UNIVERZITA KOMENSKÉHO V BRATISLAVE

Vedecká rada Rímskokatolíckej cyrilometodskej bohosloveckej fakulty

ThLic. Ing. Peter Šantavý

Autoreferát dizertačnej práce

Etické výzvy a morálne aspekty súčasných systémov umelej inteligencie

na získanie: **akademického titulu „doktor“**
v študijnom odbore doktorandského štúdia: **teológia**
študijný program: **katolícka teológia**

BRATISLAVA 2022

Dizertačná práca bola vypracovaná v externej forme doktorandského štúdia na Univerzite Komenského v Bratislave, Rímskokatolíckej cyrilometodskej bohosloveckej fakulte.

Predkladateľ: **ThLic. Ing. Peter Šantavý**
RKCMBF UK Bratislava

Školiteľ: **ThDr. Ing. Vladimír Thurzo, PhD.**
RKCMBF UK Bratislava

Oponenti: **doc. ThDr. Radoslav Lojan, PhD.**
TF KU Košice

doc. ThDr. Patrik Maturkanič, PhD.
VŠAPs Terezín

prof. ThDr. Marian Šuráb, PhD.
RKCMBF UK Bratislava

Autoreferát bol rozoslaný: 6.6.2022
Obhajoba dizertačnej práce sa koná 22.6.2022 o 12.00
pred komisiou pre obhajobu dizertačnej práce v študijnom programe doktorandského štúdia katolícka teológia
vymenovanou dekanom fakulty dňa 6.6.2022
pre študijný odbor teológia
na Univerzite Komenského v Bratislave, Rímskokatolíckej cyrilometodskej bohosloveckej fakulte.

doc. ThDr. Ing. Jozef Jančovič, PhD.
dekan RKCMBF UK,
Kapitulská 26, Bratislava

Prežívame dejinný úsek, ktorý sa hrdí takými prívlastkami ako digitálna disrupcia, rodiaca sa informačná a znalostná spoločnosť, zmena paradigmy, permanentná technologická revolúcia, atď. V tomto kontexte takmer každá technologická novinka sa pýši dodatkom, že v jej útrobách je implementovaná umelá inteligencia. Nájdeme a vieme vymenovať veľa oblastí, v ktorých prvky umelej inteligencie zohrávajú čoraz dôležitejšiu úlohu: v lekárskej diagnostike a liečbe, v riadení technologických procesov, predikcii vývoja, spracovaní obrazu, analýze a syntéze reči, kybernetickej bezpečnosti, doprave, zábave, komunikačnej technike, genetickom výskume, jadrovej fyzike, astrofyzike, atď.

Javí sa, že umelá inteligencia nie je len *buzzword*, ktorý má pomôcť technologickým firmám presadiť sa a zarobiť, či okrášliť stratégie lídrov dnešného sveta, ale ide o reálny koncept a integrálnu súčasť technologickej budúcnosti našej civilizácie. Systémy umelej inteligencie vo viacerých oblastiach posúvajú úroveň poznania a skúmanie míľovými krokmi vpred. Dokážu byť extrémne nápomocné pri záchrane ľudských životov, ochrane zdravia i zaradení sa do reálneho života po ťažkých chorobách a úrazoch. Stávajú sa takmer nepostrádateľnými asistentmi v bežnom živote, v doprave, komunikácii, zábave. Prinášajú extrémny posun vpred pri analýze a spracovaní vedeckých dát. Pridávajú ďalší stupeň ochrany pri bezpečnostných systémoch, v obrane, v boji so zločinom a pod. Takmer v každej oblasti ľudskej činnosti a života človeka nájdeme niečo, v čom sa použitie prvkov umelej inteligencie stáva veľmi osožným.

Využitie umelej inteligencie má však aj svoju temnú stranu a riziká: jednou z privilegovaných oblastí vývoja umelej inteligencie sú autonómne zbraňové systémy a samostatné nasadenie v boji; využitie prvkov umelej inteligencie v kybernetickej kriminalite je už teraz nočnou morou informačnej bezpečnosti; zneužitie umelej inteligencie pri rozpoznávaní tvárí, komplexnom monitoringu a kategorizácii ľudí, detekcii a predikcii vývoja v spoločnosti je svätým grálom akéhokoľvek autoritárskeho režimu; vplyv sociálnych médií poháňaných aktuálne nastavenými, či zneužitými algoritmami umelej inteligencie na psychiku človeka a rozvoj spoločnosti sa už teraz podľa niektorých odborníkov javí ako katastrofálny; vytváranie psychologických, zdravotných, sociologických, či iných profilov ľudí spájaním informácií z verejných zdrojov, sociálnych sietí a metadát dokáže vďaka umelej inteligencii vytvoriť nekompromisnú verejnú sondu do ľudskej duše a bezprecedentne odhaliť súkromie človeka; zneužitie falošnej identity, či vytvorenie falošných informácií o človeku bravúrne natrénovaným systémom umelej inteligencie môže danú osobu priviesť k totálnemu spoločenskému i osobnému kolapsu...

V tomto kontexte – ak také osobnosti, ako napr. Ray Kurzweil, predikujú umelej inteligencii schopnosť konkurovať, ba až nahradiť ľudskú bytosť a iní, ako napr. Douglas Hofstadter, vidia reálne riziká vyplývajúce z potenciálu napredovania v jej vývoji, **pridávame sa k hlasom, ktoré volajú po etickom zhodnotení, zvážení morálnych aspektov a definovaní pravidiel pre vývoj, používanie i samotné fungovanie systémov umelej inteligencie**.

Uvedomujeme si, že až na pár výnimiek, sa v súčasnosti umelou inteligenciu nazývajú informačné systémy, ktoré majú pramálo spoločné s víziou Kurzweila a obavami Hofstadtera, t.j. s víziou človeku konkurujúcej skutočnej umelej inteligencie. Nech však ide o akýkoľvek stupeň, či schopnosti týchto systémov, treba k nim zodpovedne pristupovať, aby sme v spoločnosti nielenže znovu nemuseli boľavo objavovať to známe „dobrý sluha a zlý pán“, ale aj nakoniec nezistili, že z umelej inteligencie dobrého sluhu vo svojej podstate ani nedokážeme reálne vytvoriť.

Môžeme povedať, že obavy zo zneužitia algoritmov umelej inteligencie nie sú ničím novým. Svet, ktorý nerozumie týmto technológiám a je naviac živený katastrofickými scenármi science fiction, je týchto obáv už niekoľko desaťročí plný.

Avšak v poslednej dekáde – v čase enormného napredovania vo vývoji a nasadení systémov umelej inteligencie – badať obavy iné:

- obavy spojené s rizikami pokročilej umelej inteligencie čoraz viac vyjadrujú ľudia, ktorí sú v tejto oblasti doma a častokrát patria k prvotriednym odborníkom, či technologickým vizionárom. Netreba však obavy spájať len s víziou nadľudskej umelej inteligencie – oveľa aktuálnejšie je riešenie mnohých výziev (medzi nimi i etických) spojených s prvkami umelej inteligencie, ktoré sú aktuálne vyvíjané a začínajú sa masovo používať naprieč rôznymi oblasťami modernej spoločnosti;
- riziká spojené s pokročilou umelou inteligenciou sú často spojené s disproporciou medzi vnímaním ľudstva a umelou inteligenciou, a tak nastavujú zrkadlo nášmu pohľadu na podstatu ľudského bytia: kým sme, čo nás definuje, do akej miery sme redukovateľní na technologické vyjadrenie, aká je hodnota a podstata ľudského bytia, atď... Teda otázky skôr filozofické, antropologické a tiež i teologické, ktoré sú tým aktuálnejšie, čím viac sa posúvame vo vývoji systémov umelej inteligencie dopredu a čím viac rastie náš apetít po vytvorení umelej inteligencie schopnej konkurovať ľudskému bytiu.

V skutočnosti sa však nachádzame vo svete, ktorý presahuje akademické úvahy o potenciáli a rizikách pokročilej umelej inteligencie. Viaceré moderné armády vyvíjajú a koketujú s nasadením autonómnych bojových systémov (USA, Rusko, Čína, Izrael,...), v niektorých krajinách sa zavádza legislatíva pre autonómne vozidlá, prvá krajina uznala práva pre inteligentného robota, prvým systémom, ktorý ešte pred prepuknutím pandémie ochorenia Covid-19 informoval, že prichádza nová epidémia koronavírusu, bola kanadská lekárska umelá inteligencia, začínajú sa realizovať lekárske operácie vedené systémami umelej inteligencie, japonská lekárska umelá inteligencia s veľkou presnosťou a prekračujúc schopnosti najlepších špecialistov deteguje vzácne typy rakoviny, umelá inteligencia sa podieľa na špičkových fyzikálnych výskumoch, prakticky celá špička firiem špecializujúcich sa na kybernetickú bezpečnosť masívne implementuje prvky umelej inteligencie do svojich riešení, pokroky v rozpoznávaní reči v spojení s ďalšími prvkami umelej inteligencie umožnili vznik inteligentných asistentov, systémy spracovania a rozpoznávania obrazu kategorizujú, titulkujú, filtrujú, či inak spracúvajú snímky a obrazový materiál v takmer všetkých cloudových riešeniach a koniec koncov primitívnejšie systémy umelej inteligencie sú súčasťou prakticky každého spracovania obrazu, navrhovania trás v mapových systémoch a marketingových nástrojov v elektronických médiách, mobilných zariadeniach a e-shopoch...

V kontexte informačnej spoločnosti sa teda už nebavíme o tom, či a kedy prvky umelej inteligencie nasadiť – veď sú tu medzi nami a ich nasadenie sa neustále rozširuje – ale ako, to znamená za akých podmienok, pre aké ciele, akým spôsobom a s akými dôsledkami by mala byť umelá inteligencia súčasťou nášho sveta. Súčasťou sveta, v ktorom má človeku a spoločnosti slúžiť (vizionár by povedal koexistovať). A preto ľudia i spoločnosť majú mať jasno v tom, ako vyzerá etický návrh, realizácia a využívanie systémov umelej inteligencie, pričom samotná umelá inteligencia v svojom autonómnom a adaptívnom fungovaní „dokáže“ etické princípy a zásady dodržať.

Pohybujúc sa už viac než dve desaťročia v oblasti kybernetickej bezpečnosti si uvedomujeme, že prakticky ešte nikdy nebolo badať toľké obavy z nasadenia novej technológie a tak seriózny záujem o etické otázky ako v prípade systémov umelej inteligencie. Kľúčové slovo či značka (hashtag) #AIEthics v komunite odborníkov neoznačuje okrajovú záležitosť, ktorá je mimo zorného uhľa pohľadu tých, čo problematike umelej inteligencii skutočne rozumejú, ale stáva sa súčasťou hlavného prúdu vývoja, implementácie a používania týchto systémov v rámci reálneho sveta.

V súčasnosti preto silnejú hlasy vyjadrujúce potrebu skúmať, navrhnúť, prijať a realizovať etický rámec vývoja, používania a fungovania umelej inteligencie – či už ide o oblasť vedeckého bádania alebo vývoja, samotnej realizácie i zodpovedného využívania. Od petícií a otvorených listov adresovaných OSN i niektorým vládam badať prechod k snahe systematicky túto oblasť podchytiť, preskúmať a legislatívne ukotviť nielen na národnej i medzinárodnej úrovni, no seriózne sa tejto téme venovať aj v rámci Katolíckej cirkvi.

V práci sme sa primárne venovali problematike etických výziev a morálnych aspektov súčasných systémov umelej inteligencie, ktorú poznáme pod spoločným označením slabá umelá inteligencia (ANI). Tento termín zahŕňa úzko špecializované systémy umelej inteligencie (narrow AI), ktoré sú optimalizované na zvládnutie konkrétnej úlohy, resp. množiny úloh. Ide súčasne o systémy slabej umelej inteligencie (weak AI), ktoré vykazujú inteligentné správanie na základe modelov a aplikovaných metód i tréningových dát. Hovoríme teda o systémoch, ktoré sú zamerané na riešenie konkrétnych úloh a sú závislé na ľudskom vstupe a konfigurácii.

Jedným z hlavných prínosov tejto práce je ponúknutý široký a primerane hlboký interdisciplinárny rámec, bez ktorého nie je možné úspešne realizovať skutočné riešenie etických problémov a výziev technológií umelej inteligencie. Ide o rámec, v ktorom sme dostatočne oboznámení aj s technologickou stránkou týchto systémov a psychologickými, sociologickými i právnymi aspektmi ich nasadenia.

V prvej kapitole sme preto ponúkli potrebný náhľad do problematiky umelej inteligencie, sumarizujúc jej základné vlastnosti, delenie a metódy. Osobitne sme sa venovali algoritmom inšpirovaným činnosťou mozgu, keďže ony sú základom prakticky všetkých pokročilých systémov umelej inteligencie. Poukázali sme na základné problémy niekoľkých dekád vývoja technológií umelej inteligencie i na masívne zavádzanie týchto prostriedkov v súčasnosti, pričom nezostal opomenutý ani futuristický presah fenoménu umelej inteligencie.

Druhá kapitola predstavuje základ pre **ďalší dôležitý prínos práce – identifikáciu, pomenovanie, analýzu a pochopenie rizík spojených s technológiami umelej inteligencie v celej možnej šírke spektra ich nasadenia**.

Pre reálne uchopenie problematiky etiky týchto technológií považujeme široko spektrálne uchopenie a komplexné pochopenie limitov a rizík súčasných systémov umelej inteligencie za podstatné. Preto sme sa v druhej kapitole pomerne obšírne venovali hlavným rizikovým faktorom a zraniteľnostiam algoritmov súčasných systémov umelej inteligencie, zaoberali sme sa bezpečnosťou procesov založených na týchto technológiách, vysvetľovali problematiku kybernetickej bezpečnosti ako integrálnej súčasti zabezpečenia funkčnosti systémov umelej inteligencie a neopomenuli sme ani riziká vyplývajúce z technologickej komplexnosti a potrebného infraštruktúrneho zázemia pre spoľahlivú činnosť v reálnom svete.

Analýza uvedených limitov a rizík bola základom pre rozoberanie negatívnych dôsledkov využívania technológií umelej inteligencie v spoločnosti, zahŕňajúc celé spektrum ich nasadenia od sociálnych sietí až po problematiku autonómnych vozidiel a rozširujúc tak interdisciplinárny rámec o psychologické a sociologické dôsledky ich využitia. Osobitne sme sa venovali oblasti dohľadových systémov, technológií využívaných v spravodajských službách a v rámci algoritmického riadenia štátu, keďže z pohľadu etických výziev ide o veľmi špecifické pole vývoja i nasadenia systémov umelej inteligencie. Azda najproblematickejšou oblasťou je adaptácia technológií umelej inteligencie vo vojenskej sfére – od využitia v armádnych spravodajských službách, cez modelovanie a simulácie technológií a procesov, virtualizačné nástroje a výcvik, až po vývoj a nasadenie smrtiacich autonómnych zbraňových systémov a kybernetických zbraní. Armádne využitie systémov umelej inteligencie v sebe obnáša celé spektrum otázok s potenciálom prevýšiť všetky ostatné etické dilemy a výzvy, preto sme tejto oblasti venovali osobitný priestor.

V prvej časti tretej kapitoly sme **ako ďalší osobitný prínos tejto práce sumarizovali naše etické postrehy a závery, ku ktorým sme dospeli v rámci analýzy limitov a rizík súčasných systémov umelej inteligencie**. Tento sumár problémov technológií umelej inteligencie s priamym či nepriamym dopadom na človeka a spoločnosť sa spolu výstupmi ďalších častí tretej kapitoly stal základom pre naše návrhy riešenia etických problémov a stanovenie všeobecných i špecifických etických zásad, ktoré následne predkladáme v štvrtej kapitole.

V ďalších častiach tretej kapitoly sme interdisciplinárny rámec rozšírili analýzou súčasných aktivít na poli etiky umelej inteligencie smerujúc k potrebným reguláciám na zabezpečenie etického rámca využívania týchto technológií. Osobitne sme sa venovali európskemu Aktu o umelej inteligencii, ktorý považujeme síce za náročnú, avšak v súčasnosti asi najprepracovanejšiu a najkomplexnejšiu (pripravovanú) reguláciu v oblasti umelej inteligencie s potenciálom ovplyvniť etiku využívania systémov umelej inteligencie vo veľkej časti sveta. Tretia kapitola bola zavŕšená analýzou súčasných aktivít Cirkvi na poli etiky umelej inteligencie, pričom sme sa osobitne zaoberali závermi rímskej konferencie renAIssance 2020, známej aj pod názvom Rome Call for Ethics.

Problematika etických výziev a morálnych aspektov súčasných systémov slabej umelej inteligencie (ANI) bola zavŕšená naším návrhom riešenia etických problémov umelej inteligencie, ktoré sme predkladali v štvrtej kapitole. Ponajprv išlo o vyjadrenie a naše chápanie základného, a to pozitívneho postoja k fenoménu umelej inteligencie vo svetle Zjavenia. Osobitne sme akcentovali interdisciplinárny rámec prístupu k problematike etiky umelej inteligencie, bez ktorého skutočné riešenie etických problémov a výziev technológií umelej inteligencie nie je možné úspešne realizovať.

K hlavným prínosom práce patrí náš návrh základnej štruktúry etických princípov a zásad, ktorý sme v štvrtej kapitole predstavili:

- rozšírili sme diapazón základného zamerania umelej inteligencie na človeka (human-centered AI) o kontext kresťanskej antropológie, inklúziu každej ľudskej bytosti bez diskriminácie so zreteľom na dobro ľudstva a spoločnosti v rozšírenej optike starostlivosti o náš spoločný a zdieľaný domov, teda o celý stvorený svet.
- definovali sme a objasnili principiálne požiadavky na dôveryhodné systémy umelej inteligencie, ktoré musia byť legálne, etické a robustné. Vo vedomí, že dokonalý systém umelej inteligencie neexistuje, sme navrhli minimálne legislatívne, etické a technologické požiadavky, resp. hranicu, od ktorej môžeme tieto technológie považovať za dôveryhodné.

- predstavili sme vlastnú sadu všeobecných a univerzálnych etických zásad.
- nami predstavené etické zásady sme porovnali so zásadami najdôležitejších súčasných aktivít a regulácií v oblasti umelej inteligencie, aby sme tak vyjadrili univerzálnosť a určitú nadčasovosť nášho návrhu.
- naším zámerom bolo navrhnúť a vytvoriť tak univerzálnu a všeobecnú množinu etických zásad, že podľa nej môžeme či už definovať – alebo ešte lepšie – z existujúcich legislatívnych rámcov a etických odporúčaní vyberať konkrétne a jasné odporúčania pre ich aplikáciu v reálnom svete. Takto sme i označili kombináciu etických odporúčaní vatikánskej konferencie renAIssance 2020 a obsahu európskeho nariadenia Akt o umelej inteligencii za v súčasnosti najvhodnejšie konkrétne a do legislatívneho rámca zasadené odporúčania pre etický vývoj, nasadenie a využívanie súčasných systémov umelej inteligencie.
- tiež sme považovali za dôležité popísať všetky oblasti nutnej implementácie etických noriem, eticko-právnych regulácií a morálnych zásad. Ide o oblasť tvorby systémov umelej inteligencie, poskytovateľov i používateľov týchto systémov a implementácie noriem i obmedzení priamo v systémoch umelej inteligencie. Akcentovali sme viaceré činitele, bez ktorých nie je možné tieto oblasti zasadiť do etického rámca vývoja, nasadenia a vyžívania – ide napr. o edukáciu a osvetu, pravidlá pre základný výskum, nutné podmienky vývoja a pod.

Vzhľadom na dôraz, ktorý sme v druhej kapitole kládli na oblasť pokročilého riadenia štátu, spravodajstva a plošného dohľadu i využívania systémov umelej inteligencie vo vojenskej oblasti, v štvrtej kapitole prinášame **naše vlastné závery, návrhy regulácií a etické odporúčania v týchto špecifických a dôležitých oblastiach, ktoré považujeme za ďalší osobitný prínos tejto práce**.

Ide o návrhy v rámci využívania technológií umelej inteligencie v oblasti pokročilého riadenia štátu, spravodajstva a plošného dohľadu; export týchto technológií do rizikových krajín; jasné pravidlá pre technológiami umelej inteligencie poháňané automatické smrtiace zbraňové systémy (LAWs), systémy automatického zameriavania a vyberania cieľov, automatické systémy schopné bez zásahu človeka rozhodnúť o smrtiacej reakcii akéhokoľvek druhu (od útoku dronu až po rozpútanie jadrového konfliktu); podmienky prevádzky akýchkoľvek systémov umelej inteligencie, ktoré môžu predstavovať riziko pre ľudskú osobu; v neposlednom rade aj stanovenie limitov, regulácií a obmedzení LAWs vo forme etického rámca postaveného na základe morálnych hodnôt ľudskej spoločnosti, a nie na základe relativistickej tzv. „následnej regulácie“.

V závere štvrtej kapitoly **sme predložili niekoľko návrhov pre využitie potenciálu Cirkvi a osobitnú angažovanosť podporujúcu etický prístup k problematike umelej inteligencie v celej jej šírke**. Ide predovšetkým o misiu zjednocovať, usmerňovať a propagovať etické aktivity vo svete a neustávajúcu snahu akcentovať a budovať univerzálne bratstvo a sociálne priateľstvo aj v oblasti digitálneho sveta a jeho technológií.

Sekundárnym aspektom nášho úsilia bolo v kontexte súčasných technológií ANI poukázať i na problematiku uvedomelej umelej inteligencie (AGI), ktorá by podľa jej protagonistov mala byť dosiahnuteľná prostredníctvom silnej (strong) a všeobecnej (general) umelej inteligencie. Všeobecnej, lebo dokáže zvládnuť akúkoľvek intelektuálnu úlohu a má schopnosť generalizovať, t.j. zovšeobecňovať a prenášať, či adaptovať naučené schopnosti na iné úlohy. Silnej, pretože aj skutočne rozumie tomu, čo rieši a vykonáva.

I keď prakticky pri všetkých témach rozoberaných v prvých štyroch kapitolách sme sa nevyhli aspoň krátkemu pohľadu za horizont – k systémom silnej a všeobecnej umelej inteligencie, až v piatej kapitole sme sa výlučne venovali jej viacerým podstatným problémom, pričom **naše uchopenie problematiky uvedomelej umelej inteligencie taktiež považujeme za dôležitý prínos tejto práce**:

- diskutovali sme teóriu mysle a zdravého rozumu so schopnosťou abstrakcie, analógie, konceptualizácie, simulácie, či chápania zmyslu, identifikujúc tak bariéru chápania zmyslu, venovali sme sa hypotéze stelesnenia a predstavili sme komplexné dôvody, pre ktoré nie sme v súčasnosti schopní vytvoriť silnú a všeobecnú umelú inteligenciu.
- v kontexte nielen klasickej definície, ale i modernistickej snahy o redefiníciu pojmu osoby sme polemizovali s víziou uvedomelej umelej inteligencie a poukázali na dôvody, prečo ani najpokročilejším systémom ANI nemôžeme priradiť štatút osoby.
- na základe kresťanskej antropológie sme vyjadrili výnimočnosť ľudskej bytosti a tým aj odmietnutie štatútu osoby pre akúkoľvek umelú inteligenciu, vrátane AGI.
- použijúc materialistický uhol pohľadu – odmietajúci duchovnú dušu ako „formu“ tela, zásadne sa podieľajúcu na vedomí a sebauvedomení – sme okrem iného poukázali na v súčasnosti neriešiteľnú problematiku implementácie mechanizmu svedomia.
- osobitne sme akcentovali reálny problém funkčného mechanizmu svedomia, ktoré má referenčný bod mimo seba. U hypotetickej analógie svedomia uvedomelej umelej inteligencie nie sme schopní garantovať, že bude zdieľať rovnaké hodnoty a rovnakým spôsobom ich aj chrániť a zachovávať.

Celkovo prácu predkladáme ako určité nóvum, snažiac sa o inovatívny prístup v nazeraní na problematiku umelej inteligencie v kontexte morálnej teológie. Interdisciplinárne uchopenie fenoménu umelej inteligencie, dôsledná analýza limitov i rizík a z nej prameniace etické závery, návrh základnej štruktúry všeobecných i špecifických etických princípov a zásad, vyjadrenie diskutabilných faktorov uvedomelej umelej inteligencie a zdôvodnenie nášho jasného postoja k jej porovnávaniu s ľudskou bytosťou – to všetko tvorí náš vklad do prebiehajúceho celosvetového etického diskurzu, ktorý tým viac naberá na dôležitosti, čím viac sa technológie umelej inteligencie rozvíjajú a jej sofistikované implementácie sa do reálneho života masívne zavádzajú, ovplyvňujúc tak životy miliónov ľudí.

Nástojčivosť tejto témy vynikne, ak si z našej rozpravy o oblastiach implementácie etických princípov (kap. 4.3.4) pripomenieme, že v prípade pokročilých systémov umelej inteligencie – na rozdiel od iných informačných systémov a technológií – prakticky nie je možné na existujúce a už nasadené systémy spoľahlivo aplikovať etické zásady a regulácie prostredníctvom doplnenia technických úprav alebo procesných postupov. Jednoducho povedané, s etikou musíme vývoj týchto systémov sprevádzať a ich realizáciu predchádzať, inak nám – nielen jednotlivcom, ale celej spoločnosti digitálneho veku – tieto technológie prerastú cez hlavu.

Dúfame, že táto práca prispeje k rozšíreniu (nielen) etických obzorov v jednej z najdynamickejšie sa rozvíjajúcich oblastí digitálneho sveta a láskavý čitateľ bude zhovievavý k prípadným nepresnostiam a nedostatkom nášho pohľadu na problematiku etických výziev a morálnych aspektov súčasných systémov umelej inteligencie.

SUMMARY

Digital disruption, emerging information and knowledge society, paradigm shift, permanent technological revolution, etc. are terms exceedingly prominent in the current historical era. In this context, the core of almost every technological innovation takes pride in having artificial intelligence implemented in it. It seems, however, that artificial intelligence is not only a buzzword used to help technology companies to achieve breakthrough and earn money, nor to improve the strategies of the world's leaders. It has become a real concept and will play an integral part in the future of our civilization.

AI systems cause the levels of knowledge and research in various fields to progress rapidly. Moreover, they can be helpful and very useful in almost the whole spectrum of human activity, equally, they are becoming of bigger importance to the life of society.

Nevertheless, the application of AI technology has its dark side and risks. Their failure, misuse or direct exploitation are becoming a nightmare of security, democracy, sociology, and psychology. In order to attain conscious AI systems, the society endeavours to go beyond the current AI systems. This brings about a dilemma, whether this effort will lead to a "golden age" of the civilization's development, or it will become a downfall of all that has been built and achieved for the good of humanity so far.

In this regard, we take the side of voices calling for ethical evaluation, consideration of moral aspects, as well as for defining rules for development, use, and operation of AI systems as such. Due to the progress of information society and the outset of digital age, the questions *if* to use the components of AI technology and *when*, are not taken into consideration because they eventually already exist among us and their presence is expanding constantly. Rather the following question needs to be asked: *how* to use them, i.e. under what conditions, for what purposes, in what manner, and with what consequences the phenomenon of AI should become a part of our world.

With over twenty years of experience in the field of cyber security, we come to realize that AI systems have caused many reasons to worry about implementing new technology and a serious concern for ethical queries like nothing before. The use of #AIEthics as a hashtag and a key word among AI experts is becoming a part of the main flow of AI systems development, implementation, and use within the real world. It is not of peripheral importance for those who are truly experienced in problems associated with AI.

The aim of this study is to propose the basic principles to be upheld during ethics proposal, execution and usage of any AI systems, based on the analysis of the current state of AI systems development and existing ethical rules.

The study is primarily focused on ethical challenges and moral aspects of modern systems of artificial intelligence referred to under the common term Artificial Narrow Intelligence (ANI). This term covers narrow AI systems which are optimized to perform a specific task, or rather a set of tasks. At the same time, they are called weak AI systems and display intelligent behaviour based on some models, applied methods and training data. The goal of such systems is therefore a solution of specific tasks; they are dependent on interference of humans as well as on human configuration.

A broad and suitably deep interdisciplinary framework proposed is one of a significant contribution of this paper. Without it, no real and successful solution to ethical issues and AI technology challenges

would be possible. This framework covers a satisfactory amount of information about technology of the systems as well as the psychological, sociological, and legal aspects of their use.

After having been informed of the nature of this framework, another contribution of the paper is formed: identification, naming, analysis and comprehension of risks connected to AI technologies in respect to using them to the extent possible. It is key to have a broader spectrum of knowledge and complex comprehension of the limits and risks of modern AI systems in order to truly understand the ethical issues of such technologies.

The following distinctive contribution made, is a summary of ethical observations, formed during the analysis of limits and risks of present AI systems.

In an effort to fulfil the purpose of this study we gradually broadened the interdisciplinary framework by analysing the present activities in the AI ethics field, both in examining existing activities and upcoming regulations which aim to secure the ethical ambit of using such technologies.

The main finding of our study is the proposal of the basic ethical principles and guidelines for development, implementing and usage of AI systems. Considering the analysis emphasizing the progressive management of a state, intelligence, global surveillance and AI systems integrated in military, the paper presents our own conclusions, proposals for regulations and recommendations for ethical principles; even for such specific and important areas listed above.

Several proposals for making the most of the Church's potential are presented, as well as the involvement facilitating the support of ethical approach to the issues with the use of AI in a complete extent of range. Mostly it applies to the mission to unite, guide and promote ethical activities in the world; as well as to a constant effort to emphasize and build universal fraternity and social friendship, even in the digital world and its technologies.

The secondary aspect of our effort is to point to the problems of conscious Artificial General Intelligence (AGI) in respect of current ANI technologies. According to protagonists, AGI should be achievable by strong and general artificial intelligence. The term general stands for its ability to generalize and to cope with any intellectual task; to transfer abilities across tasks or to adapt them for another tasks. The term strong stands for its real understanding of everything it deals and is tasked with.

The fifth chapter discusses various fundamental problems exclusively, though the topics analysed in the previous four chapters provide at least a quick glance beyond the horizon towards strong and general artificial intelligence. Another important contribution of this study is represented by our grasp of the ethical issues of conscious AI (e.g. theory of mind, barrier of meaning, cracks in intelligence, embodiment hypothesis, consciousness, the term "person", implementation of mechanism of conscience, etc.).

Altogether, we present the study as a novelty in a sense of striving for innovative approach to exploring AI issues with respect to moral theology. Our contribution into the global ethical discourse is comprised of the interdisciplinary understanding of AI phenomenon, the thorough analysis of limits and risks and its ethical conclusion, the proposal of basic ethical principles (general and specific), the reference to arguable factors of conscious AI and the reasons given to support our clear view on comparing it to a human being. This debate is becoming more significant with the evolutionary progress of AI technologies, along with the increasingly growing sophisticated implementation of them into the real world which affects the lives of millions.

BIBLIOGRAFIA – výber z použitej literatúry

ADIB-MOGHADDAM, A. *Artificial intelligence must not be allowed to replace the imperfection of human empathy*. [on-line]. [cit. 20. februára 2022].
Dostupné na internete: <https://theconversation.com/artificial-intelligence-must-not-be-allowed-to-replace-the-imperfection-of-human-empathy-151636>

AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense. [on-line]. [cit. 9. marca 2022].
Dostupné na internete: <https://admin.govexec.com/media/dib_ai_principles_-_supporting_document_-_embargoed_copy_(oct_2019).pdf>

ALFONSEC, M., CEBRIAN, M. et al. *Superintelligence Cannot be Contained: Lessons from Computability Theory.* [on-line]. [cit. 26. januára 2021].
Dostupné na internete: <https://jair.org/index.php/jair/article/view/12202>

Algoritmy strojového učenia I. [on-line]. [cit. 3. januára 2022].
Dostupné na internete: <https://umelainteligencia.sk/algoritmy-strojoveho-ucenia/>

Algoritmy strojového učenia II. [on-line]. [cit. 3. januára 2022].
Dostupné na internete: <https://umelainteligencia.sk/algoritmy-strojoveho-ucenia-ii-ucenie-bez-ucitela/>

Algoritmy strojového učenia III. [on-line]. [cit. 4. januára 2022].
Dostupné na internete: <https://umelainteligencia.sk/algoritmy-strojoveho-ucenia-iii-ucenie-formou-odmenovania/>

AMODEI, D., HERNANDEZ, D. *AI and Compute*. [on-line]. [cit. 19. januára 2022].
Dostupné na internete: <https://openai.com/blog/ai-and-compute/>

An Open Letter to the United Nations: Convention on Certain Conventional Weapons. [on-line]. [cit. 21. marca 2022].
Dostupné na internete: <https://futureoflife.org/2017/08/20/autonomous-weapons-open-letter-2017/>

ANDERSON, K., WAXMAN, M. C. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*. In: *Jean Perkins Task Force on National Security and Law Essay Series*. Stanford University: Hoover Institution Press, 2013.

ANDERSON, K., WAXMAN, M. C. *Law and Ethics for Robot Soldiers*. In: *Policy Review*. 2012, 176, 12.

Artificial Intelligence: Principles, laws, and frameworks. OneTrust DataGuidance Limited, 2022. ISSN 2398-9955.

Autonomous Weapons: An Open Letter from AI [Artificial Intelligence] & Robotics Researchers. [on-line]. Future of Life Institute, 2015. [cit. 8. marca 2022].
Dostupné na internete: <http://futureoflife.org/open-letter-autonomous-weapons/>

BENGIO, Y. *Machines Dream*. In: BEYER D. ed. *The Future of Machine Intelligence: Perspectives from Leading Practitioners*. Sebastopol, Calif.: O'Reilly Media, 2016.

BOGERT, E., SCHECTER, A., WATSON, R. T. *Humans rely more on algorithms than social influence as a task becomes more difficult*. In: *Sci Rep*. [on-line]. 2021, 11, 8028. [cit. 20. februára 2022].
DOI: 10.1038/s41598-021-87480-9
Dostupné na internete: <https://doi.org/10.1038/s41598-021-87480-9>

BOSTROM, N. *A history of transhumanist thought*. In: *Journal of Evolution and Technology*. [on-line]. 2005, roč. 14, vyd. 1. [cit. 8. januára 2022].
Dostupné na internete: <http://www.nickbostrom.com/papers/history.pdf>

CLAPPER, J. R. Jr. et al. *Unmanned Systems Roadmap: 2007-2032*. [on-line]. Washington, DC: Department of Defense [DOD], 2007. [cit. 5. marca 2022].

Dostupné na internete: <http://www.globalsecurity.org/intell/library/reports/2007/dod-unmanned-systems-roadmap_2007-2032.pdf>

CLARK, A. *Being There: Putting Brain, Body, and World Together Again*. Cambridge, Mass.: MIT Press, 1996.

Compilation of open letters against autonomous weapons. [on-line]. [cit. 19. augusta 2020].
Dostupné na internete: <https://autonomousweapons.org/compilation-of-open-letters-against-autonomous-weapons/>

Conversations That Matter: The Crossroads of Science and Human Dignity. Fall 2021. [on-line]. [cit. 28. marca 2022].
Dostupné na internete: <https://mcgrath.nd.edu/conferences/academic-pastoral/conversations-that-matter-the-crossroads-of-science-and-human-dignity/conversations-that-matter-the-crossroads-of-science-and-human-dignity-fall-2021/>

Defense Science Board. *Task Force Report: The Role of Autonomy in DoD Systems*. Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, 2012.

DOBBINS, J., COHEN, R. S., CHANDLER, N. et al. *Overextending and Unbalancing Russia: Assessing the Impact of Cost-Imposing Options*. [on-line]. Santa Monica, CA: RAND Corporation, 2019. [cit. 18. marca 2022].
Dostupné na internete: <https://www.rand.org/pubs/research_briefs/RB10014.html>

Ethics guidelines for trustworthy AI. [on-line]. [cit. 28. marca 2022].
Dostupné na internete: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

ETZIONI, A., ETZIONI, O. *Keeping AI Legal*. In: *Vanderbilt Journal of Entertainment & Technology Law*. [on-line]. 2016, 19, no. 1, s. 133-146. [cit. 8. marca 2017].
Dostupné na internete: <http://www.jetlaw.org/wp-content/uploads/2016/12/Etzioni_Final.pdf>

ETZIONI, A., ETZIONI, O. *Pros and Cons of Autonomous Weapons Systems.* [on-line]. [cit. 5. marca 2022].
Dostupné na internete: <https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/May-June-2017/Pros-and-Cons-of-Autonomous-Weapons-Systems/>

FELDSTEIN, S. *The Global Expansion of AI Surveillance*. [on-line]. [cit. 28. februára 2022].
Dostupné na internete: <https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847>

First Turing Test success marks milestone in computing history. [on-line]. [cit. 29. januára 2021].
Dostupné na internete: <https://phys.org/news/2014-06-turing-success-milestone-history.html>

CHADHA, S. *"Common Sense" is the Dark Matter of Artificial Intelligence*. [on-line]. [cit. 4. februára 2022].
Dostupné na internete: <https://hackernoon.com/the-dark-matter-of-ai-common-sense-is-not-so-common>

GARAMONE, J. 9*/11 Drove Change in Intelligence Community, NSA Chief Says*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://www.defense.gov/News/News-Stories/Article/Article/945544/911-drove-change-in-intelligence-community-nsa-chief-says/>

Gartner Top Strategic Predictions For 2020 And Beyond. [on-line]. [cit. 23. marca 2022].
Dostupné na internete: <https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2020-and-beyond>

GEIST, E., LOHN, A. J. *How Might Artificial Intelligence Affect the Risk of Nuclear War?* [on-line]. Santa Monica, CA: RAND Corporation, 2018. [cit. 21. marca 2022].
Dostupné na internete: <https://www.rand.org/pubs/perspectives/PE296.html>

GIBNEY, A. *Zero Days*. [filmový dokument]. [cit. 10. marca 2022].
Dostupné na internete: <http://www.zerodaysfilm.com/>

GIDDA, M. *Edward Snowden and the NSA files – timeline.* [on-line]. In: *The Gurdian*. 2013, 23. 6. [cit. 23. februára 2022].
Dostupné na internete: <http://www.guardian.co.uk/world/2013/jun/23/edward-snowden-nsa-files-timeline>

Google » Tensorflow: Vulnerability Statistics. [on-line]. [cit. 11. januára 2022].
Dostupné na internete: <https://www.cvedetails.com/product/53738/Google-Tensorflow.html?vendor_id=1224>

Google » Tensorflow: Security Vulnerabilities. [on-line]. [cit. 11. januára 2022].
Dostupné na internete: <https://www.cvedetails.com/vulnerability-list/vendor_id-1224/product_id-53738/Google-Tensorflow.html>

HE, K., CHEN, X., XIE, S., LI, Y., DOLLÁR, P., GIRSHICK, R.: *Masked Autoencoders Are Scalable Vision Learners.* [on-line]. Facebook AI Research (FAIR), 2021. [cit. 18. novembra 2021].
Dostupné na internete: <https://arxiv.org/abs/2111.06377>

HE, S. et al. *Learning to predict the cosmological structure formation.* [online]. In: *Proceedings of the National Academy of Sciences.* 2019, roč. 116, č. 28, s. 13825. [cit. 6. augusta 2020].
DOI: 10.1073/pnas.1821458116
Dostupné na internete: <https://www.pnas.org/content/116/28/13825>

HOFSTADTER, D. R. *Analogy as the Core of Cognition*. [on-line]. Presidential Lecture, Stanford University, 2009. [cit. 7. apríla 2022].
Dostupné na internete: <https://www.youtube.com/watch?v=n8m7lFQ3njk>

HOFSTADTER, D. R. *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books, 1979. ISBN: 978-0-465-02656-2.

HOFSTADTER D. R. *Staring Emmy Straight in the Eye – and Doing My Best Not to Flinch*. In: DARTNELL T. *Creativity, Cognition, and Knowledge*. Westport, Conn.: Praeger, 2002.

HOFSTADTER, D. R, SANDER, E. *Surfaces and Essences*. New York: Basic Books, 2013.

How 9/11 Changed the Course of Personal Data Collection and Surveillance. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://www.startpage.com/privacy-please/startpage-articles/how-9-11-changed-the-course-of-personal-data-collection-and-surveillance>

China wants to make its own rules for AI ethics. [on-line]. [cit. 21. marca 2022].
Dostupné na internete: <https://www.scmp.com/abacus/tech/article/3029194/china-wants-make-its-own-rules-ai-ethics>

CHOI, Q. CH. *Superintelligent AI May Be Impossible to Control; That's the Good News.* [on-line]. [cit. 26. januára 2021].
Dostupné na internete: <https://spectrum.ieee.org/tech-talk/robotics/artificial-intelligence/super-artificialintelligence>

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. [on-line]. [cit. 31. marca 2022].
Dostupné na internete: <https://standards.ieee.org/industry-connections/ec/autonomous-systems/>

It's Getting Harder to Spot a Deep Fake Video. [on-line]. [cit. 21. marca 2022].
Dostupné na internete: <https://www.youtube.com/watch?v=gLoI9hAX9dw>

JÁN PAVOL II. *Redemptoris Missio*. Praha: Zvon, 1994.

JOBIN, A., IENCA, M., VAYENA, E. *The global landscape of AI ethics guidelines*. In: *Nat Mach Intell*. [on-line]. 2019, 1, s. 389–399. [cit. 28. marca 2022].
Dostupné na internete: <https://doi.org/10.1038/s42256-019-0088-2>

JOBIN, A., IENCA, M., VAYENA, E. *Artificial Intelligence: the global landscape of ethics guidelines*. [on-line]. [cit. 19. augusta 2020].
Dostupné na internete: <https://arxiv.org/pdf/1906.11668>

Katechizmus Katolíckej cirkvi. [on-line]. [cit. 10. apríla 2022].
Dostupné na internete: <https://katechizmus.sk/>

KNIGHT, W. *The Dark Secret at the Heart of AI*. In: *Technology Review*. [on-line]. 2017, 11. 4.. [cit. 10. februára 2022].
Dostupné na internete: <https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/>

KORAKOVOUNIS, D. *Spiking Neural Networks: where neuroscience meets artificial intelligence*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://theaisummer.com/spiking-neural-networks/>

KOUSHIK, J. *Understanding Convolutional Neural Networks*. [on-line]. [cit. 31. januára 2022].
Dostupné na internete: <https://arxiv.org/abs/1605.09081>

LECUN, Y., MISRA, I. *Self-supervised learning: The dark matter of intelligence.* [on-line]. [cit. 4. februára 2022].
Dostupné na internete: <https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence>

LEE, I. *Equalism: Paradise Regained*. In: LEE N. (ed.). *The Transhumanist Handbook*. Springer, 2019, s. 849 – 863.

LEHMAN, J., CLUNE, J., RISI, S. *An Anarchy of Methods: Current Trends in How Intelligence Is Abstracted in AI*. In: *IEEE Intelligent Systems*. 2014, 29, č. 6.

MAASS, W. *Networks of spiking neurons: The third generation of neural network models*. [on-line]. [cit. 28. februára 2022].
Dostupné na internete: <https://www.sciencedirect.com/science/article/abs/pii/S0893608097000117>

MARCUS, G. *Deep Learning: A Critical Appraisal*. [on-line]. [cit. 7. apríla 2022].
Dostupné na internete: <https://arxiv.org/abs/1801.00631>

MARCHANT, G. E. et al. *International Governance of Autonomous Military Robots*. In: Columbia Science and Technology Law Review. [on-line]. 2011, 12. 6., s. 272–276. [cit. 27. marca 2017].
Dostupné na internete: <http://stlr.org/download/volumes/volume12/marchant.pdf>

MCCARTHY, J. et al. *Proposal for the Dartmouth Summer Research Project in Artificial Intelligence*. [on-line]. In: *AI Magazine*. 1955, 27(4). [cit. 3. februára 2021].
Dostupné na internete: <https://doi.org/10.1609/aimag.v27i4.1904>

Measuring trends in Artificial Intelligence – 2021 AI Index Report. [on-line]. [cit. 28. marca 2022].
Dostupné na internete: <https://aiindex.stanford.edu/ai-index-report-2021/>

MERCER, C., TROTHEN, T. J. *Religion and Transhumanism: The Unknown Future of Human Enhancement*. Praeger, 2014.

MINSKY, M. L. *Computation: Finite and Infinite Machines*. Upper Saddle River, N.J.: Prentice Hall, 1967.

MINSKY, M. L. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster, 2006.

MINSKY, M. L., PAPERT, S. L. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, Mass: MIT Press, 1969.

MINSKY, M. *Society of Mind*. New York: Simon & Schuster, 1986.

MITCHELL, M. *An Introduction to Genetic Algorithms*. Cambridge, Mas: MIT Press, 1996.

MITCHELL, M. *Artificial Intelligence*. Farrar, Straus and Giroux, 2019, ISBN: 978-0-374-71523-6.

MITCHELL, M. *Conceptual Abstraction and Analogy in Artificial Intelligence*. [on-line]. In: *ALIFE 2020: The 2020 Conference on Artificial Life*. 2020. [cit. 5. apríla 2022].
Dostupné na internete: <https://doi.org/10.1162/isal_a_00354>

MOEWES, C., NÜRNBERGER, A. *Computational Intelligence in Intelligent Data Analysis*. New York: Springer, 2013.

MORAVEC, H. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, Mass.: Hardvard University Press, 1988.

MÜLLER, V. C., BOSTROM, N. *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*. In: *Fundamental Issues of Artificial Intelligence*. Cham, Switzerland: Springer International, 2016, s. 555-572.

Nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 z 27. apríla 2016 o ochrane fyzických osôb pri spracúvaní osobných údajov a o voľnom pohybe takýchto údajov, ktorým sa zrušuje smernica 95/46/ES (všeobecné nariadenie o ochrane údajov). [on-line]. [cit. 19. februára 2022].
Dostupné na internete: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=celex%3A32016R0679>

Nariadenie Európskeho parlamentu a Rady, ktorým sa stanovujú harmonizované pravidlá v oblasti umelej inteligencie (Akt o umelej inteligencii) a menia niektoré legislatívne akty únie. [on-line]. [cit. 24. marca 2022].
Dostupné na internete: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX:52021PC0206>

NG, A. *Deep Learning in Practice: Speech Recognition and Beyond*. In: *EmTech Digital*. [on-line]. 2016, 23. 5. [cit. 3. februára 2022].
Dostupné na internete: <https://events.technologyreview.com/video/watch/andrew-ng-deep-learning/>

NGUYEN, A., YOSINSKI, J., CLUNE, J. *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*. [on-line]. CVPR, 2015. [cit. 12. februára 2022].
Dostupné na internete: <https:// cv-foundation.org/openaccess/con
tent_cvpr_2015/papers/Nguyen_Deep_Neural_Networks_2015_CVPR_paper.pdf>

NIELSEN, M. *Neural Networks and Deep Learning*. [on-line]. [cit. 19. januára 2022].
Dostupné na internete: <http://neuralnetworksanddeeplearning.com/>

NILSSON, N. J., MCCARTHY, J. *A Biographical Memoir*. Washington D.C.: National Academy of Sciences, 2012.

One Hundred Year Study on Artificial Intelligence. [on-line]. AI100, 2016. [cit. 14. novembra 2021].
Dostupné na internete: <https://ai100.stanford.edu/2016-report>

One Hundred Year Study on Artificial Intelligence. [on-line]. AI100, 2021. [cit. 14. novembra 2021].
Dostupné na internete: <https://ai100.stanford.edu/2021-report>

Open Letter on Research Priorities for Robust and Beneficial Artificial Intelligence. [on-line]. [cit. 21. marca 2022].
Dostupné na internete: <https://futureoflife.org/2015/10/27/ai-open-letter/>

ORLOWSKI, J. *The Social Dilemma*. Netflix, 2020. [filmový dokument]. [cit. 7. decembra 2021].
Dostupné na internete: <https://www.netflix.com/sk/title/81254224>

PATTYNOVÁ, J. *Výzvy a právní aspekty umělé inteligence*. In: *Umělá inteligence 2021*. Praha: TUESDAY Business Network, 2021.
Dostupné na internete: <https://www.tuesday.cz/akce/umela-inteligence-1/zaznam-akce/>

PEARL, J., MACKENZIE, D. *The Book of Why: THe New Science of Cause and Effect*. New York: Basic Books, 2018.

PFEIFFER, M., PFEIL, T. *Deep Learning With Spiking Neurons: Opportunities and Challenges*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00774/full>

POITRAS, L., ROSENBACH, M., SCHMID, F., STARK, H. *NSA horcht EU-Vertretungen mit Wanzen aus.* [on-line]. In: *Der Spiegel*. 2013, 29. 6. [cit. 23. februára 2022].
Dostupné na internete: <http://www.spiegel.de/netzwelt/netzpolitik/nsa-hat-wanzen-in-eu-gebaeuden-installiert-a-908515.html>

POITRAS, L., ROSENBACH, M., STARK, H. *NSA überwacht 500 Millionen Verbindungen in Deutschland.* [on-line]. In: *Der Spiegel*. 2013, 30. 6. [cit. 23. februára 2022].
Dostupné na internete: <http://www.spiegel.de/netzwelt/netzpolitik/nsa-ueberwacht-500-millionen-verbindungen-in-deutschland-a-908517.html>

PRESS, G. 12 *Observations About Artificial Intelligence From The O'Reilly AI Conference*. [on-line]. In: *Forbes*. 2016, 31. 10. [cit. 7. augusta 2020].
Dostupné na internete: <https://www.forbes.com/sites/gilpress/2016/10/31/12-observations-about-artificial-intelligence-from-the-oreilly-ai-conference/>

REHÁK, M. *Útoky na systémy umělé inteligence a jejich obrana*. In: *Umělá inteligence 2021*. Praha: TUESDAY Business Network, 2021.
Dostupné na internete: <https://www.tuesday.cz/akce/umela-inteligence-1/zaznam-akce/>

Rome Call for AI Ethics. [on-line]. [cit. 19. augusta 2020].
Dostupné na internete: <http://www.academyforlife.va/content/pav/en/events/intelligenza-artificiale.html>

Rome Call for AI Ethics (document). [on-line]. [cit. 28. marca 2022].
Dostupné na internete: <https://www.romecall.org/wp-content/uploads/2022/03/RomeCall_Paper_web.pdf>

ROSENBLATT, F. *The Perceptron: A Probabilistic Model of Information Storage and Organization in the Brain*. In: *Psychological Review*. 1958, 65, č. 6.

ROTA, G.C. *Indiscrete Thoughts*. Boston: Berkhäuser, 1997.

RUMELHART, D. E., MCCLELLAND, J. L. *Parallel Distributed Processing*. Vol 1/2. Bradford Book, 1986.

RUSSELL, S. *Human Compatible*. Penguin Books, 2020, ISBN: 978-0-241-33524-6.

SAMPLE, I. *Thousands of leading AI researchers sign pledge against killer robots*. [on-line]. [cit. 21. marca 2022].
Dostupné na internete: <https://www.theguardian.com/science/2018/jul/18/thousands-of-scientists-pledge-not-to-help-build-killer-ai-robots>

SAMUEL, A. L. *Some Studies in Machine Learning Using the Game of Checkers*. In: *IBM Journal of Research and Development*. 1959, č. 3.

Scientists' Call to Ban Autonomous Lethal Robots. ICRAC, October 2013. [on-line]. [cit. 8. marca 2022].
Dostupné na internete: <http://www.icrac.net/>

SEARLE, R. J. *Minds, Brains, and Programs In: The Behavioral and Brain Sciences*. [on-line]. Cambridge University Press, 1980, zv. 3. [cit. 30. januára 2021].

Dostupné na internete:
<https://web.archive.org/web/20071210043312/http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>

SHARKEY, N. *Saying 'No!' to Lethal Autonomous Targeting*. In: *Journal of Military Ethics*. 2010, 9, č. 4, s. 369–383.

SIMON, H. A. *Artificial Intelligence: An Empirical Science*. In: *Artificial Intelligence*. 1955, 77, č. 2.

SIMON, H. A. *The Shape of Automation for Men and Management*. New York: Harper & Row, 1965.

SMITH, R. *Why No Science?* [on-line]. [cit. 6. novembra 2021].
Dostupné na internete: <https://www.thecatholicthing.org/2021/11/02/why-no-science/>

SPARROW, R. *Killer Robots*. In: *Journal of Applied Philosophy*. 2007, 24, č. 1, s. 62–77.

STRAHOVNIK, V. *Virtues and transhumanist human enhancement*. In: PETROUŠEK, R., ŽALEC, B. *Transhumanism as a Chalange for Ethics and Religion*. 2021, s. 37 – 44.

Stratégia kybernetickej obrany Slovenskej republiky. [on-line]. Ministerstvo obrany SR, Vojenské spravodajstvo. [cit. 17. marca 2022].
Dostupné na internete: <https://www.slov-lex.sk/legislativne-procesy/-/SK/dokumenty/LP-2022-128>

SZEGEDY, CH. et al. *Intriguing Properties of Neural Networks*. In: *Proceedings of the International Conference on Learning Representations*. 2014.

ŠANTAVÝ, P. *Analýza virtuálneho sveta v kontexte náboženskej formácie, pôsobenia a života Cirkvi* [licenciátska práca]. [on-line]. Bratislava: RKCMBF UK, 2017. [cit. 19. augusta 2020 – 7. decembra 2021].
Dostupné na internete:
<https://peter.santavy.cloud/docs/Analyza_virtualneho_sveta_v_kontexte_nabozenskej_formacie_posobenia_a_zivota_Cirkvi.pdf>

ŠANTAVÝ, P. *Dejiny spásy: zmluvy Starého zákona a Nová zmluva* [diplomová práca]. [on-line]. Bratislava: RKCMBF UK, 2000. [cit. 6. novembra 2021].
Dostupné na internete: <https://peter.santavy.cloud/docs/Dejiny_spasy-zmluvy_SZ_a_NZ.pdf>

ŠANTAVÝ, P. *Niektoré výzvy informačnej spoločnosti v oblasti morálnej teológie*. In *Doctorandum dies 2018: Varia historia et moralia*. RKCMBF UK, Bratislava 2018.

ŠTARHA, Š., GAŠPAROVIČ, R. *AI z pohľadu práva*. [on-line]. [cit. 29. marca 2022].
Dostupné na internete: <https://www.epravo.sk/top/clanky/ai-z-pohladu-prava-4483.html>

The AI Powered State. [on-line]. [cit. 26. marca 2022].
Dostupné na internete: <https://www.nesta.org.uk/feature/ai-powered-state/>

The „good" algorithm. [on-line]. [cit. 28. marca 2022].
Dostupné na internete:
<https://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/Assemblea/Atti_Assemblea_e_28febbraio/Atti%20completi_PAV_2020_.pdf>

The Nonhuman Rights Project: Frequently Asked Questions. [on-line]. [cit. 10. apríla 2022].
Dostupné na internete: <https://www.nonhumanrights.org/frequently-asked-questions/>

The Price of Privacy: Re-Evaluating the NSA. [on-line]. [cit. 23. februára 2022].
Dostupné na internete: <https://www.youtube.com/watch?v=kV2HDM86XgI&t=18m>

The Turing Digital Archive. [on-line]. [cit. 30. januára 2021].
Dostupné na internete: <http://www.turingarchive.org/>

THURZO, V. *The Influence of Existentialism and Subjectivism on the Concept of the Human Person*. In: *Spiritual and Social Experience in the Context of Modernism and Postmodernism*. Morrisville, 2021.

Top Strategic Predictions for 2020 and Beyond: Technology Changes the Human Condition. [on-line]. [cit. 24. marca 2022].
Dostupné na internete: <https://www.gartner.com/document/3970846>

THOMPSON, N. C., GREENEWALD, K., LEE, K., MANSO, G. F. *Deep learning computational cost*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://spectrum.ieee.org/deep-learning-computational-cost>

THURNHER, J. S. *Legal Implications of Fully Autonomous Targeting*. In: *Joint Force Quarterly*. [on-line]. 2012, 67, 4, s. 83. [cit. 7. marca 2022].
Dostupné na internete: <http://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-67/JFQ-67_77-84_Thurnher.pdf>

Transhumanist Declaration. [on-line]. [cit. 26. januára 2022].
Dostupné na internete: <https://hpluspedia.org/wiki/Transhumanist_Declaration>

TROTT, D. *The hard things are easy, but the easy things are hard*. [on-line]. [cit. 8. januára 2022].
Dostupné na internete: <https://www.campaignlive.com/article/hard-things-easy-easy-things-hard/1498154>

Trustworthy AI is human-centered. [on-line]. [cit. 20. februára 2022].
Dostupné na internete: <https://www.ibm.com/watson/trustworthy-ai>

TUCKER, P. *SecDef: China Is Exporting Killer Robots to the Mideast*. [on-line]. [cit. 6. decembra 2021].
Dostupné na internete: <https://www.defenseone.com/technology/2019/11/secdef-china-exporting-killer-robots-mideast/161100/>

TUCKER, P. *The Pentagon's AI Ethics Draft Is Actually Pretty Good*. [on-line]. [cit. 9. marca 2022].
Dostupné na internete: <https://www.defenseone.com/technology/2019/10/pentagons-ai-ethics-draft-actually-pretty-good/161005/>

Understanding China's AI Strategy. [on-line]. [cit. 21. marca 2022].
Dostupné na internete: <https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy>

URBINA, F., LENTZOS, F., INVERNIZZI, C. et al. *Dual use of artificial-intelligence-powered drug discovery*. In: *Nat Mach Intell*. [on-line]. 2022, 4, s. 189–191. [cit. 22. marca 2022].
Dostupné na internete: <https://doi.org/10.1038/s42256-022-00465-9>

Uznesenie Európskeho parlamentu zo dňa 16. 2. 2017 obsahujúce odporúčania pre Komisiu k normám občianskeho práva v oblasti robotiky (2015/2103(INL)). [on-line]. [cit. 29. marca 2022].
Dostupné na internete: <https://eur-lex.europa.eu/legal-content/SK/TXT/PDF/?uri=CELEX:52017IP0051&from=EN>

Vatican Hackathon – harnessing youth, technology to serve common good. [on-line]. [cit. 28. marca 2022].
Dostupné na internete: <https://www.vaticannews.va/en/vatican-city/news/2018-03/vatican-hackathon-.html>

VIGNARD, K. *Manifestos and open letters: Back to the future?* [on-line]. [cit. 21. marca 2022].
Dostupné na internete: <https://thebulletin.org/2018/04/manifestos-and-open-letters-back-to-the-future/>

WOODS, E. T. Jr. *Ako Katolícka cirkev budovala západnú civilizáciu*. Bratislava: Redemptoristi - Slovo medzi nami, 2010.

ZUBOFF, SH. *The real reason why Facebook and Google won't change*. [on-line]. [cit. 21. marca 2022].
Dostupné na internete: <https://www.theguardian.com/science/2018/jul/18/thousands-of-scientists-pledge-not-to-help-build-killer-ai-robots>