

# POSELSTVÍ SVATÉHO OTCE FRANTIŠKA K 57. SVĚTOVÉMU DNI MÍRU (1. LEDNA 2024)

## Umělá inteligence a mír

Na začátku nového roku, v čase milosti, kterou Pán uděluje každému z nás, bych se chtěl obrátit na Boží lid, na země, hlavy států a vlád, představitele různých náboženství a občanské společnosti a na všechny muže a ženy naší doby, abych jim odevzdal své přání pokoje.

### 1. Rozvoj vědy a techniky jako cesta k míru

Písmo svaté dosvědčuje, že Bůh dal lidem svého Ducha, aby je naplnil „moudrostí, chápavostí a znalostí každého díla“ (Ex 35,31). Chápavost je výrazem důstojnosti, kterou nám dal Stvořitel, jenž nás stvořil k svému obrazu a podobě (srov. Gn 1,26) a umožnil nám svobodně a uvědoměle odpovídat na jeho lásku. Věda a technika projevují zvláštním způsobem tuto zásadní vztahovou vlastnost lidské chápavosti a inteligence; jsou mimořádnými výtvoři jejího tvůrčího potenciálu.

Druhý vatikánský koncil v pastorální konstituci *Gaudium et spes* tuto pravdu potvrdil a prohlásil, že „člověk se vždy snažil rozvíjet svůj život prací a důvtipem“.<sup>1</sup> Když lidé „pomocí techniky“ usilují o to, aby se země stala „důstojným příbytkem celé lidské rodiny“,<sup>2</sup> jednají v souladu s Božím záměrem a spolupracují s jeho vůlí, aby bylo stvoření dovršeno a aby se mezi národy šířil mír. Také pokrok vědy a techniky, nakolik přispívá k lepšímu uspořádání lidské společnosti, k rostoucí svobodě a bratrskému společenství, vede ke zdokonalování člověka a k proměně světa.

Právem se radujeme a jsme vděční za mimořádné úspěchy vědy a techniky, díky nimž se podařilo odstranit nesčetné druhy zla, které sužovaly lidský život a způsobovaly velká utrpení. Vědeckotechnický pokrok, který umožňuje vykonávat dosud nevídanou kontrolu nad naší skutečností, zároveň vkládá do lidských rukou širokou škálu možností, z nichž některé mohou představovat riziko pro přežití a nebezpečí pro společný domov.<sup>3</sup>

Výrazný pokrok v nových informačních technologiích, zejména v digitální oblasti, proto představuje vzrušující příležitosti i vážná rizika, která mají závažné důsledky pro úsilí o spravedlnost a harmonii mezi národy. Je proto třeba si položit několik naléhavých otázek. Jaké budou střednědobé a dlouhodobé důsledky nových digitálních technologií? A jaký dopad budou mít na život jednotlivců a společnosti, na mezinárodní stabilitu a na mír?

### 2. Budoucnost umělé inteligence mezi přísliby a riziky

Pokrok v informatice a rozvoj digitálních technologií v posledních desetiletích již začaly způsobovat hluboké proměny v globální společnosti a její dynamice. Nové digitální nástroje mění tvář komunikace, veřejné správy, vzdělávání, spotřeby, osobních interakcí a bezpočtu dalších stránek každodenního života.

Kromě toho mohou technologie využívající množství algoritmů získávat z digitálních stop zanechaných na internetu údaje, které umožňují pro komerční nebo politické cíle kontrolovat mentální a vztahové návyky lidí, aniž by si toho byli vědomi, a omezovat tak jejich vědomé uplatňování svobodné volby. V prostoru, jako je web, který se vyznačuje přehlcením informacemi, totiž mohou strukturovat tok dat podle výběrových kritérií, jež nejsou uživatelem vždy vnímána.

<sup>1</sup> Druhý vatikánský koncil, Pastorální konstituce o církvi v dnešním světě *Gaudium et spes*, č. 33, in: *Dokumenty II. vatikánského koncilu*, Karmelitánské nakladatelství, Kostelní Vydří 2002.

<sup>2</sup> Tamtéž, č. 57.

<sup>3</sup> Srov. František, enc. *Laudato si'* (24. května 2015), č. 104, Paulínky, Praha 2018.

Musíme mít na paměti, že vědecký výzkum a technologické inovace nejsou odtržené od reality a nejsou „neutrální“,<sup>4</sup> ale podléhají kulturním vlivům. Tyto činnosti jsou plně lidské, proto směry, kterými se ubírají, odrážejí volby podmíněné osobními, společenskými a kulturními hodnotami každé doby. Totéž platí pro výsledky, jichž dosahují. Jelikož jsou plodem specificky lidských přístupů k okolnímu světu, mají vždy svůj etický rozměr, úzce spjatý s rozhodnutími těch, kdo navrhují experimentaci a orientují výrobu k určitým cílům.

To platí i pro různé podoby umělé inteligence. Ve světě vědy a techniky dosud neexistuje její jednoznačná definice. Samotný termín, který se dnes běžně užívá, zahrnuje celou řadu vědních oborů, teorií a technologií, jejichž cílem je, aby stroje při svém fungování napodobovaly nebo reprodukovaly kognitivní schopnosti člověka. Hovořit v množném čísle o „podobách inteligence“ napomáhá zdůraznit nepřeklenutelnou propast, která existuje mezi těmito systémy, jakkoli úžasnými a výkonnými, a lidskou osobou; v konečném důsledku jsou „fragmentární“ v tom smyslu, že mohou napodobovat nebo reprodukovat pouze určité funkce lidské inteligence. Použití množného čísla rovněž upozorňuje na fakt, že tato od sebe velmi odlišná zařízení je třeba vždy považovat za „sociotechnické systémy“. Jejich dopad totiž nezávisí pouze na jejich základní technologii, ale také na cílech a zájmech jejich vlastníků a výrojců, jakož i na situacích, v nichž jsou používány.

Umělou inteligenci je proto třeba chápat jako plejádu různých skutečností a nelze *a priori* předpokládat, že její rozvoj bude přínosem pro budoucnost lidstva a mír mezi národy. Takový pozitivní výsledek bude možný pouze tehdy, pokud prokážeme, že jsme schopni jednat odpovědně a respektovat základní lidské hodnoty, jako je „inkluzivita, transparentnost, bezpečnost, spravedlnost, diskrétnost a důvěryhodnost“.<sup>5</sup>

Nestačí ani předpokládat, že ti, kdo navrhují algoritmy a digitální technologie, se zavážou jednat eticky a odpovědně. Je třeba posílit nebo v případě potřeby zřídit orgány, které by se zabývaly nově vznikajícími etickými otázkami a ochraňovaly práva těch, kdo umělou inteligenci používají nebo jsou jí ovlivňováni.<sup>6</sup>

Obrovský rozmach technologií proto musí být doprovázen náležitým vzděláváním v oblasti odpovědnosti za jejich rozvoj. Svoboda a mírové soužití jsou ohroženy tehdy, když lidé podlehnou pokušení sobectví, zájmu jen o sebe samé, touhy po zisku a prahnutí po moci. Máme proto povinnost rozšířit svůj pohled a zaměřit technický a vědecký výzkum na úsilí o mír a společné dobro ve službách integrálního rozvoje člověka a lidského společenství.<sup>7</sup>

Přirozená důstojnost každého člověka a bratrství, které nás spojuje jako členy jedné lidské rodiny, musí být základem vývoje nových technologií a sloužit jako nezpochybnitelné kritérium pro jejich hodnocení, než budou použity, aby digitální pokrok mohl probíhat s respektem ke spravedlnosti a přispívat k míru. Technologický rozvoj, který nevede ke zlepšení kvality života celého lidstva, ale naopak prohlubuje nerovnosti a konflikty, nelze nikdy považovat za skutečný pokrok.<sup>8</sup>

Umělá inteligence bude stále získávat na významu. Výzvy, které s sebou přináší, jsou nejen technické, ale také antropologické, vzdělávací, sociální a politické. Slibuje například úsporu práce, efektivnější výrobu, snazší dopravu a dynamičtější trhy, stejně jako revoluci ve sběru, uspořádání a ověřování dat. Musíme si být vědomi probíhajících rychlých transformací a řídit je způsobem, který zaručuje dodržování základních lidských práv a respektuje instituce a zákony, které podporují integrální lidský rozvoj. Umělá inteligence by měla sloužit tomu nejlepšímu lidskému potenciálu a našim nejvyšším aspiracím, nikoli jim konkurovat.

---

<sup>4</sup> Srov. tamtéž, č. 114.

<sup>5</sup> *Audience pro účastníky setkání „Minerva Dialogues“* (27. března 2023).

<sup>6</sup> Srov. tamtéž.

<sup>7</sup> Srov. *Poselství výkonnému prezidentovi „World Economic Forum“ v Davosu-Klostesu* (12. ledna 2018).

<sup>8</sup> Srov. enc. *Laudato si'*, č. 194; *Projev k účastníkům semináře „Společné dobro v digitální epoše“* (27. září 2019).

### 3. Technologie budoucnosti: stroje, které se samy učí

Ačkoli se umělá inteligence založená na technikách strojového učení (*machine learning*) stále ještě nalézá v průkopnické fázi, v mnoha svých podobách již přináší významné změny do struktury společnosti a výrazně ovlivňuje kulturu, sociální chování a upevňování míru. Takový vývoj, jako je strojové učení nebo hluboké učení (*deep learning*), vyvolává otázky, které přesahují oblast technologie a inženýrství a souvisejí s porozuměním, jež je úzce spjato se smyslem lidského života, základními procesy poznání a schopností mysli dosáhnout pravdy.

Například schopnost některých zařízení vytvářet syntakticky a sémanticky koherentní texty není zárukou spolehlivosti. Říká se, že mají schopnost „halucinovat“, tj. generovat výroky, které na první pohled vypadají věrohodně, ale ve skutečnosti jsou nepodložené nebo prozrazují předsudky. Pokud je umělá inteligence využívána v dezinformačních kampaních, které šíří falešné zprávy a vedou k rostoucí nedůvěře ke sdělovacím prostředkům, nastoluje to vážný problém. Důvěrnost, vlastnictví údajů a duševní vlastnictví jsou dalšími oblastmi, v nichž dotyčné technologie představují vážná rizika. K nim se přidávají další negativní důsledky jejich zneužití, jakými jsou diskriminace, zasahování do volebních procesů, upevňování společnosti, která sleduje a kontroluje osoby, digitální vyloučenost a vyostření individualismu, jenž je stále více odtržen od společnosti. Všechny tyto faktory mohou podněcovat konflikty a bránit míru.

### 4. Smysl pro omezení v technokratickém paradigmatu

Náš svět je příliš rozsáhlý, rozmanitý a složitý na to, abychom jej mohli zcela poznat a klasifikovat. Lidská mysl nikdy nemůže vyčerpat jeho bohatství, a to ani za pomoci nejmodernějších algoritmů. Ty totiž nenabízejí zaručené předpovědi budoucnosti, ale pouze statistické přibližné odhady. Ne všechno lze předpovědět, ne všechno lze vypočítat; v konečném důsledku je „realita nadřazena ideji“<sup>9</sup> a bez ohledu na to, jak zázračná může být naše schopnost výpočtu, vždy bude existovat nedostupné reziduum, které uniká jakémukoli pokusu o změření.

Velké množství dat analyzovaných umělými inteligencemi navíc samo o sobě není zárukou nestrannosti. Když algoritmy extrapolují informace, vždy hrozí, že je zkreslí a budou replikovat nespravedlnost a zažitá předsudky z prostředí, ve kterém vznikly. Čím se stávají rychlejšími a složitějšími, tím obtížnější je pochopit, proč dospěly k určitému výsledku.

„Inteligentní“ stroje mohou plnit zadané úkoly s rostoucí efektivitou, ale účel a smysl jejich činnosti bude i nadále určován nebo umožňován lidmi, kteří mají své vlastní hodnotové spektrum. Riziko spočívá v tom, že kritéria určitých rozhodnutí budou méně jasná, že odpovědnost za rozhodování bude skrytá a že se tvůrci mohou vyhýbat své povinnosti jednat pro dobro společnosti. Svým způsobem tomu napomáhá technokratický systém, který spojuje ekonomiku s technologií a upřednostňuje kritérium efektivnosti při tendenci ignorovat vše, co nesouvisí s jeho bezprostředními zájmy.<sup>10</sup>

To by nás mělo přimět k zamyšlení nad aspektem, který je v dnešní technokratické a na výkon orientované mentalitě tak často přehlížen, ale který je rozhodující pro osobní a společenský rozvoj, a sice „smysl pro omezení“.

Pokud se lidská bytost, která je z definice smrtelná, domnívá, že díky technologii překročí všechna omezení, riskuje, že v posedlosti touhou mít vše pod kontrolou ztratí kontrolu sama nad sebou; v touze po absolutní svobodě se dostane do spirály technologické diktatury. Uznání a přijetí vlastní omezenosti jako tvora je nezbytnou podmínkou pro to, aby člověk dosáhl plnosti, či spíše ji přijal jako dar. Naopak v ideologickém kontextu technokratického paradigmatu, oživaného prométheovskou domněnkou soběstačnosti, by nerovnosti mohly až neúměrně narůstat a znalosti a bohatství hromadit v rukou několika málo lidí, což by znamenalo vážná rizika pro demokratické společnosti a pro mírové soužití.<sup>11</sup>

<sup>9</sup> František, apoštol. exhort. *Evangelii gaudium* (24. listopadu 2013), č. 233, Paulínky, Praha 2014.

<sup>10</sup> Srov. enc. *Laudato si'*, č. 54.

<sup>11</sup> Srov. *Projev k účastníkům plenárního zasedání Papežské akademie pro život* (28. února 2020).

## 5. Žhavá témata etiky

V budoucnu by mohly systémy umělé inteligence určovat spolehlivost žadatele o hypotéku, vhodnost osoby pro přijetí do práce, pravděpodobnost recidivy odsouzeného nebo právo na politický azyl či sociální pomoc. Nedostatek různých úrovní mediačního procesu, který tyto systémy zavádějí, je obzvláště náchylný k projevům předsudků a diskriminace; systémové chyby se mohou snadno znásobit a způsobit nejen nespravedlnost v jednotlivých případech, ale také dominovým efektem vyvolat různé formy skutečné sociální nerovnosti.

Někdy se navíc zdá, že umělá inteligence je schopna ovlivňovat rozhodnutí jednotlivců prostřednictvím předem stanovených možností volby, podněcujících či odrazujících, nebo prostřednictvím systémů pro regulaci osobních rozhodnutí na základě uspořádání informací. Tyto způsoby manipulace nebo sociální kontroly vyžadují pečlivou pozornost a dohled a předpokládají jasnou právní odpovědnost ze strany producentů, jejich zadavatelů i vládních orgánů.

Spoléhání se na automatické procesy, které kategorizují jednotlivce, například prostřednictvím všudypřítomného sledování nebo přijetí systémů sociálního kreditu, by také mohlo závažně ovlivnit občanskou strukturu a vytvořit nesprávnou klasifikaci občanů. Tyto umělé procesy třídění by rovněž mohly vést k mocenským konfliktům, protože se netýkají pouze virtuálních příjemců, ale lidí z masa a kostí. Základní úcta k lidské důstojnosti vyžaduje nepřipustit, aby byla jedinečnost osoby ztotožňována se souborem údajů. Nelze svolit, aby algoritmy určovaly způsob, jakým chápeme lidská práva, aby odsouvaly stranou základní hodnoty, jako je soucit, milosrdenství a odpuštění, nebo aby vylučovaly možnost, že se jednatel změní a nechá svou minulost za zády.

V této souvislosti se nemůžeme nezamyslet nad dopadem nových technologií na pracovní prostředí. Úkolů, které byly dříve výhradní doménou lidské práce, se rychle ujímají průmyslové aplikace umělé inteligence. I zde existuje značné riziko neúměrného zvýhodnění několika málo osob na úkor ochuzení řady jiných. Respektování důstojnosti pracovníků a významu zaměstnání pro ekonomickou prosperitu jednotlivců, rodin a společností, jistota zaměstnání a spravedlivá mzda by měly být pro mezinárodní společenství vysokou prioritou, protože uvedené druhy technologií pronikají stále více do pracovních prostředí.

## 6. Proměníme meče v radlice?

Při pohledu na svět kolem nás v dnešní době nelze přehlédnout vážné etické problémy spojené se zbrojním průmyslem. Možnost vést vojenské operace prostřednictvím dálkově řízených systémů přivodila oslabené vnímání devastace, kterou způsobují, a odpovědnosti za jejich použití, což přispívá k ještě chladnějšímu a odtažitějšímu přístupu k nesmírné tragédii války. Výzkum nových technologií používaných v oblasti takzvaných „autonomních vražedných zbraňových systémů“, včetně válečného využití umělé inteligence, představuje vážný etický problém. Autonomní zbraňové systémy nemohou být nikdy morálně odpovědnými subjekty. Výlučná schopnost člověka morálně usuzovat a eticky se rozhodovat je něčím víc než pouhým složitým souborem algoritmů a tuto schopnost nelze redukovat na programování stroje, který, ať je jakkoli „inteligentní“, zůstává stále jen strojem. Z tohoto důvodu je nezbytné zajistit odpovídající, smysluplný a důsledný lidský dohled nad zbraňovými systémy.

Nemůžeme také opomíjet možnost, že se sofistikované zbraně dostanou do nesprávných rukou a usnadní například teroristické útoky nebo zásahy zaměřené na destabilizaci legitimních vládních institucí. Svět skutečně nepotřebuje nové technologie, které by přispívaly k nekalému rozvoji trhu se zbraněmi a obchodu s nimi a podporovaly by šílenství války. Hrozí přitom, že se nejen inteligence, ale i samotné srdce člověka bude stávat stále více „umělým“. Nejpokročilejší technické aplikace by neměly sloužit k usnadnění násilného řešení konfliktů, ale k tomu, aby dláždily cesty k míru.

Viděno z pozitivnějšího hlediska by umělá inteligence, využívaná k podpoře integrálního lidského rozvoje, mohla přinést zásadní inovace v zemědělství, vzdělávání a kultuře, zlepšení životní úrovně celých zemí a národů a růst lidského bratrství a sociálního přátelství. Způsob, jakým ji využijeme k inkluzi těch nejmenších, tedy našich nejslabších a nejpotřebnějších bratrů a sester, je konečným měřítkem naší lidskosti.

Lidský pohled a touha po lepší budoucnosti našeho světa vedou k potřebě mezioborového dialogu zaměřeného na etický vývoj algoritmů, na *algor-etiku*, v níž by vodítkem orientujícím směřování nových technologií byly hodnoty.<sup>12</sup> Etické otázky by se měly brát v úvahu od samého počátku při výzkumu, stejně jako ve fázi testování, projektování, výroby, distribuce a uvádění na trh. Takový je přístup etiky projektového řízení, v němž zásadní roli mají hrát vzdělávací instituce a lidé zodpovědní za rozhodovací proces.

## 7. Výzvy pro vzdělávání

Vývoj technologií, které respektují a slouží lidské důstojnosti, má jasné důsledky pro vzdělávací instituce a pro svět kultury. Znásobením komunikačních možností nám digitální technologie zpřístupnily nové způsoby setkávání. Nadále je však třeba pokračovat v reflexi o tom, k jakým vztahům nás vedou. Mladí lidé vyrůstají v kulturním prostředí prostoupeném technologiemi, a to se nemůže nepromítnout do výukových a vzdělávacích metod.

Vzdělávání v oblasti využívání umělé inteligence by se mělo zaměřit především na podporu kritického myšlení. Je nezbytné, aby uživatelé všech věkových kategorií, zejména však mladí lidé, při používání údajů a obsahu shromážděných na webu nebo vytvořených systémy umělé inteligence si vypěstovali schopnost rozlišování. Školy, univerzity a akademické instituce by měly pomáhat studentům a odborníkům, aby si osvojili sociální a etické aspekty vývoje a využívání technologií.

Výuka k používání nových komunikačních nástrojů by měla brát v úvahu nejen dezinformace, falešné zprávy (*fake news*), ale také znepokojivý návrat „zdeděných obav [...], které byly schopny se skrýt za novými technologiemi a tam zesílit“.<sup>13</sup> Bohužel se opět setkáváme s tím, že musíme bojovat s „pokusením budovat kulturu hradeb, vytyčovat hradby, abychom zabránili setkání s jinými kulturami, s jinými lidmi“<sup>14</sup> a rozvoji mírového a bratrského soužití.

## 8. Výzvy pro rozvoj mezinárodního práva

Vzhledem ke globálnímu dosahu umělé inteligence je zřejmé, že kromě odpovědnosti suverénních států za regulaci jejího používání v jejich domácím prostředí mohou při uzavírání mnohostranných dohod a koordinaci jejich uplatňování a provádění sehrát rozhodující úlohu mezinárodní organizace.<sup>15</sup> V tomto smyslu vyzývám Společenství národů, aby pracovalo společně na přijetí závazné mezinárodní smlouvy upravující vývoj a používání umělé inteligence v jejích mnoha podobách. Cílem regulace by samozřejmě neměla být pouze prevence nesprávných postupů, ale také podpora těch správných, podněcování nových a tvůrčích přístupů a usnadňování osobních a kolektivních iniciativ.<sup>16</sup>

Při hledání normativních modelů, které by mohly vývojářům digitálních technologií poskytnout etické vodítko, je tedy nezbytné identifikovat lidské hodnoty, jež by měly tvořit základ snah společností formulovat, přijmout a prosazovat potřebné legislativní rámce. Při práci na přípravě etických směrnic pro tvorbu podob umělé inteligence nelze opomenout zohlednění hlubších otázek týkajících se smyslu lidského života, ochrany základních lidských práv a úsilí o spravedlnost a mír. Tento proces etického a právního rozlišování se může ukázat jako cenná příležitost ke společnému zamyšlení nad tím, jakou roli by měla technologie hrát v našich životech jednotlivců i společností a jak může její využívání přispět k vytváření spravedlivějšího a humánnějšího světa. Z tohoto důvodu by v debatách o regulaci umělé inteligence měly být zohledněny hlasy všech zúčastněných stran, včetně lidí chudých, stojících na okraji a dalších, kteří často zůstávají v globálních rozhodovacích procesech nevyšlyšeni.

---

<sup>12</sup> Srov. tamtéž.

<sup>13</sup> František, enc. *Fratelli tutti* (3. října 2020), č. 27, Karmelitánské nakladatelství, Praha 2021.

<sup>14</sup> Srov. tamtéž.

<sup>15</sup> Srov. tamtéž, č. 170–175.

<sup>16</sup> Srov. enc. *Laudato si'*, č. 177.

\* \* \* \* \*

Doufám, že nás tato úvaha povzbudí k tomu, abychom zajistili, že pokrok ve vývoji různých podob umělé inteligence bude nakonec sloužit lidskému bratrství a míru. To není na zodpovědnosti několika málo lidí, ale celé lidské rodiny. Mír je totiž plodem vztahů, které uznávají a přijímají druhého v jeho nezcizitelné důstojnosti, a plodem spolupráce a angažovanosti ve snaze o integrální rozvoj všech lidí a všech národů.

Na začátku nového roku se modlím za to, aby rychlý rozvoj umělé inteligence nezesílil mnohé nerovnosti a nespravedlnosti, jež jsou ve světě již přítomny, ale aby pomohl ukončit války a konflikty a zmírnit mnohé formy utrpení, které sužují lidskou rodinu. Kéž věřící křesťané, věřící různých náboženství i muži a ženy dobré vůle spolupracují ve vzájemné harmonii, aby využili příležitostí a čelili výzvám, které přináší digitální revoluce, a odevzdali budoucím generacím solidárnější, spravedlivější a mírumilovnější svět.

Ve Vatikánu dne 8. prosince 2023

*FRANTIŠEK*