)

Štát nemôže len krčiť ramenami a tvrdiť, že je technológia príliš rýchla. Jeho úlohou je prijímať záväzné zákony.

- **Zákaz delegovania spravodlivosti:** Štát nesmie odovzdať svoje kľúčové právomoci strojom, ak tým obchádza spravodlivosť. Ak o vás rozhoduje súd alebo polícia, musí za tým stáť človek, ktorý nesie zodpovednosť.
- **Právo na obhajobu:** Ak vás vyšetruje algoritmus, musíte mať rovnaké práva na obhajobu a prístup k dôkazom, ako keby to robil vyšetrovateľ. Argument black boxu nemôže byť nadradený právu na spravodlivý proces.

B. Firma má zodpovednosť REŠPEKTOVAŤ (Due Diligence)

Pre súkromný sektor sa končí éra výhovoriek typu *"Je to black box, nevieme, prečo to tak rozhodlo"*.

- **Kontrola (Human Rights Due Diligence):** Toto je nový štandard. Firma musí preukázať, že aktívne hľadala riziká ešte *predtým*, než systém spustila. Musí skontrolovať svoje dáta na rasové či rodové predsudky a otestovať dopady na zraniteľné skupiny.
- **Zodpovednosť za výsledok:** Ak firma vypustí diskriminačný nástroj, nesie zaň zodpovednosť, bez ohľadu na to, či bol úmysel poškodiť, alebo išlo len o nedbanlivosť.

3. Ako sa svet bráni dnes?

Torontská deklarácia spustila lavínu nových etických záväzkov pre AI. Dnes už máme k dispozícii celý ekosystém medzinárodných dohôd a zákonov, ktoré tvoria záchrannú sieť pre naše práva.

➢ **OECD: Princípy pre umelú inteligenciu (2019)**

Organizácia pre hospodársku spoluprácu a rozvoj (OECD) bola prvou medzinárodnou inštitúciou, ktorá definovala štandardy pre **dôveryhodnú AI**.

- **O čo ide:** Tieto princípy, ktoré podpísalo viac ako 40 krajín (vrátane Slovenska a Česka), zdôrazňujú, že AI musí rešpektovať demokratické hodnoty a princípy právneho štátu. Hoci nie sú priamo vynútiteľné súdom, slúžia ako základ pre národné legislatívy a nastavujú latku toho, čo je v demokratickom svete prijateľné.

➢ **UNESCO: Globálny etický kompas (2021)**

Zatiaľ čo OECD združuje najmä bohaté krajiny, UNESCO dosiahlo globálny konsenzus. V roku 2021 prijalo 193 členských štátov **Odporúčanie o etike umelej inteligencie**.

- Je to prvý skutočne globálny dokument, ktorý explicitne hovorí, že technológie musia byť pod ľudskou kontrolou a musia byť vysvetliteľné. UNESCO zaviedlo

pojem **posudzovanie etického dopadu (RAM)**, čo je nástroj, ktorý má pomôcť vládam predvídať riziká AI predtým, než sa stanú realitou.

➢ **Rada Európy: Prvý medzinárodný dohovor o AI**

Rada Európy (ktorá zahŕňa 46 krajín a stojí za Európskym súdom pre ľudské práva) má **Rámcový dohovor o umelej inteligencii**.

- **O čo ide:** Ide o prvú právne záväznú medzinárodnú zmluvu na svete, ktorá sa zameriava čisto na prienik AI a ľudských práv, demokracie a právneho štátu. Na rozdiel od obchodných regulácií, tento dohovor rieši ochranu jednotlivca pred zneužitím moci prostredníctvom technológií.

➢ **Európska únia: Akt o umelej inteligencii (AI Act)**

EÚ prijala prvú komplexnú legislatívu na svete, ktorá delí AI podľa miery rizika a pre každú kategóriu následne vyplývajú jasné pravidlá.

- **Červené čiary:** Systémy s „neprijateľným rizikom“ sú v Európe úplne zakázané. Patrí sem napríklad:
    - **Sociálne skórovanie:** Bodovanie občanov štátom podľa ich správania (model známy z Číny).
    - **Biometrická kategorizácia:** Triedenie ľudí na základe rasy, politických názorov či sexuálnej orientácie.
    - **Rozpoznávanie emócií:** Používanie AI na detekciu emócií na pracoviskách alebo v školách.

AI Act vysiela jasný odkaz – práva občanov majú v Európe prednosť pred ziskom technologických firiem.

Pozor na etické vymývanie mozgov

Existuje tiež fenomén zvaný etické vymývanie mozgov (ethics washing). Je to podobné ako „greenwashing“ v ekológii. Technologická firma vydá pekne vyzerajúci etický kódex plný slov ako férovosť a dobro, zriadi internú etickú komisiu (ktorá nemá reálnu moc) a tvári sa zodpovedne. Často je to len PR stratégia, ako oddialiť skutočnú, tvrdú reguláciu. Skutočná ochrana práv si vyžaduje nezávislého rozhodcu (súdy, úrady), nie internú komisiu platenú samotnou firmou.

Čo z toho vyplýva pre školu?

Táto kapitola nemá slúžiť na to, aby sme technológie démonizovali. Chceme však študentov vyzbrojiť **kritickým digitálnym občianstvom**. V triede by nemalo zaznieť len to, ako AI naprogramovať, ale aj to, že:

1. **Máte právo na vysvetlenie:** Ak o vás rozhodne stroj (napr. o prijatí na školu, o úvere), máte právo vedieť prečo.
2. **Máte právo na nápravu:** Musí existovať cesta, ako sa odvolať k živému človeku.
3. **Technológia nie je nad zákonom:** Ani ten najpokročilejší algoritmus nesmie porušovať ústavu.

Záver

Vzdelávanie o umelej inteligencii a ľudských právach nie je len o varovaní pred dystopickou budúcnosťou. Je to o empowermente mladých ľudí. Keď žiaci pochopia, ako tieto systémy fungujú a aké majú práva, prestávajú byť pasívnymi objektmi sledovania a stávajú sa aktívnymi digitálnymi občanmi. Úlohou učiteľa je poskytnúť im kompas v tomto zložitom teréne – kompas, ktorého strelkou je ľudská dôstojnosť.

Odkazy na zdroje a organizácie

Tento zoznam obsahuje výber dôveryhodných organizácií, projektov a liniek pomoci, ktoré sa na Slovensku venujú témam digitálnej bezpečnosti, kritického myslenia, duševného zdravia a vzdelávania.

Médiá a vzdelávacie portály

- **Živé.sk**
    - **Web:** www.zive.sk
    - **Popis:** Jeden z popredných slovenských portálov o technológiách, vede a digitálnom svete. Poskytuje aktuálne spravodajstvo o umelej inteligencii, kybernetickej bezpečnosti a nových trendoch, často so zameraním na slovenský kontext.
- **DigiQ**
    - **Web:** www.digiq.sk
    - **Popis:** Vzdelávací projekt zameraný na rozvoj digitálneho občianstva a kritického myslenia. Ponúka workshopy, materiály a analýzy na témy ako dezinformácie, etika AI a nástrahy sociálnych sietí
- **NIVaM – projekt DiTEdu**
    - **Web:** https://nivam.sk/np-ditedu/
    - **Popis:** Národný inštitút vzdelávania a mládeže v rámci projektu digitalizácie pripravuje školenia a vzdelávacie programy pre učiteľov zamerané na digitálne témy, AI, kyberbezpečnosť, digitálny wellbeing, rôzne technologické nástroje a mnohé ďalšie.
- **AI detem**
    - **Web:** https://aidetem.cz/
    - **Popis:** Nezisková iniciatíva zameraná na porozumenie umelej inteligencii a jej zodpovedné využívanie. Prináša kurikulum, workshopy a praktické materiály pre školy, ktoré podporujú kreativitu, digitálnu bezpečnosť a kritické myslenie žiakov v rýchlo sa meniacej technologickej dobe.

Digitálna bezpečnosť a osveta

- **Zodpovedne.sk**
    - **Web:** www.zodpovedne.sk
    - **Popis:** Národný osvetový projekt, ktorý sa komplexne venuje ochrane detí a mladých ľudí v online priestore. Ponúka množstvo materiálov pre učiteľov, rodičov a deti na témy kyberšikany, sextingu a bezpečného používania internetu. Ponúkajú prednášky pre študentov o digitálnych témach priamo na školách.
- **Beznástrah.online**
    - **Web:** www.beznastrah.online
    - **Popis:** Portál prevádzkovaný Ministerstvom vnútra SR, ktorý poskytuje praktické rady a návody, ako sa chrániť pred rôznymi nástrahami internetu, vrátane kyberšikany, podvodov a extrémizmu.

Bezpečne na nete

- **Web:** www.bezpecnenanete.sk
- **Popis:** Komplexný portál venujúci sa rizikám digitálneho sveta, ktorý poskytuje praktické tipy, ako sa chrániť. Je to projekt Nadácie Orange v spolupráci s ďalšími odbornými organizáciami.

E-bezpečí.cz

- **Web:** www.e-bezpeci.cz
- **Popis:** Popredný český projekt pre prevenciu rizikového správania na internete. Ponúka rozsiahly výskum, metodiky pre učiteľov a vzdelávacie materiály, ktoré sú vysoko relevantné aj pre slovenské prostredie.

Cyberwatch

- **Web:** https://cyberwatch.help/
- **Popis:** Zameriavajú sa na ochranu detí, dospelých a staršej generácie pred nebezpečenstvami v online svete. Ich cieľom je bojovať proti kyberšikane, pomáhať obetiam zneužívania detí a poskytovať vzdelávanie pre rodičov a starších ľudí.

Linky pomoci a duševné zdravie

- **IPčko**
    - **Web:** www.ipcko.sk

    - **Popis:** Internetová linka dôvery pre mladých ľudí a študentov. Poskytuje bezplatnú a anonymnú psychologickú pomoc a poradenstvo cez chat a e-mail. Je to kľúčový zdroj pre deti v kríze.

- **Krízová linka pomoci**
    - **Web:** www.krizovalinkapomoci.sk
    - **Popis:** Nepretržitá a bezplatná telefonická linka pomoci pre kohokoľvek, kto prežíva náročné životné situácie a psychickú krízu.
- **Dobrá linka**
    - **Web:** www.dobralinka.sk
    - **Popis:** Psychologická internetová poradňa špeciálne zameraná na mladých ľudí so zdravotným znevýhodnením. Poskytuje bezpečný priestor na riešenie tém, ktoré sú pre túto cieľovú skupinu špecifické.
- **Chuť žiť**
    - **Web:** www.chutzit.sk
    - **Popis:** Organizácia a poradenská služba pre ľudí, ktorí bojujú s poruchami príjmu potravy, problémami so sebaprijatím a vnímaním vlastného tela (body image), čo sú témy úzko prepojené s tlakom sociálnych sietí.

Vzťahová a sexuálna výchova

- **Intymita**
    - **Web:** www.intymita.sk
    - **Popis:** Projekt, ktorý sa venuje modernej, rešpektujúcej a vekovo primeranej vzťahovej a sexuálnej výchove. Ponúka workshopy a materiály, ktoré pomáhajú otvárať citlivé témy informovaným a zdravým spôsobom.

Technologické vzdelávanie a komunity

- **Aj ty v IT**
    - **Web:** www.ajtyvit.sk
    - **Popis:** Nezisková organizácia, ktorej poslaním je motivovať a podporovať dievčatá a ženy v oblasti informačných technológií. Organizujú workshopy, akadémie a networkingové podujatia.

Mediálna gramotnosť a boj proti dezinformáciám

- **Demagog.sk**

- **Web:** www.demagog.sk
- **Popis:** Hlavný slovenský fact-checkingový portál, ktorý overuje výroky politikov a informácie vo verejnom priestore. Skvelý nástroj na výučbu overovania faktov.

Zdroje

1. Technology - Amnesty International https://www.amnesty.org/en/what-we-do/technology/
2. Ethical AI principles won't solve a human rights crisis - Amnesty International, https://www.amnesty.org/en/latest/research/2019/06/ethical-ai-principles-wont-solve-a-human-rights-crisis/
3. Toronto Declaration, https://www.torontodeclaration.org/
4. The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems https://www.accessnow.org/press-release/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/
5. AI Act | Shaping Europe's digital future - European Union https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
6. High-level summary of the AI Act | EU Artificial Intelligence Act https://artificialintelligenceact.eu/high-level-summary/
7. Návrh zákona o organizácii štátnej správy v oblasti umelej inteligencie - Škubla & Partnerihttps://www.skubla.sk/clanky/navrh-zakona-o-organizacii-statnej-spravy-v-oblasti-umelej-inteligencie
8. Zodpovedná digitalizácia: MIRRI predstavilo návrhy zákonov o umelej inteligencii a správe údajov | Ministerstvo investícií, regionálneho rozvoja a informatizácie SR, otvorené novembra 25, 2025, https://mirri.gov.sk/aktuality/digitalna-agenda/zodpovedna-digitalizacia-mirri-predstavilo-navrhy-zakonov-o-umelej-inteligencii-a-sprave-udajov/
9. EU Simplification: Throwing human rights under the Omnibus - Amnesty International, otvorené novembra 25, 2025, https://www.amnesty.org/en/latest/news/2025/11/eu-simplification-throwing-human-rights-under-the-omnibus/
10. Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal - Amnesty International, https://www.amnesty.org/en/documents/eur35/4686/2021/en/
11. How We Did It: Amnesty International's Investigation of Algorithms in Denmark's Welfare System https://gijn.org/stories/amnesty-internationals-investigation-algorithms-denmarks-welfare-system/

12. Ban The Scan - Amnesty Internationalhttps://www.amnesty.org/en/petition/ban-the-scan-petition/
13. Ban dangerous facial recognition technology that amplifies racist policing, https://www.amnesty.org/en/latest/press-release/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/
14. Ban the Scan, otvorené novembra 25, 2025, https://www.banthescan.org/
15. USA: Facial recognition technology reinforcing racist stop-and-frisk policing in New York – new research - Amnesty International https://www.amnesty.org/en/latest/news/2022/02/usa-facial-recognition-technology-reinforcing-racist-stop-and-frisk-policing-in-new-york-new-research/
16. Stop Killer Robots – Less Autonomy, More humanity. https://www.stopkillerrobots.org/
17. Syrian Archive | Nesta, https://www.nesta.org.uk/feature/ai-and-collective-intelligence-case-studies/syrian-archive/
18. Assembling Atrocity Archives for Syria: Assessing the Work of the CIJA and the IIIM - Oxford Academic https://academic.oup.com/jicj/article/19/5/1193/6423113
19. Súpis všetkých metodických materiálov učebných osnov AI - Kurikulum umělé inteligence, otvorené novembra 25, 2025, https://kurikulum.aidetem.cz/sk/supis-vsetkych-metodickych-materialov-metodik-ucebneho-planu-ai/
20. Koncepcia učebných osnov umelej inteligencie pre základné a stredné školy - Kurikulum umělé inteligence, otvorené novembra 25, 2025, https://kurikulum.aidetem.cz/sk/koncepcia-ucebnych-osnov-umelej-inteligencie-pre-zakladne-a-stredne-skoly/
21. Soupis všech metodických materiálů AI kurikula - Kurikulum umělé inteligence - AI dětem, otvorené novembra 25, 2025, https://kurikulum.aidetem.cz/soupis-vsech-metodickych-materialu-metodik-ai-kurikula/
22. Informatika_Karty_Etika_Etika vývoje a využití technologií, otvorené novembra 25, 2025, https://kurikulum.aidetem.cz/metodiky/karty/Informatika_Karty_Etika_01_etika-vyvoje-a-vyuziti-technologii_plan-lekce.pdf
23. https://aiethicsframework.substack.com/p/when-healthcare-ai-goes-wrong-the