ANTHROPOLOGIA INTEGRA
13/2022/2
ČASOPIS PRO OBECNOU ANTROPOLOGII A PŘÍBUZNÉ OBORY
JOURNAL FOR GENERAL ANTHROPOLOGY AND RELATED DISCIPLINES

Etické aspekty rozvoje umělé inteligence

Anetta Jedličková

Fakulta humanitních studií UK Praha, Pracoviště doktorských studií, Obor Aplikovaná etika, Pátkova 2137/5, 182 00 Praha 8, email: Anetta.Jedlickova@fhs.cuni.cz

Do redakce doručeno 20. října 2022; k publikaci přijato 23. listopadu 2022

ETHICAL ASPECTS OF THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

ABSTRACT Computer data processing using artificial intelligence is becoming a daily routine in the public sector, as well as in the commercial sphere, and thus, essentially affects various personal areas of life. With the significant development of algorithmic decision-making in data processing by means of the machine learning technology, ethical aspects represent an important part of professional discussions in the development, deployment and use of systems with various levels of autonomous decision-making processes of artificial intelligence. The paper aims to acquaint professionals and the general public with ethical principles that are discussed across the various artificial intelligence ethics guidance literature, with some ethical implications that arise from the development, implementation and use of autonomous and intelligent systems, and introduces challenges related to establishing the appropriate ethical framework from a practical perspective.

KEY WORDS algorithmic decision-making, artificial intelligence; autonomous and intelligent systems; autonomous decision-making; ethical aspects; ethical dilemmas; ethical principles

ABSTRAKT Počítačové zpracovávání dat s využitím umělé inteligence se ve veřejném sektoru i v komerční sféře postupně stává každodenní rutinou, a zásadně tak zasahuje do rozličných osobních oblastí života lidí. S výrazným rozvojem algoritmického rozhodování při zpracování dat prostřednictvím technologie strojového učení neboli machine learning se do popředí odborných diskusí dostávají také etické aspekty při vývoji, zavádění a používání systémů využívajících při své činnosti různou míru autonomního rozhodování umělé inteligence. Článek si v jednotlivých částech klade za cíl obeznámit odbornou i laickou veřejnost s etickými principy, které jsou diskutovány specialisty v různých etických pokynech zabývajících se umělou inteligencí, s etickými aspekty, jež provází vývoj, implementaci a využívání autonomních a inteligentních systémů, a v závěru obeznamuje s problematikou formování příslušného etického rámce z praktického pohledu.

KLÍČOVÁ SLOVA algoritmické rozhodování; autonomní a inteligentní systémy; autonomní rozhodování; etické aspekty; etická dilemata; etické principy; umělá inteligence

UMĚLÁ INTELIGENCE A JEJÍ SCHOPNOSTI

Je zcela neoddiskutovatelné, že v mnoha oborech je již v současné době využívání počítačových technologií založených na autonomních a inteligentních systémech nejen neocenitelným pomocníkem, ale je nutné také zdůraznit, že při zpracovávání informací úroveň výkonu těchto technologií v některých parametrech neporovnatelně překonává výkon lidské inteligence. Vývoj v této oblasti jednoznačně směřuje k co největší míře zastoupení lidské intelektuální činnosti umělou inteligencí počítačových či robotických systémů. Pod pojmem umělá inteligence rozumíme systémy vykazující inteligentní chování na základě analýzy prostředí a následného rozhodování, přijímaní a provádění opatření s určitou mírou autonomie k dosažení konkrétních cílů (Evropská komise, 2018a). Pro umělou inteligenci je často používán její anglický ekvivalent Artificial Intelligence (AI), případně také pojem autonomní a inteligentní systémy (anglicky autonomous and intelligent systems – A/IS), který ve svých dokumentech používá Institut pro elektrotechnické a elektronické inženýrství (Institute of Electrical and Electronics Engineers – IEEE).
Podle úrovně schopností nalézt řešení či dosáhnout stanoveného cíle lze umělou inteligenci rozdělit na tři základní skupiny:

1. Artificial narrow intelligence (ANI), tzv. úzká umělá inteligence, která má úzký, přesně vymezený rozsah schopností a zaměřuje se na nejkvalitnější možné splnění specifického úkolu (např. rozpoznávání obličejů, hlasů, kontextoví asistenti, implementace informačních a optimalizačních systémů v průmyslu, roboticky asistovaná chirurgie, parkovací asistent automobilů a další). V současné praxi je již ANI běžně rozšířená a její využití nadále nabývá na významu v různých oblastech.
2. Artificial general intelligence (AGI), neboli obecná umělá inteligence, jejíž rozsah schopností je na stejné úrovni jako jsou schopnosti přirozené pro člověka. Jejím cílem je realizace úkonů na úrovni komplexní lidské inteligence a myšlení takovým způsobem, aby byla stejně kompetentní, jako je lidská mysl. Vývojem AGI se zabývá řada výzkumných projektů, v současnosti však neexistuje žádný funkční systém, který by byl svými rekurzivními algoritmy vylepšován na úroveň lidské mysli.
3. Artificial superintelligence (ASI), tzv. super inteligence, která výrazně přesahuje schopnosti člověka, tedy lidskou inteligenci převyšuje, a to v různých kognitivních oblastech. Zatím se jedná o neexistující úroveň inteligence. ASI lze dále teoreticky členit na rychlostní, kolektivní a kvalitativní super inteligenci. Rychlostní super inteligence odpovídá lidskému intelektu, je však řádově neporovnatelně rychlejší, čímž by překonala schopnosti člověka. Kolektivní super inteligence je tvořena soustavou spojení množství menších intelektových výkonů v celek s vyšším celkovým výkonem, který předčí kognitivní systém tvořen souborem všech kognitivních procesů člověka. Kvalitativní super inteligence představuje systém, který je alespoň stejně rychlý jako kognitivní procesy lidské mysli, ale kvalitativně je značně převyšuje. Je ovšem diskutabilní, zdali je vůbec možné přesně stanovit horní limit kvality lidské inteligence a zda schopností posuzování kvality kognitivní inteligence skutečně disponujeme (Bostrom 2014).

Umělá inteligence umožňuje prostřednictvím strojového učení vysoký stupeň autonomních aktivit v mnoha oblastech, je schopna se zcela autonomně přizpůsobovat různým změnám a předvídat vývoj. Mezi další její přednosti patří nejen schopnost rychle zpracovat množství různorodých informací, ale na základě jejich analýzy a vyhodnocení také především schopnost řešit rozmanité úkoly a problémy v poměrně krátkém časovém intervalu. Již dnes předčí ANI člověka obzvláště strategickým sekvenčním procesováním, tedy rychlostí, výkonem i kvalitou analytického rozboru v řešení specifického úkolu. Schopnosti umělé inteligence nabízejí ve srovnání s limity biologické výbavy lidské inteligence, která je určená vlastnostmi neuronů, axonů či synapsí, výhody zásadního významu. Patří mezi ně například rychlost a množství výpočetních operací, rychlost přenosů informací, paměťová kapacita, možnosti úprav a další.

ETICKÉ PRINCIPY PŘI VÝVOJI, ZAVÁDĚNÍ A POUŽÍVÁNÍ SYSTÉMŮ UMĚLÉ INTELIGENCE

K nejvýznamnějším tématům expertních disputací odborné veřejnosti, jež v souvislosti s progresivním rozvojem umělé inteligence vyžadují prioritní přístup, patří etické aspekty algoritmického rozhodování a strojového učení při zpracování dat. Zavádění technologií aplikované umělé inteligence a expertních systémů do běžné každodenní praxe s sebou přináší nové naléhavé výzvy a je nutné zajistit, aby algoritmické rozhodování nezpůsobovalo žádné újmy, ale naopak, aby byly moderní technologie zcela využívány jen ve prospěch společnosti i jednotlivců. Je potřeba si uvědomit, že jakkoli je umělá inteligence racionální, eticky jednat neumí a nečiní žádná morální rozhodnutí. Neumí definovat, co je spravedlivé, ani nezvládne posoudit následky svého rozhodování. Algoritmické rozhodování se může významně mýlit v kategorizaci rozličných údajů a charakteristik, a tedy může také závažně škodit, způsobovat újmu či diskriminovat, může se stát zdrojem nástrojů umožňujících manipulativní praktiky např. prostřednictvím podprahových technik, může taktéž vytvářet algoritmy, které jsou úmyslně nastaveny nesprávně, a autonomní inteligentní systémy proto mohou být rovněž zneužitelné, nebo mohou dokonce sloužit autoritativním režimům. Rozvoj umělé inteligence, robotika a související technologie budou mít za následek také výrazné změny na trhu práce a budou požadovat zásadní transformace z hlediska reorganizace pracovní síly, což může rovněž přinášet závažné etické implikace.

Významný dopad nových technologií umělé inteligence na jednotlivce i společnost vyžaduje kritickou reflexi potenciálních etických důsledků. Experti v dané problematice se jednoznačně shodují, že je zcela zásadní jasně stanovit požadavky, které musí vývojáři autonomních a inteligentních systémů splňovat již při jejich návrhu a vytváření, provozovatelé následně dodržovat při jejich zavádění a používání a neméně důležitá je také efektivní kontrola jejich dodržování a náležitý dohled.

POŽADAVKY NA ETICKÉ PRINCIPY V EVROPĚ

Důležitou roli při zajištění ochrany práv, soukromí, lidské důstojnosti či svobody všech zúčastněných má dodržování morálních zásad při vývoji, zavádění a používání autonomních a inteligentních systémů. V souvislosti s umělou inteligencí se jedná o známé čtyři základní etické principy[1], které jsou v praxi již běžně aplikovány v medicíně a v biomedicínském výzkumu, tj. princip beneficence, princip nonmaleficence, princip spravedlnosti a princip respektu k autonomii člověka

1 V souvislosti s etickým rozhodováním v medicíně vydali Tom L. Beauchamp a James F. Childress již v roce 1979 první vydání Principles of Biomedical Ethics. Aplikovaný etický přístup k řešení etických dilemat založen na aplikaci určitých etických principů se označuje anglickým termínem Principlism.

(Beauchamp – Childress 2019). Odborná skupina na vysoké úrovni pro umělou inteligenci (High-Level Expert Group on Artificial Intelligence – AI HLEG), která byla zřízena Evropskou komisí v červnu 2018, v dokumentu s názvem Etické pokyny pro zajištění důvěryhodnosti umělé inteligence sice neuvádí důležitý princip beneficence, který je často slučován s principem nonmaleficence, přidala ovšem další etický princip, který má základní etické principy v souvislosti s moderními technologiemi vhodně doplnit. Jde o princip vysvětlitelnosti, a to především prostřednictvím transparentnosti, srozumitelnosti a odpovědnosti (High-Level Expert Group 2019). Každý z principů obsahuje v souvislosti s autonomními a inteligentními systémy konkrétní požadavky:

- Princip beneficence (jako příklad uveďme využívání technologií pouze ve prospěch jedince a společnosti a snahu o maximalizaci přínosů pro jedince a společnost)
- Princip nonmaleficence (nezpůsobit ani nezhoršit individuální a kolektivní újmy, které zahrnují také nehmotné újmy, např. morální, sociální, společenské či psychické újmy, ochrana fyzické a duševní nedotknutelnosti, minimalizace rizik, vyhodnocení poměru risk/benefit s tím, že v případech, v nichž rizika pro jednotlivce či společnost převyšují přínos, nesmí být autonomní a inteligentní systém provozován, dále zamezení mocenské či informační asymetrie)
- Princip spravedlnosti (spravedlivé rozdělení přínosů i rizik pro všechny zúčastněné strany, zajištění spravedlivého přístupu, stejných práv a rovných příležitostí, rovnost podmínek při rozhodování, zamezení nespravedlivé podjatosti, diskriminace a stigmatizace, dodržování zásad proporcionality mezi prostředkem a účelem a také rovnováhy mezi protichůdnými zájmy a cíli, možnost zpochybnit nespravedlivá rozhodnutí přijatá systémem a možnost domoci se účinné nápravy, u relevantních systémů patří k principu spravedlnosti taktéž alokační rozhodování a distribuční spravedlnost)
- Princip respektu k autonomii člověka (respektování svobodné vůle a rozhodnutí s vyloučením ovlivňování či manipulace, umožnění jednotlivcům činit odůvodněná informovaná rozhodnutí, respekt k jejich hodnotovému systému, zavedení jasných pravidel a postupů pro racionální rozhodování umělé inteligence při zachování autonomie člověka, ochrana fyzických osob v souvislosti se zpracováním osobních údajů, volným pohybem těchto údajů a jejich správou, dále problematika mlčenlivosti. V souvislosti s principem autonomie je důležité uvědomit si, že respektování lidské autonomie je úzce spojeno s právem na lidskou důstojnost a svobodu.)
- Princip vysvětlitelnosti je prezentován následujícími požadavky:
  - transparentnost dat, datových souborů a procesů: důvody vytvoření určitého algoritmického rozhodnutí, identifikace jeho využití a důsledků
  - srozumitelnost rozhodovacích procesů systému: např. způsob fungování určitého algoritmického rozhodování, stanovení míry autonomního rozhodování
  - odpovědnost: stanovení konkrétní odpovědnosti za navržené algoritmické rozhodování, za stanovený stupeň autonomního rozhodování a za důsledky rozhodování autonomních a inteligentních systémů.

V praxi je nutné včas detekovat potenciální konflikty mezi jednotlivými etickými principy a náležitě je řešit individuálním vyvažováním na základě posouzení poměru risk/benefit nabízejících se postupů v každé konkrétní situaci. Přestože principy nemají hierarchický charakter, při řešení některých etických dilemat lze při jejich vyvažování vnímat, že některý z principů se v danou chvíli jeví jako významnější: například princip nonmaleficence bude nadřazen principu beneficence v situacích, které vyhodnotíme, že je důležité především nikdy neuškodit[2], když už v dané situaci nelze pomoci či prospět. Konfliktní, nebo dokonce antagonistické cíle a hodnoty vždy vyvolávají morální obavy, protože rozhodnutí, který cíl či která hodnota by měly být upřednostněny, se musí zvažovat velmi opatrně a individuálně a může přinést celou řadu následných otázek či pochybností. V bioetické praxi se nejčastěji dostávají do vzájemného střetu princip beneficence a princip autonomie. Princip autonomie je ve většině situací dominantnější, nejedná-li se o osobu nekompetentní autonomního rozhodování (tj. o nedostatek kognitivních schopností k individuální rozvaze a rozhodování), případně o neadekvátní či neoprávněné požadavky, nebo dokonce o ochranu veřejného zdraví či o ochranu zdraví třetích osob (Jedličková 2020). Analogicky lze tedy předpokládat, že rovněž při řešení etických dilemat autonomních a inteligentních systémů bude nejčastěji docházet ke konfliktu, ne-li dokonce k rozporu, právě mezi těmito dvěma etickými principy. V této souvislosti je nutné zdůraznit, že při konfrontaci a uplatňování etických principů v praxi je nutný jak racionální diskurs konkrétních okolností daného případu, tak důraz na důsledné respektování lidské důstojnosti.

Odborná skupina AI HLEG dále vypracovala následujících sedm požadavků, které je nutné dodržet při vývoji, zavádění a používání důvěryhodných systémů umělé inteligence:

1. Lidský faktor a dohled (např. podpora lidské autonomie a rozhodování, možnost zvrátit rozhodnutí učiněné umělou inteligencí či systém v určité situaci vůbec nepoužít)
2. Technická robustnost a bezpečnost (zahrnuje odolnost vůči útokům, zabezpečení, nouzový plán, bezpečnost, přesnost, spolehlivost a reprodukovatelnost)
3. Ochrana soukromí a správa dat (včetně kvality a integrity údajů a přístup k údajům)
4. Transparentnost (zahrnuje sledovatelnost, vysvětlitelnost a komunikaci)
5. Rozmanitost, nediskriminace a spravedlnost (zahrnuje předcházení nespravedlivé podjatosti, přístupnost, univerzální design a zapojení zúčastněných stran)
6. Dobré environmentální a sociální podmínky (zahrnuje

2 Z latiny známé Primum non nocere.

udržitelnost a šetrnost k životnímu prostředí, sociální dopad na společnost a demokracii)
7. Odpovědnost (zahrnuje auditovatelnost, minimalizaci negativních dopadů a podávání zpráv o těchto dopadech, možnost zjednání náprav) (High-Level Expert Group 2019).

Evropská komise iniciovala v dubnu 2018 vznik projektu Umělá inteligence pro Evropu, který má zajistit koordinovaný přístup členských států Evropské unie (EU) k co největšímu využití příležitostí plynoucích z umělé inteligence a k řešení nových výzev, které s tím souvisí. Cílem této iniciativy je výrazné posílení technických kapacit EU, nárůst využívání umělé inteligence, příprava na významné socioekonomické změny způsobené rozvojem umělé inteligence a zajištění příslušného právního a etického rámce pro související procesy (Evropská komise 2018a).

POŽADAVKY NA ETICKÉ PRINCIPY V MEZINÁRODNÍM MĚŘÍTKU

AI HLEG ve svých doporučeních vycházela také z dokumentů Globální iniciativy pro etiku autonomních a inteligentních systémů (Global Initiative on Ethics of Autonomous and Intelligent Systems), která působí v rámci mezinárodní instituce IEEE. Tato iniciativa si klade za cíl zajistit, že všechny zúčastněné strany podílející se na návrhu a vývoji autonomních a inteligentních systémů budou odborně vzdělávány, školeny a oprávněny upřednostňovat etické aspekty tak, aby moderní technologie mohly být využívány ve prospěch lidstva. Iniciativa proto vytvořila následující obecné zásady pro etický a důvěryhodný design při vývoji a provozování autonomních a inteligentních systémů:

- Lidská práva – A/IS budou vytvářena a provozována tak, aby respektovala, propagovala a chránila mezinárodně uznávaná lidská práva
- Well-being – vývojáři A/IS přijmou za primární kritérium úspěchu vývoje systému zvýšení lidské pohody a prosperity
- Správa dat – vývojáři A/IS umožní jednotlivci bezpečný přístup k osobním údajům, aby byla zachována možnost osob mít kontrolu nad vlastní identitou
- Účinnost – vývojáři a provozovatelé A/IS poskytnou důkaz o účinnosti a vhodnosti A/IS
- Transparentnost – základ konkrétního rozhodnutí A/IS by měl být vždy zjistitelný
- Odpovědnost – A/IS bude vytvořen a provozován tak, aby poskytoval jednoznačnost odůvodnění všech přijatých rozhodnutí
- Povědomí o zneužívání – vývojáři musí chránit A/IS před potenciálním zneužitím a riziky
- Kompetence – A/IS musí specifikovat a provozovatelé musí udržovat znalosti a dovednosti potřebné pro bezpečné a efektivní používání (IEEE Global Initiative 2019).

Také jiní autoři se věnují formulování morálních zásad a etických principů, jejichž prostřednictvím by bylo možné zajistit bezpečné a etické využívání technologií strojového učení a umělé inteligence v praxi. Uveďme například M. Petersona, který determinoval následujících pět morálních principů, které by měly tvořit základ při posuzování a rozhodování, zda je aplikace určitého technologického systému morálně správná:

- Princip poměru nákladů a přínosů (technologická intervence je morálně správná pouze tehdy, je-li čistý přebytek přínosů nad náklady u všech dotčených osob přinejmenším stejně velký jako u každé jiné alternativy)
- Princip předběžné opatrnosti (technologická intervence je morálně správná pouze tehdy, jsou-li přijata přiměřená předběžná opatření k ochraně před nejistými a nezanedbatelnými hrozbami)
- Princip udržitelnosti (technologická intervence je morálně správná pouze tehdy, nevede-li k žádnému významnému dlouhodobému vyčerpání přírodních, sociálních nebo ekonomických zdrojů)
- Princip autonomie (technologická intervence je morálně správná pouze tehdy, nesnižuje-li nezávislost, svrchovanost nebo svobodu lidí, kterých se určitá intervence týká)
- Princip spravedlnosti (technologická intervence je morálně správná pouze tehdy, nevede-li k nespravedlivým nerovnostem mezi lidmi).

Peterson ve své práci zpochybňuje způsobilost obecných etických teorií nalézt uspokojivou odpověď na konkrétní etická dilemata, a formuluje proto argumenty pro geometrickou konstrukci etických principů pomocí matematické teorie Voronoiových diagramů. Geometrické koncepty mohou být použity k vytvoření uvedených pěti principů jako abstraktních oblastí v morálním prostoru a autor se věnuje identifikaci vhodného způsobu interpretace dimenzí morálního prostoru spojených s geometrickou konstrukcí těchto principů a správnému umístění konkrétního případu v morálním prostoru. Tyto principy, jsou-li geometricky konstruovány pro konkrétní specifickou oblast (v našem případě se jedná o etiku technologií), jsou společně dostačující pro analýzu jakýchkoli současných etických problémů v relevantní oblasti, ačkoli autor připouští, že nebude-li jejich počet pro nějaké nové situace postačovat, může se podle potřeby jejich počet navyšovat (Peterson 2017). Návrh geometrické metody pro etické rozhodování je předmětem kritiky jiných autorů,[3] avšak Peterson ve své odpovědi na kritiku opakovaně zdůraznil, že použití geometrické metody umožňuje etikům vést diskuse o morálních principech způsoby, které byly dříve za hranicemi disciplíny, a zároveň vyvažovat případné konflikty mezi jednotlivými principy (Peterson 2018).

3 Viz např.: Lokhorst, G. J. C.: Review of Martin Peterson: The Ethics of Technology: A Geometric Analysis of Five Moral Principles (Lokhorst 2018), případně: Shrader-Frechette, K.: Review of Martin Peterson: The ethics of technology (Shrader-Frechette 2017).

POŽADAVKY NA ETICKÉ PRINCIPY V ČESKÉ REPUBLICE

Ministerstvo průmyslu a obchodu České republiky vydalo v květnu 2019 dokument s názvem Národní strategie umělé inteligence v České republice (NAIS). NAIS je součástí naplňování Inovační strategie České republiky 2019–2030 a přímo navazuje na iniciativy Evropské komise, především na Koordinovaný plán k umělé inteligenci, vydaný Evropskou komisí v prosinci 2018.[4] V souvislosti s etickými požadavky specifikuje NAIS pouze ve velmi obecné rovině cíle a nástroje na podporu výzkumu na vývoj odpovědné a důvěryhodné umělé inteligence se zaměřením na člověka a etické standardy. Mezi ně patří např. vytvoření jednotného systému pro vyhodnocování dopadů právních předpisů a etických pravidel a jejich adaptace či zřízení expertní platformy a fóra pro průběžnou revizi právních a etických pravidel umělé inteligence (Ministerstvo průmyslu a obchodu, 2019). Vytvoření etických kodexů pro jednotlivé sektory průmyslu, což je v dokumentu rovněž uvedeno mezi střednědobými cíli, však nelze považovat za účinný způsob prevence závažných etických implikací či zajištění fungujících etických standard při vývoji, zavádění a používání autonomních a inteligentních systémů. Praxe ukazuje, že pouhá existence etických kodexů je neúčinná a většinou se jedná o formalitu.

Z uvedeného přehledu vybraných etických pokynů je zřejmé, že experti různých profesí, kteří se věnují problematice etiky technologií a specifikují nároky na etické principy, jež je nutné v souvislosti s autonomními a inteligentními systémy dodržovat, stanovili velmi podobné, ne-li identické požadavky nehledě na geografickou lokalitu svého působení. Zároveň s požadavkem prospěchu a přínosu pro jedince a společnost či bezpečnosti a ochrany práv, svobod a soukromí (tj. principy beneficence/nonmaleficence a respektu k autonomii) akcentují ve větší či menší míře zejména spravedlnost, transparentnost a odpovědnost. Je pozitivní, že se skupiny odborníků v této oblasti snaží o vzájemnou spolupráci, je však nutné, aby příslušné etické požadavky vycházející z obecných etických principů byly co nejvíce konkretizovány, protože jen v konkrétní formě mohou být využitelné v praxi, a využití moderních technologií tak může být v souladu s individuálními i společenskými hodnotami a potřebami.

ETICKÉ IMPLIKACE VYUŽÍVÁNÍ AUTONOMNÍCH A INTELIGENTNÍCH SYSTÉMŮ

Možnosti uplatnění autonomních a inteligentních systémů v praxi se neustále rozšiřují, čímž přinášejí jak pozitivní, tak negativní dopady na různé oblasti života lidí, jejich blaho či prosperitu společnosti. Zvyšuje se tudíž rovněž riziko neetického provozování či využívání algoritmického rozhodování, a je tedy nutná etická reflexe se zaměřením na potenciální negativní aspekty. Lze vůbec účinně posuzovat, monitorovat, měřit a minimalizovat negativní dopady na lidské blaho, které v sobě zahrnuje celé spektrum osobních, sociálních a environmentálních faktorů?

Problematiku etiky umělé inteligence lze velmi obecně rozdělit na dvě základní skupiny. První se týká etických aspektů autonomních a inteligentních systémů při jejich vývoji, zavádění a používání v současnosti. Tedy, jakým způsobem člověk zachází s možnostmi, které nám umělá inteligence poskytuje, jak je umělá inteligence navržena, k čemu slouží, jak se využívají výsledky jejího algoritmického rozhodování, zda není zneužívána k neetickým účelům, nediskriminuje či nezpůsobuje újmy, jaká opatření jsou nezbytná k zajištění bezpečnosti, k bezpečnému uchování informací, k zabránění neoprávněnému přístupu, změnám či zveřejnění informací, k ochraně soukromí, svobod a práv,[5] zda je dodržována zásada transparentnosti, respekt k lidské důstojnosti apod. Všechny tyto etické aspekty lze ovlivnit rozhodnutím člověka, proto je nutné mít přesně vymezena pravidla pro každý krok od samotného návrhu a vývoje přes zavádění do praxe a používání až po účinnou kontrolu v průběhu všech fází „života" daného systému umělé inteligence. Jako příklad z praxe může posloužit zákaz používání vzdálené biometrické identifikace osob v reálném čase na veřejně přístupných místech pro účely prosazování práva (s určitými omezenými výjimkami), jak stanoví tzv. Akt o umělé inteligenci (Evropská komise 2021).

Druhou skupinu etických otázek, kterými se je nutné v oblasti etiky umělé inteligence také zabývat, představuje možný budoucí vývoj autonomních a inteligentních systémů a míra jejich autonomního rozhodování. V tomto článku se nebudeme zabývat variantami, při nichž se umělá inteligence stane záměrně zlovolnou či je předem naprogramovaná na ničení a zabíjení (například autonomní zbraně s rizikem ztráty kontroly člověka nad destruktivními činnostmi umělé inteligence k dosažení naprogramovaného cíle). Budeme se zamýšlet nad neetickými dopady prospěšně vytvořené a naprogramované umělé inteligence, u níž neplynou obavy ze zlovolnosti, ale z kompetencí, tedy nad potenciálními nezamýšlenými negativními dopady rozvoje autonomních a inteligentních systémů, které případně mohou také samostatně morálně rozhodovat, a nad potenciálními etickými důsledky takových rozhodnutí. Podstatnou součást diskusí o etice a bezpečnosti autonomních a inteligentních systémů tvoří také specifická témata jako využití elektronických sle-

4 Podle koordinovaného plánu je jednou z klíčových zásad pro umělou inteligenci vytvořenou v Evropě „etika již od návrhu", podle níž jsou etické zásady zakomponovány do výrobků a služeb umělé inteligence již na začátku procesu návrhu. Více viz (Evropská komise 2018b).

5 Například právo získat přístup k osobním údajům a k dalším informacím o zpracování, právo vznést námitku proti zpracování osobních údajů, právo na opravu nepřesných osobních údajů a doplnění neúplných osobních údajů, právo odvolat souhlas se zpracováním údajů, právo na výmaz osobních údajů neexistují-li žádné právní či oprávněné důvody pro zpracování či právo podat stížnost u dozorového orgánu. Více viz (Evropský parlament a Rada EU 2016).

dovacích zařízení, dronů či navržení etických principů pro algoritmické rozhodování technologie autonomního řízení vozidel, u nichž dosažená autonomizace a autonomnost vozidla představují na jedné straně účinný nástroj k významnému snížení počtu smrtelných dopravních nehod, na druhé straně však přinášejí zásadní morální výzvy. Jako příklad uveďme morální dilema autonomního vozidla, při němž se vozidlo před nadcházející neodvratnou kolizní situací bude muset rozhodnout pro výběr ze dvou negativních řešení: buď svým rozhodnutím poškodí či obětuje posádku vozidla, nebo jiné účastníky silničního provozu. Společensky odpovědnější přístup předpokládá utilitaristickou etiku, tj. příslušný algoritmus by byl založen na minimalizaci újmy, tedy na minimalizaci počtu obětí bez ohledu na to, zda se potenciální oběti nacházejí ve vozidle nebo mimo něj. Jednalo by se o vozidlo typu utilitarista a jak uvádí R. Kopecký, bylo by vybaveno řídícím softwarem počtář. Cílem jiného přístupu, tj. přístupu orientovaného na posádku ve vozidle, by byl algoritmus založen na ochraně posádky vozidla bez ohledu na počet obětí, tedy vozidlo typu egoista s řídícím softwarem tank. Dalším přístupem by mohlo být zachovat směr jízdy a aktivně ho nezměnit, hrozí-li ohrožení účastníků dopravní situace, kteří by jinak touto změnou ohroženi nebyli. Jde o typ ohleduplného vozidla s řídícím softwarem rytíř (Kopecký 2019). Popřípadě by si algoritmus autonomního vozidla v kolizních situacích mohl řidič předem nastavit podle své morální orientace. Nebo by snad bylo eticky korektnější upřednostnit k ochraně před kolizí konkrétní osoby podle určených kritérií (např. dítě, společensky či vědecky důležitou osobu, nebo naopak vystavit následkům kolize osobu nedodržující silniční předpisy, případně kriminálníka)? Má některý lidský život větší hodnotu než jiný? Nebo je přijatelnější určitá sofistikovaná kombinace uvedených strategií? V roce 2017 vypracovala etická komise pro autonomní řízení vozidel německého ministerstva dopravy základní etické zásady, kde je mezi jiným uvedeno, že ochrana jednotlivců má přednost před utilitárními hledisky (Di Fabio et al. 2017). Expertní skupina Evropské komise vydala v roce 2020 dokument s názvem Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility s dvaceti doporučeními pro etiku rozvoje datově propojených a autonomních vozidel v kontextu následujících tří hlavních oblastí:

- Silniční bezpečnost, rizika a dilemata
- Etika dat a algoritmů s ohledem na soukromí, spravedlnost a srozumitelnost
- Odpovědnost

Doporučení jsou založena na etickém rozdělení rizika mezi účastníky silničního provozu a jsou v souladu s pokyny AI HLEG pro důvěryhodnou umělou inteligenci (Horizon 2020). Mezi další etické aspekty technologie autonomního řízení vozidel patří například kybernetická rizika spojená s možnou chybou algoritmu při provozu autonomních vozidel, s výpadky systému nebo jeho částečnou nefunkčností v extrémních situacích, s kybernetickou kriminalitou spojenou se záměrným kybernetickým útokem za účelem převzetí kontroly nad řízením vozidla či jeho ovlivnění (Ministerstvo dopravy 2017) a v neposlední řadě také odpovědnost za vývoj konkrétního softwarového programu morálního rozhodování autonomních vozidel či kvalifikování právní odpovědnosti.

Současné právní předpisy týkající se odpovědnosti za způsobenou škodu však nejsou přizpůsobeny pro řešení nároků na náhradu škody způsobenou produkty či službami založenými na umělé inteligenci, a to zejména v oblasti prokazování protiprávního jednání či opomenutí osoby, která škodu způsobila. Specifické vlastnosti autonomních a inteligentních systémů, včetně autonomie a neprůhlednosti (tzv. efekt „černé skříňky"[6]), ztěžují identifikaci odpovědné osoby a prokázání podmínek pro úspěšné uplatnění nároku na náhradu škody. Ke snížení právní nejistoty týkající se odpovědnosti a k posílení důvěryhodnosti k technologiím využívajícím umělou inteligenci vydala Evropská komise v září 2022 návrh směrnice Evropského parlamentu a Rady o přizpůsobení pravidel mimosmluvní občanskoprávní odpovědnosti umělé inteligenci (směrnice o odpovědnosti za umělou inteligenci), která se zaměřuje na poskytnutí stejné úrovně ochrany obětem poškozeným systémy umělé inteligence, jako by byly poškozeny za jakýchkoli jiných okolností či jinými technologiemi, a zmírňuje důkazní břemeno obětí zavedením „domněnky příčinné souvislosti". V praxi to bude znamenat, že mohou-li poškozené oběti prokázat nedodržení určité povinnosti, jež se vztahuje k jejich újmě, a existenci přiměřené pravděpodobnosti příčinné souvislosti s výkonem systému založeném na umělé inteligenci, může soud předpokládat, že tímto nedodržením povinnosti jakékoli osoby (vývojář, provozovatel, uživatel) byla způsobena škoda (Evropská komise 2022). Vývoj, zavádění a používání umělé inteligence, robotiky a souvisejících technologií včetně softwaru, algoritmů a dat, které tyto technologie používají či produkují, nesmí záměrně působit nebo již od návrhu vědomě přijímat újmu jednotlivcům či společnosti. Každý by měl proto v souladu s právními předpisy týkajícími se odpovědnosti, s etickými zásadami a v míře, v níž pracuje s autonomními a inteligentními systémy, odpovídat za veškerou újmu způsobenou jednotlivcům nebo společnosti. Důležité je také stanovení konkrétní odpovědnosti za přijetí odpovídajících opatření k zamezení újmy a jejich důsledné dodržování, jak je stanoveno v dokumentu s názvem Rámec pro etické aspekty umělé inteligence, robotiky a souvisejících technologií, který byl přijat Evropským parlamentem v říjnu 2020 (Evropský parlament 2020).

6 Etický princip vysvětlitelnosti má pro budování a udržení důvěry uživatelů v systémy umělé inteligence zásadní význam. Vysvětlení, proč autonomní a inteligentní systém vytvořil určité rozhodnutí nebo výstup, případně jaká kombinace vstupních faktorů k nim přispěla, není vždy možné. Tyto případy se nazývají „černé skříňky" (High-Level Expert Group, 2019).

ZÁVĚR

Nebudou-li autonomní a inteligentní systémy disponovat schopností morálně a eticky řešit etická dilemata, nelze je nechat autonomně rozhodovat bez kontroly člověka. Skutečná podstata využívání autonomních a inteligentních systémů v praxi by měla být postavena na zásadě sloužit člověku s cílem zlepšit způsob života lidí či jejich bezpečnost, například zdokonalováním výrobních procesů, řešením komplikovaných situací, syntézou nových léčiv, zajištěním bezpečnější dopravy či zvýšením přesnosti chirurgických zákroků.

Budoucí zaměření technologického vývoje předpokládá, že umělá inteligence má potenciál stát se inteligentnějším než člověk. My bychom však v takovém případě nedisponovali žádnými spolehlivými možnostmi, jak anticipovat její následný autonomní vývoj, motivace a cíle a s tím související jednání k člověku. Bude nám chtít pomáhat nebo nás chtít přelstít, či dokonce ovládat? Pakliže již nebudeme nejinteligentnější, máme jistotu, že budeme mít algoritmická rozhodování těchto autonomních inteligentních systémů stále pod kontrolou? Kdo ponese odpovědnost za jejich autonomní rozhodování? Future of Life Institute (FLI) zastává názor, že rostoucí sílu technologií umělé inteligence můžeme přemoci moudrostí, s níž ji zvládáme, a to prostřednictvím podpory výzkumu bezpečnosti umělé inteligence. Některé výzkumné projekty v této oblasti FLI také financuje (Future of Life 2021).

V oblasti etiky umělé inteligence je potřebné včas vytvořit etický rámec s konkrétními postupy a spolehlivou metodikou, jednoznačně specifikovat priority a usilovat o minimalizaci nepříznivých důsledků, a to za spolupráce odborníků z různých vědních disciplín. Formovat etický rámec je nutné prostřednictvím aplikace filosoficko-etické metodologie[7] konvenčních i novějších etických teorií, tedy prostřednictvím aplikované etiky, také uplatňováním etických principů, a to vše v souladu s příslušnými hodnotami lidství a s respektem k lidské důstojnosti. Při jejich tvorbě nepostačují pouze aspekty pragmatického a deskriptivního charakteru či racionální pravidla logiky, navíc je nutné společně s legalitou uplatňovat také legitimnost, tj. oprávněnost záměru, korektnost cíle, odůvodněnost postupů, ohled na humánnost, soulad s lidskými hodnotami. Stojíme před mnoha potenciálními situacemi, které předem neumíme detekovat, nelze tedy spoléhat na univerzální popisy a postupy, byť budou pečlivě připraveny odborníky. Etika je filosofická disciplína a filosofie není vědou zabývající se popisem. Podstatou filosofického zkoumání je poznání světa či smyslu lidského bytí. Při detekci správnosti cíle konkrétního poznávání a vhodné cesty je nutné především vnímat lidskost, bytostně se tázat a odhalovat vhled do podstaty, odkud k nám vyplouvají etické hodnoty. Jinak nelze podstatu pochopit a zůstaneme u popisu. Účelem vytvoření etického rámce ovšem není vytvoření vědecké deskripce, ale hluboké pochopení smyslu a poslání budoucího rozvoje umělé inteligence. A to nelze prostřednictvím předem připravených univerzálních postupů a procedur, nebo dokonce etických kodexů, vytvořených a schválených různorodými institucemi a komisemi. Je pozitivní, že odborný diskurs o reálné podobě zajišťování kybernetické bezpečnosti z etického pohledu intenzivně probíhá, čímž je zajištěna možnost významně ovlivnit formování výsledku rozvoje umělé inteligence v obecné rovině. V kontextu výše uvedeného je však potřeba dbát na riziko nadbytečného zobecňování a vytváření pravidel, která jsou v praxi neuchopitelná. V jednotlivých projektech je zapojení etiků do týmů příslušných expertů při vytváření pravidel algoritmického rozhodování autonomních a inteligentních systémů čím dále častější, přestože je tato praxe leckdy ještě stále považována za nutné zlo či formalitu. Začlenění etiků do odborných týmů v konkrétních projektech je zásadní. Z vlastní praxe s aktuálně probíhajícími výzkumnými projekty však lze potvrdit, že s etiky probíhají konstruktivní diskuse o relevantních etických aspektech a konkrétní etické požadavky jsou při návrhu a vývoji autonomních a inteligentních systémů respektovány.

7 Tradičními filosoficko-etickými metodologickými východisky řešení etických dilemat v jednotlivých oblastech aplikované etiky jsou deontologická etika, utilitaristická etika (se svými variantami konsekvencialismu) a také etika ctností.

LITERATURA

Beauchamp, Tom L. – Childress, James F. (2019): *Principles of Biomedical Ethics*. 8th edition. New York: Oxford University Press.

Bostrom, Nick (2014): *SUPERINTELLIGENCE: Paths, Dangers, Strategies*. New York: Oxford University Press.

Di Fabio, Udo – Broy, Manfred – Brüngger, Renata J. (2017): *Ethics Commission. Automated and Connected Driving*. Federal Ministry of Transport and Digital Infrastructure of the Federal Republic of Germany. (online). https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

Evropská komise (2018a): *Umělá inteligence pro Evropu*. SDĚLENÍ KOMISE EVROPSKÉMU PARLAMENTU, RADĚ, EVROPSKÉMU HOSPODÁŘSKÉMU A SOCIÁLNÍMU VÝBORU A VÝBORU REGIONŮ, COM/2018/237. Brusel. (online). https://www.vlada.cz/assets/evropske-zalezitosti/umela-inteligence/Sdeleni-EK-k-AI.PDF

Evropská komise (2018b): *Koordinovaný plán v oblasti umělé inteligence*. SDĚLENÍ KOMISE EVROPSKÉMU PARLAMENTU, EVROPSKÉ RADĚ, RADĚ, EVROPSKÉMU HOSPODÁŘSKÉMU A SOCIÁLNÍMU VÝBORU A VÝBORU REGIONŮ. Brusel. (online). https://www.vlada.cz/assets/evropske-zalezitosti/umela-inteligence/Koordinovany-plan-k-AI_1.pdf

Evropská komise (2021): *Návrh NAŘÍZENÍ EVROPSKÉHO PARLAMENTU A RADY, KTERÝM SE STANOVÍ HARMONIZOVANÁ PRAVIDLA PRO UMĚLOU INTELIGENCI (AKT O UMĚLÉ INTELIGENCI) A MĚNÍ URČITÉ LEGISLATIVNÍ AKTY UNIE*. COM/2021/206. Brusel. (online). https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0002.02/DOC_1&format=PDF

Evropská komise (2022): *Návrh směrnice Evropského parlamentu a Rady o přizpůsobení pravidel mimosmluvní občanskoprávní odpovědnosti umělé inteligenci (směrnice o odpovědnosti za umělou inteligenci)*. COM/2022/496. Brusel. (online). https://ec.europa.eu/info/sites/default/files/1_1_197605_prop_dir_ai_en.pdf

Evropský parlament a Rada EU (2016): NAŘÍZENÍ EVROPSKÉHO PARLAMENTU A RADY 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (obecné nařízení o ochraně osobních údajů). In: *Úřední věstník Evropské unie*. (online). https://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:32016R0679&from=EN

Evropský parlament (2020): *Rámec pro etické aspekty umělé inteligence, robotiky a souvisejících technologií.* (online). https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_CS.pdf

Future of Life Institute (2021): *Benefits & Risks of Artificial Intelligence*. (online). https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/

High-Level Expert Group on Artificial Intelligence (2019): *Etické pokyny pro zajištění důvěryhodnosti UI.* (online). https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_CS.pdf

Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (2020): *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility*. Publication Office of the European Union. (online). https://op.europa.eu/en/publication-detail/-/publication/89624e2c--f98c-11ea-b44f-01aa75ed71a1/language-en/format-PDF/source--search

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019): *Ethically Aligned Design: A Vision for Prioritizing Human Well--being with Autonomous and Intelligent Systems*. 1st edition. (online). https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf?utm_medium=alias&utm_source=-LI&utm_campaign=EAD1e&utm_content=report&utm_term=undefined

Jedličková, Anetta (2020): Etické konotace léčby onemocnění covid-19. *Vnitřní Lék.*, 66(7), 8–12. doi: 10.36290/vnl.2020.132

Kopecký, Robin (2019): Morální problémy autonomních vozidel. *Filosofický časopis*, 67(2), 263–276.

Lokhorst, Gert-Jan C. (2018): Review of Martin Peterson: The Ethics of Technology: A Geometric Analysis of Five Moral Principles. *Sci Eng Ethics,* 24,1641–1643. (online). https://doi.org/10.1007/s11948-017-0014-0

Ministerstvo dopravy (2017): *VIZE ROZVOJE AUTONOMNÍ MOBILITY*. (online). https://www.databaze-strategie.cz/cz/md/strategie/vize-rozvoje-autonomni-mobility?typ=download

Ministerstvo průmyslu a obchodu (2019): *Národní strategie umělé inteligence v České republice.* (online). https://www.vlada.cz/assets/evropske-zalezitosti/umela-inteligence/NAIS_kveten_2019.pdf

Peterson, Martin (2017): *The Ethics of Technology: A Geometric Analysis of Five Moral Principles*. New York: Oxford University Press

Peterson, Martin (2018): The Ethics of Technology: Response to Critics. *Sci Eng Ethics,* 24, 1645–1652. (online). https://doi.org/10.1007/s11948-018-0062-0

Shrader-Frechette, Kristin (2017): Review of Martin Peterson: The ethics of technology. *Notre Dame Philosophical Reviews*. (online). http://ndpr.nd.edu/news/the-ethics-of-technology-a-geometric-analysis-of-five--moral-principles/