

## Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems

### Vyjádření k umělé inteligenci, robotice a „autonomním“ systémům

Brusel, 9. března 2018

<https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-84689254>

Pracovní překlad.

Z angličtiny přeložil Boleslav Vraný, [bolek@bolekvrany.cz](mailto:bolek@bolekvrany.cz), 16. března 2020

# Shrnutí

Pokroky v umělé inteligenci, robotice a tzv. „autonomních“<sup>1</sup> technologiích vedly k řadě stále urgentnějších a složitějších morálních otázek. Současné úsilí k nalezení odpovědí na jimi přinášené etické, společenské a legální výzvy a k jejich zaměření na společné dobro představuje slepenec nesourodých iniciativ. To podtrhuje potřebu společného, široce zaměřeného a inkluzivního procesu reflexe a dialogu, dialogu, který se zaměřuje na hodnoty, okolo kterých chceme organizovat společnost a na roli, které by v této společnosti měly hrát technologie.

Toto vyjádření volá po spuštění procesu, který by vystavěl cestu ke společnému, mezinárodně uznávanému etickému a legálnímu rámci pro návrh, výrobu, použití a správu umělé inteligence, robotiky a „autonomních“ systémů. Toto vyjádření také navrhuje skupinu základních etických principů, založených na hodnotách zakotvených ve smlouvách EU a v Evropské chartě základních práv, které mohou vést tento vývoj.

---

1 Toto vyjádření se vztahuje ke skupině chytrých digitálních technologií, které rychle konvergují a často jsou vzájemně provázány, propojeny nebo plně integrovány, např. Klasická umělá inteligence, algoritmy strojového učení, hluboké učení a konektivních sítí, generative adversarial networks, mechatronika a robotika. Samoříditelná auta, robotické zbraňové systémy, chatboty a systémy rozpoznávající řeč a obraz jsou některé z dobře známých příkladů kombinace takových technologií.

# Pozadí

První dvě dekády 21. století přinesly pozoruhodné příklady toho, čemu se se obecně říká „autonomní technologie“ a „umělá inteligence.“ Samoříditelná auta a drony, roboty v průzkumu hlubokých moří a vesmíru, zbraňové systémy, softwaroví agenti, jako boty na finančních trzích, a hluboké učení v oblasti lékařských diagnóz patří mezi nejnápadnější příklady, ale rozhodně nejde o příklady jediné. Významnými hnacími mechanismy tu je umělá inteligence (AI), zejména v podobě strojového učení, a rostoucí dostupnost velkých skupin dat z různých oblastí života. Souběh těchto digitálních technologií je rychle dělá výkonnějšími, jsou aplikovány ve stále větším počtu nových produktů a služeb, ve veřejném i soukromém sektoru, a mohou mít vojenské i civilní použití. AI schovaná v těchto systémech může předefinovat práci nebo zlepšit pracovní podmínky lidí a snížit potřebu lidského příspěví, vstupu a zasahování do provozu. Může lidem pomoci nebo nahradit lidi chytrou technologií ve složitých, špinavých, jednotvárných nebo nebezpečných pracích i dále.

Chytré systémy dnes bez lidského zásahu a řízení zvenčí provádí rozhovory se zákazníky v on-line call centrech, přesně a nepřetržitě řídí robotické ruce a manipulují s objekty, během milisekund ve velkém nakupují a prodávají akcie, řídí auta, aby zahýbala nebo brzdila a zabránila srážce, klasifikují lidi a jejich chování nebo dávají pokuty.

Je nešťastné, že některé z těch nejmocnějších kognitivních nástrojů jsou zároveň ty nejméně průhledné. Jejich akce již nejsou lineárně programovány lidmi. Google Brain vyvíjí AI, která údajně staví další AI lépe a rychleji, než to dovedou lidé. AlphaZero se za čtyři hodiny dokáže naučit šachy od úrovně, kdy je nezná vůbec, na úroveň velmistra. Nelze pochopit, jak přesně AlphaGo dokázala porazit lidského Go World šampióna. Hluboké učení a přístupy založené na „generative adversarial network“ (GAN) dovolují strojům, aby se samy učily nové strategie a vyhledávaly nové údaje k analýze. V tomto smyslu už jejich akce nejsou srozumitelné a otevřené podrobnému zkoumání lidmi. Tak je tomu proto, že jednak nelze určit, jak k výsledku dospěly, kromě znalosti původních algoritmů. Za druhé je jejich chování založeno na datech, na kterých se učily a která už ani nemusí být k dispozici. Takže šum a chyby v datech, na kterých se učily v minulosti, se hluboce zakořenily do systému.

Systémy, které se naučí vykonávat tyto úlohy bez lidského vedení nebo dozoru, se dnes často nazývají „autonomní.“ Tyto tzv. „autonomní“ systémy se mohou projevovat jako high-tech robotické systémy nebo jako inteligentní software jako např. boty. Mnohé z nich jsou uvolněny do světa bez lidského dohledu a mohou dosáhnout věcí, které nebyly předvídaný jejich lidskými tvůrci nebo vlastníky.

Proto vidíme tyto relevantní vývoje v oblasti technologie

1. AI ve formě strojového učení, zejména hlubokého učení, krmeného BigData se rychle stává čím dál výkonnější. Je aplikována ve stále větším množství nových digitálních produktů a služeb ve veřejném i soukromém sektoru a může mít vojenské i civilní využití. Jak už bylo řečeno, vnitřní fungování AI může být velice těžko, pokud vůbec, sledovatelné, vysvětlitelné a kriticky zkoumatelné. Tyto pokročilé schopnosti se z velké části hromadí u soukromých subjektů a jsou z velké části proprietární.
2. Pokročilá mechatronika (kombinace AI, hlubokého učení, datové vědy, senzorové technologie, internetu věcí, mechanického a elektrického inženýrství) přináší širokou paletu stále sofistikovanějších robotických a high-tech systémů pro praktické aplikace ve službách, výrobě, zdravotní péči, maloobchodu, logistice, domotice (automatizaci domácností) a

bezpečnosti (security i safety). Ve veřejných debatách vyčnívají dvě aplikační domény, a to robotické zbraňové systémy a „autonomní“ vozidla.

3. Jsou vyráběny stále chytřejší systémy, které vykazují velkou dávku toho, čemu se často říká „autonomie“, což znamená, že se vyvíjejí a umí vykonávat úlohy bez lidských operátorů a jejich dohledu.
4. Zdá se, že existuje tlak na stále větší úroveň automatizace a „autonomie“ v robotice, AI a mechatronice. Investice států a velkých společností do této oblasti jsou enormní a vedoucí úloha v AI patří mezi hlavní cíle světových supervelmocí.
5. Existuje vývoj směrem ke stále těsnější spolupráci mezi lidmi a stroji (co-boti, cyber-posádky, digitální dvojčata a dokonce integrace chytrých mašin do lidského těla v podobě rozhraní počítač-mozek nebo kyborgů). Dobře sehrané týmy AI a lidských profesionálů v některých oblastech dosahují lepších výsledků než lidé nebo stroje samostatně.

# Morální reflexe

## Klíčové otázky

Nástup high-tech systémů a softwaru, které mohou fungovat stále více nezávisle na lidech a mohou vykonávat úlohy, které, pokud by je vykonávali lidé, by vyžadovaly inteligenci, vyžaduje další zvláštní rozlišování. Tyto systémy dávají povstat celé řadě důležitých a složitých morálních otázek.

Za prvé jsou to otázky o bezpečnosti a zabezpečení, prevenci ublížení a zmírnění rizik. Jak můžeme učinit svět s propojenou AI a „autonomními“ zařízeními bezpečný a zabezpečený a jak můžeme tato rizika měřit a posoudit?

Za druhé jsou tu otázky po lidské mravní odpovědnosti. Kde se v dynamických a složitých socio-technických systémech s pokročilou AI a robotickými komponentami nachází mravně relevantní schopnost rozhodování? Jak by měla být přiřazena a rozdělena mravní odpovědnost a kdo a v jakém smyslu je zodpovědný za nešťastné, nečekané výsledky? Má smysl hovořit o „sdíleném řízení“ a „sdílené odpovědnosti“ mezi lidmi a chytrými stroji? Budou lidé částmi ekosystému „autonomních“ zařízení jako morální „deformační zóny,“ vložené jen proto, aby absorbovaly odpovědnost, nebo budou mít dobrou možnost nést odpovědnost za to, co dělají?

Za třetí vznikají o správě, regulaci, návrhu, vývoji, inspekci, dohledu, testování a certifikaci. Jak by měly být naše instituce a zákony transformovány, aby sloužily dobru jedinců a společnosti a aby učinily společnost bezpečnou pro tuto technologii?

Za čtvrté, jsou tu otázky týkající se demokratického rozhodování, včetně rozhodování o institucích, zásadách a hodnotách, které tvoří základy pro otázky shora uvedené. Po celém světě se konají vyšetřování, aby se zjistila míra, ve kterém jsou občané zneužívání s využitím pokročilých technik pozitivní zpětné vazby a nepřímých doporučení (angl. nudging), založených na strojovém učení, big data a behaviorální vědě, které umožňují jemné profilování, mikro-cílení, šití na míru a manipulaci toho, jak jsou jim prezentovány možnosti výběru, s ohledem na komerční nebo politické cíle.

Nakonec jsou zde otázky po vysvětlitelnosti a transparentnosti AI a „autonomních“ systémů. Jakým hodnotám tyto systémy efektivně a demonstratelně slouží? Na základě jakých hodnot je navrhujeme? Okolo jakých hodnot chceme organizovat společnost? A u kterých hodnot dovolujeme, otevřeně nebo skrytě, aby byly podkopány technologickým pokrokem a utilitárním trade-offs? „Optimalizace“ společenských procesů za pomoci AI, založená na systémech sociálního skóre, se kterými experimentují některé země, porušuje základní myšlenku rovnosti a svobody stejně jako kastovní systémy, protože vytváří „různé druhy lidí“ tam, kde ve skutečnosti jsou jen jiné „vlastnosti lidí.“ Jak může být odvrácen útok na demokratické systémy a využití skórovacích systémů jako základu dominance těch, kde mají přístup k těmto mocným technologiím?

## Klíčové úvahy

Z etického úhlu pohledu je důležité mít na paměti, že:

Termín „autonomie“ vychází z filosofie a odkazuje na schopnost lidských osob vydávat pro sebe zákony, formulovat, promýšlet a volit si normy, pravidla a zákony, kterými se budou řídit. Zahrnuje právo svobodně nastavit své standardy a zvolit si své životní cíle a účely (???). Kognitivní procesy, které toto podporují a ulehčují, patří mezi ty nejvíce spojované s důstojností člověka a lidskou schopností konat *par excellence*. Obvykle s sebou nesou prvky uvědomování si a vědomí sebe sama a formování sebe sama podle rozumu a hodnot. Autonomie v eticky relevantním smyslu slova proto

může být připsána jen lidským bytostem. Proto je označování pouhých lidských výtvorů, byť velmi pokročilých, komplexních adaptivních nebo dokonce „inteligentních“ systémů, za „autonomní“, poněkud nešťastné. Terminologie „autonomních“ systémů se ale široce rozšířila ve vědecké literatuře i ve veřejné debatě k označení nejvyššího stupně automatizace a nejvyššího stupně nezávislosti na lidských bytostech ve smyslu operační a rozhodovací „autonomie.“ Ale autonomie ve svém původním významu je důležitá stránka lidské důstojnosti, která by neměla být relativizována.

Protože žádný lidský výtvor nebo systém – jakkoliv pokročilý a sofistikovaný – nemůže být sám o sobě nazýván „autonomním“ v původním, etickém významu slova, nemůže mu být poskytnuto mravní postavení lidské osoby a nemůže zdědit lidskou důstojnost. Lidská důstojnost jako základ pro lidská práva implikuje, že musí být možná smysluplná lidská intervence ve věcech, které se týkají lidských bytostí a jejich prostředí. Proto, v kontrastu k automatizaci výroby, není náležité řídit a rozhodovat o lidech tím způsobem, jakým rozhodujeme o předmětech nebo datech, a to ani tehdy, když je to technicky představitelné. Takový „autonomní“ management lidských bytostí by byl neetický a podkopával by hluboce zakořeněné základní evropské hodnoty. Lidské bytosti by měly být schopné určit (nebo zjistit – oba překlady možné), kterým hodnotám technologie slouží, co je mravně relevantní a o které konečné cíle a koncepce dobra stojí za to usilovat. To nemůže být ponecháno sebevýkonnějším strojům.

Schopnost a ochota brát a přisuzovat mravní odpovědnost je integrální součástí konceptu osoby, na kterém stojí všechny naše mravní, společenské a právní instituce. Mravní odpovědnost je zde chápána v širokém smyslu, ve kterém může odkazovat na několik aspektů lidské schopnosti konat, např. kauzalitu, zodpovědnost ve smyslu povinnosti poskytnout příčinu, pohnutku, odpovědnost ve smyslu povinnosti kompenzovat škody, reaktivní postoje jako pochvala a vina (vhodnost celé řady mravních citů), a povinností spojených se společenskými rolmi. Mravní odpovědnost, v jakémkoliv významu, nemůže být přidělena „autonomní“ technologii nebo na ni být přesunuta.

V nedávných debatách o smrtících autonomních zbraňových systémech (Lethal Autonomous Weapons Systems, LAWS) a autonomních autech se zdá existovat široký konsenzus, že Smysluplná Lidská Kontrola (Meaningful Human Control, MHC) je základní pro mravní odpovědnost. Princip MHC byl poprvé navržen pro omezení vývoje a využití budoucích zbraňových systémů. Znamená, že lidé – a ne počítače a jejich algoritmy – mají v konečném důsledku zůstat u řízení a tak být morálně zodpovědní.<sup>2</sup>

## Za úzce etický rámec

Dvě oblasti vývoje „autonomních“ systémů už vedly k vysokoprofilovým etickým debatám, a to smrtící autonomní zbraňové systémy (LAWS) a samořiditelná auta. Ačkoli zcela samořiditelná auta ještě nejsou na trhu, některé země se už připravují na legislativní možnost zavedení takových aut na veřejných cestách. Morální debata byla rozvířena v roce 2016, když první člověk zahynul pod koly auta v „autonomním“ režimu. Morální debaty jsou dnes většinou omezeny na diskuzi výjimečných případů s použitím myšlenkových experimentů typu Tramvajové dilema. Ty se zabývají dilematy neodvratitelných nehod, kdy jediný výběr je mezi možnostmi zahrnujícími ztráty na životech. Tato úzká konstrukce etických problémů vede k početnému přístupu a často příliš zjednodušující metrice v lidských záležitostech. V tomto rámci se ústřední otázky týkají odpovědnosti „autonomních“ systémů, jejich dopadů a jak by měly být naprogramovány, aby jejich nasazení vedlo k morálně akceptovatelným výsledkům se smyslu ztracených, resp. zachráněných životů. To ale zanedbává

---

<sup>2</sup> NGO článek 36, 2015

hraniční otázky jako „jaké design decisions byly přijaty v minulosti, které nás přivedly k tomuto morálnímu dilematu,“ „které hodnoty by měly mít vliv na design“, „jak by měly být hodnoty v návrhu váženy v případě konfliktu a kým“, „jaký je status obrovské empirické evidence týkající se toho, jak se lidé skutečně rozhodují v případě Tramvajového dilematu a jak by měla být přenesena na automatizovaná vozidla?“

Druhým polem kontroverze jsou „autonomní“ zbraňové systémy. Tyto vojenské systémy nesou smrtící zbraně, ale z hlediska softwaru nejsou příliš odlišné od „autonomních“ systémů, které můžeme najít v řadě civilních aplikací blíže domovu. Velká část této debaty se odehrává v rámci Conference on Certain Conventional Weapons v Ženevě a týká se morální přípustnosti takových systémů a legální a morální odpovědnosti za jejich nasazení. Nyní se pozornost musí nasměřovat na otázky ohledně povahy a významu „Meaningful human control“ nad těmito systémy a jak zavést morálně žádoucí formy kontroly.

Třetí oblastí aplikací je „autonomní“ software včetně botů. Obchod, finance a akciové trhy jsou z velké části řízeny algoritmy a sw. Chytré systémy dnes bez lidské intervence a dohledu zvenčí provádějí dialogy se zákazníky v on-line call centrech; systémy rozpoznávání řeči a doporučující systémy online platforem, např. Siri, Alexa, Cortana, dávají doporučení uživatelům. Kromě přímočarých otázek po ochraně dat a soukromí se můžeme ptát, zda lidé mají právo vědět, zda interagují s člověkem nebo AI strojem. Dále vzniká otázka, zda by měly být limity v tom, co AI systém může osobě navrhnout, na základě konstrukce, jak daná osoba sama chápe svoji identitu  
**(překlad???)**

Přestože roste vědomí potřeby věnovat se těmto otázkám, pokrok v AI a robotice jde rychleji než proces nacházení odpovědí na tyto trnitě etické, legální a společenské otázky. Současné úsilí představuje mozaiku nesourodých iniciativ. Je tu jasná potřeba kolektivního, širokého a inkluzivního procesu, který by dláždil cestu ke společnému, mezinárodně uznávanému etickému rámci pro návrh, produkci, použití a správu AI, robotů a „autonomních“ systémů.

Toto prohlášení volá po zahájení takového procesu a navrhuje množinu fundamentálních etických principů a demokratických předpokladů, které mohou také vést reflexi patřičných zákonů. EGE je toho názoru, že Evropa by v tom měla hrát aktivní a prominentní roli. Dohlížejíc na debaty o morální odpovědnosti AI a tzv. „autonomních“ systémů, EGE volá po více systematickém přemýšlení a výzkumu ohledně etických, legálních a správních aspektů high-tech systémů které mohou konat ve světě bez přímé kontroly lidských uživatelů, ku prospěchu či neprospěchu lidstva. To je velmi urgentní věcí.

# Směrem ke sdílenému etickému rámci pro umělou inteligenci, robotiku a „autonomní“ systémy

Některé z nejvíce prominentních iniciativ k formulování etických principů ohledně AI a „autonomních“ systémů vzešly z průmyslu, od praktiků a profesních skupin, jako politický dokument IEEE (Institute of Electrical and Electronics Engineers) „Ethically Aligned Design“,<sup>3</sup> globální summit ITU (International Telecommunication Union) „AI for Good“<sup>4</sup> v létě 2017 a práce ACM (Association for Computing Machinery) na tomto tématu, včetně velké AAAI/ACM konference o AI, etice a společnosti<sup>5</sup> v únoru 2018. V soukromém sektoru společnosti jako IBM, Microsoft a Google DeepMind vytvořily své vlastní etické zásady pro AI a spojily síly k vytvoření širších iniciativ jako „Partnership on AI“<sup>6</sup> nebo „OpenAI“,<sup>7</sup> které spojují průmysl, neziskové a akademické organizace.

Jedna z vedoucích iniciativ volajících po zodpovědném vývoji AI byla spuštěna Future of Life Institute a vyvrcholila vytvořením „Asilomar AI Principles.“ Tento seznam 23 základních principů vedoucích výzkum a aplikace AI byl podepsán stovkami účastníků,<sup>8</sup> kde signatáři reprezentovali převážně vědce, výzkumníky v oboru AI a průmysl. Podobný participativní proces byl spuštěn po iniciativě Fóra pro společensky odpovědný vývoj umělé inteligence, které proběhlo na Univerzitě v Montrealu v listopadu 2017 v reakci na první návrh možné „Deklarace pro zodpovědný vývoj umělé inteligence.“ Nyní je veřejně dostupné jako on-line platforma, kde jsou všechny části společnosti zvány, aby text komentovaly.<sup>9</sup>

Celosvětová debata o vojenském využití AI byla zahájena OSN a na setkáních Convention on Certain Conventional Weapons (CCW) v Ženevě, kde několik významných stran podpořilo princip „smysluplné lidské kontroly na LAWS.“ Říkájí: „Autonomní zbraňové systémy, které nevyžadují žádnou smysluplnou/významnou lidskou kontrolu, mají být zakázány“ (Valné shromáždění OSN, 2016). OSN také vytvořila speciální výzkumný institut v Haagu pro studium správy robotiky a AI (UNICRI).<sup>10</sup> Za zákaz „autonomních“ zbraní bojuje i několik iniciativ a nevládních organizací, které se snaží o použití AI a „autonomních“ systémů pro „dobro“ např. Foundation for Responsible Robotics.

Iniciativy na národní úrovni jsou mezitím nestejně, když některé země dávají priority vývoji pravidel pro robotiku a AI a jdou až k přijetí legislativy (např. pro regulace samořiditelných aut na veřejných cestách), zatímco jiné země se ještě musí s tímto tématem poprat. Tento nedostatek harmonizovaného evropského přístupu vedl Evropský parlament, aby volal po různých prostředcích,

---

<sup>3</sup> [http://standards.ieee.org/news/2016/ethically\\_aligned\\_design.html](http://standards.ieee.org/news/2016/ethically_aligned_design.html)

<sup>4</sup> <https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx>

<sup>5</sup> <http://www.aies-conference.com/>

<sup>6</sup> <https://www.partnershiponai.org/>

<sup>7</sup> <https://openai.com/>

<sup>8</sup> <https://futureoflife.org/ai-principles/>

<sup>9</sup> <http://nouvelles.umontreal.ca/en/article/2017/11/03/montreal-declaration-for-a-responsible-development-of-artificial-intelligence/>

<sup>10</sup> V tomto směru také stojí za pozornost, pod záštitou UNESCO: „COMEST Report on robotics ethics“ a „IBC Report on big data and health,“ obě přijaté v září 2017

jak připravit regulaci pokročilé robotiky,<sup>11</sup> včetně vývoje směr udávajících etických rámců pro návrh, výrobu a použití robotů.

Na tomto pozadí EGE upozorňuje na rizika nekoordinovaných, nevyvážených přístupů k regulaci AI a „autonomních“ technologií. Regulatorní mozaiky mohou vést k „ethics shopping“, vedoucímu k přemístění vývoje a použití AI do oblastí s nižšími etickými standardy. Pokud bychom dovolili, aby debata byla dominována některými regiony, disciplínami, demografiemi (???) nebo průmyslovými aktéry, riskujeme, že budou vyloučeny širší společenské zájmy a perspektivy. Současným diskusím také někdy chybí přehled „autonomních“ technologií, které pravděpodobně budou studovány, vyvíjeny a implementovány v příští dekádě, což ponechává slepou skvrnu ohledně regulatorní předvídatosti.

---

<sup>11</sup> European Parliament, Committee on Legal Affairs 2015/2103 (INL) Report with Recommendations to the Commission on Civil Law Rules on Robotics, Rapporteur Mady Delvaux.

EGE volá po širokém a systematickém veřejném zapojení se a debatě o etice AI, robotiky a „autonomních“ technologií a o množině hodnot, které se společnosti rozhodnou vložit do vývoje a správy těchto technologií. Tento proces, ve kterém je EGE připravena sehrát svoji úlohu, by měl připravit platformu pro spojení shora zmíněných různých globálních iniciativ. Měl by zahrnovat širokou, inkluzivní a dalekosáhlou společenskou debatu, čerpající z podnětů různých perspektiv, kde budou slyšet lidé s různou expertizou a různými hodnotami. EGE vybízí Evropskou unii, aby zaujala místo v předvoji takového procesu a vybízí Evropskou komisi, aby spustila a podporovala jeho implementaci.

Jako první krok k formulaci skupiny etických zásad, které mohou sloužit jako základ pro ustavení globálních standardů a legislativních akcí, navrhuje EGE skupinu základních principů a demokratických předpokladů, založených na základních hodnotách ustanovených ve Smlouvách EU a Listině základních práv Evropské unie.

# Etické principy a demokratické předpoklady

## Lidská důstojnost

Princip lidské důstojnosti, chápaný jako uznání, že člověk si sám o osobě zaslouží respekt, nesmí být přestupován „autonomními“ technologiemi. To například znamená, že existují limity na rozhodnutí a klasifikace týkající se osob, učiněná na základě algoritmů a „autonomních“ systémů, obzvláště když ti, kterých se týkají, o nich nejsou informováni. Také to znamená, že musí být (legislativní) limity pro způsoby, kterými jsou lidé vedeni věřit, že mluví s člověkem, i když ve skutečnosti interagují s algoritmy a chytrými stroji. Vztahová koncepce lidské důstojnosti, která je charakterizována našimi vztahy, vyžaduje, abychom věděli, zda a kdy interagujeme se strojem nebo jinou lidskou bytostí, a že si vyhrazujeme právo svěřit některé úkoly člověku nebo stroji.

## Autonomie

Princip autonomie implikuje svobodu lidské bytosti. To se překládá do lidské odpovědnosti a tím také kontroly a znalosti „autonomních“ systémů, protože ty nesmí narušit svobodu lidských bytostí stanovit si své vlastní standardy a normy a být schopni podle nich žít. Všechny „autonomní“ technologie proto musí uznávat lidskou schopnost vybrat si, zda, kdy a jak na ně delegovat svá rozhodnutí a akce. To také zahrnuje transparentnost a předvídatelnost „autonomních“ systémů, bez které by uživatelé nebyli schopni intervenovat nebo ukončit jejich činnost, pokud by to považovali a mravně nezbytné.

## Odpovědnost

Princip odpovědnosti musí být základem pro výzkum a aplikace AI. „Autonomní“ systémy by měly být vyvíjeny a používány jen způsoby, které slouží dobru společnosti a životního prostředí, jak jsou určena výstupy deliberativních demokratických procesů. To implikuje, že musí být navrženy tak, aby se jejich efekty shodovaly s pluralitou základních lidských hodnot a práv. Jelikož možné zneužití „autonomních“ systémů představuje velkou hrozbu, je klíčové znalost rizik a předběžná opatrnost. Aplikace AI a robotiky by neměly znamenat nepřijatelná rizika pro lidské bytosti a neohrožovat lidskou svobodu a autonomii tím, že by neoprávněně a podloudně snižovaly možnosti a znalosti občanů. Místo toho by měly ve vývoji být nasměrovány k rozšíření přístupu jednotlivců ke znalostem a příležitostem.

Výzkum, návrh a vývoj AI, robotiky a „autonomních“ systémů by měl být veden autentickým zájmem o etiku výzkumu, společenskou odpovědnost výzkumníků a celosvětovou akademickou spoluprací k ochraně základních práv a svobod a cílen na technologie, které je podporují a nepodkopávají.

## Spravedlnost, rovnost a solidarita

AI by měla přispívat ke globální spravedlnosti, rovnému přístupu k dobrům a výhodám, které AI, robotika a „autonomní“ systémy mohou přinést. Diskriminující zkreslení v datech používaných k tréninku a běhu systémů umělé inteligence by mělo být detekováno, oznámeno a neutralizováno co nejdříve.

Potřebujeme koordinované globální úsilí k rovnému přístupu k „autonomním“ technologiím a spravedlivé rozdělení benefitů a rovnost příležitostí mezi společnostmi a v rámci společností. To zahrnuje formulování nových modelů spravedlivého rozdělení a sdílení benefitů, schopné odpovědět na ekonomické transformace způsobené automatizací, digitalizací a AI, zajištění přístupu k hlavním

technologíím AI a podporu výcviku ve STEM a digitální vzdělávání, obzvláště ve vztahu ke znevýhodněným regionům a společenským skupinám. Je potřeba ostražitosti směrem k nevýhodám (rizikům) detailních a masivních dat o jednotlivcích, která se hromadí, a která vytvoří tlak na myšlenku solidarity, např. V systémech vzájemné pomoci jako sociální pojištění a zdravotní péče. Tyto procesy mohou nabourat společenskou kohezi a způsobit rozšíření radikálního individualismu.

## Demokracie

Klíčová rozhodnutí ohledně AI by měla být výsledkem demokratické debaty a veřejného angažmá. Duch globální spolupráce a veřejného dialogu zajistí, že budou přijata inkluzivním, informovaným a do dálky vidícím způsobem. Právo získat vzdělání nebo informace o nových technologiích a jejich etických dopadech umožní, aby každý chápal rizika a příležitosti a měl možnost se zapojit do rozhodovacích procesů, které zásadně ovlivňují naši budoucnost.

Principy lidské důstojnosti a autonomie zahrnují ve svém středu lidské právo na seburčení pomocí demokracie. Pro naše demokratické politické systémy mají klíčový význam hodnoty plurality, různosti a přijetí různých konceptů dobrého života občanů. Ty nesmí být ohroženy, podvráceny nebo zarovnány novými technologiemi, které potlačují nebo ovlivňují politické rozhodování a porušují svobodu projevu a právo získat a sdělit informace bez zásahů. Digitální technologie by raději měly být použity k zapřažení kolektivní inteligence a k podpoře a vylepšení společenských procesů, na kterých naše demokratické společnosti závisí.

## Vláda práva a odpovědnosti

Vláda práva, přístup ke spravedlnosti a právo na nápravu a na spravedlivý proces poskytují potřebný rámec pro dodržování standardů lidských práv a případných regulací AI. To zahrnuje i ochrany proti rizikům plynoucím z „autonomních“ systémů, které nedodržují lidská práva jako bezpečnost a soukromí.

Celý rozsah legislativních výzev vznikajících na tomto poli by měl být adresován včasnou investicí do vývoje robustních řešení, která poskytnou spravedlivé a jasné stanovení odpovědnosti a efektivní mechanismy aplikovatelného práva.

V tomto ohledu by vlády a mezinárodní organizace měly zvýšit své úsilí k vyjasnění, na kom leží zodpovědnost za škody způsobené nežádoucím chováním „autonomních“ systémů. Nad to by měly existovat efektivní systémy zabránění škodám.

## Bezpečnost, integrita tělesná i duševní

Bezpečnost a zabezpečení (safety and security) „autonomních“ systémů se projevuje ve třech formách. Za prvé, vnější bezpečnost pro prostředí a uživatele, za druhé spolehlivost a vnitřní robustnost např. proti hackerským útokům, za třetí emoční bezpečnost vzhledem k interakci člověka a stroje. Všechny rozměry bezpečnosti musí být vývojáři AI brány do úvahy a striktně otestovány před vydáním produktu, aby bylo zajištěno, že „autonomní“ systémy neporuší lidské právo na tělesnou a mentální integritu a na bezpečné a zabezpečené prostředí. Zvláštní pozornost by měla být věnována osobám, které se nachází ve zranitelné pozici a také potenciálnímu dvojímu užití AI a jejímu zbrojnímu využití, například v cybersecurity, financích, infrastruktuře a ozbrojeném konfliktu.

## Ochrana dat a soukromí

V době všudypřítomného a masivního sběru dat skrze digitální komunikační technologie je právo na ochranu osobních údajů a právo na respekt k soukromí kriticky zpochybňováno. Jak fyzické AI roboty jako součásti internetu věcí, tak AI softboty, které operují skrze World Wide Web, musí splňovat regulace na ochranu dat a nesbírat a nešířit data nebo být spouštěny na datech, k jejichž sběru a šíření nebyl dán informovaný souhlas.

„Autonomní“ systémy nesmí interferovat s právem na soukromý život, které zahrnuje právo svobody od technologií ovlivňujících osobní rozvoj a názory, právo vytvářet a rozvíjet vztahy s dalšími lidskými bytostmi, a právo svobody od sledování. V tomto směru také mají být definována a ustanovena přesná kritéria a mechanismy, aby se zajistil eticky správný vývoj a aplikace „autonomních“ systémů.

Ve světle obav týkajících se dopadů „autonomních“ systémů na soukromý život a soukromí, může být brán zřetel na probíhající debatu o zavedení dvou nových práv: práva na smysluplný lidský kontakt a práva nebýt profilován, měřen, analyzován, trénován nebo veden prostřednictvím pozitivní zpětné vazby a nepřímých doporučení (angl. nudging).

## Udržitelnost

Technologie AI musí být v souladu s lidskou odpovědností za základní předpoklady života na naší planetě, pokračujícího vzkvétání lidstva a udržení dobrého prostředí pro další generace. Strategie, jak zabránit budoucím technologiím, aby negativně ovlivnily lidský život a přírodu, musí být založeny na politikách, které zajišťují prioritu environmentální ochrany a udržitelnosti.

Umělá inteligence, robotika a „autonomní“ systémy, pokud budou moudře navrženy a nasazeny, mohou přinést prosperitu, přispět k blahobytu a pomoci dosáhnout evropských morálních a socio-ekonomických cílů. Etické úvahy a sdílené morální hodnoty mohou být využity k formování zítřejšího světa a měly by být chápány jako stimulus a příležitosti k inovaci, ne jako překážky a bariéry.

EGE vyzývá Evropskou komisi, aby prozkoumal, které existující právní nástroje jsou k dispozici pro efektivní zabývání se problémy diskutovanými v tomto vyjádření a zda jsou potřeba nové správní a regulační instrumenty.

EGE požaduje spuštění procesu, který by připravil cestu ke společnému, mezinárodně uznávanému etickému a právnímu rámci pro návrh, produkci, použití a správu umělé inteligence, robotiky a „autonomních“ systémů.