



Evropská komise:

# ODBORNÁ SKUPINA NA VYSOKÉ ÚROVNI PRO UMĚLOU INTELIGENCI



## NÁVRH ĚTICKÉ POKYNY PRO ZAJIŠTĚNÍ DŮVĚRYHODNOSTI UI SHRNUTÍ

Pracovní dokument určený ke konzultaci zúčastněných stran

Brusel 18. prosince 2018

# NÁVRH ETICKÝCH POKYNŮ PRO ZAJIŠTĚNÍ DŮVĚRYHODNOSTI UI



Odborná skupina na vysoké úrovni pro umělou inteligenci  
**Návrh etických pokynů pro zajištění důvěryhodnosti UI**

Evropská komise  
Generální ředitelství pro komunikaci

Kontakt: Nathalie Smuha – Koordinátor skupiny AI HLEG  
E-mail: CNECT-HLG-AI@ec.europa.eu

Evropská komise  
B-1049 Bruxelles/Brussel

Zveřejnění dokumentu proběhlo 18. prosince 2018, v angličtině.

**Tento pracovní dokument zformulovala Odborná skupina na vysoké úrovni pro umělou inteligenci (AI HLEG), přičemž jeho obsahem není dotčen individuální postoj členů skupiny k jednotlivým bodům, ani konečné znění pokynů. Práce na formulaci konečného znění bude pokračovat do března 2019, kdy bude finální verze zveřejněna v návaznosti na konzultaci se zúčastněnými stranami realizovanou Evropskou aliancí pro umělou inteligenci.**

Komise ani jiné osoby jednající jejím jménem neodpovídají za použití níže uvedených informací. Odpovědnost za obsah tohoto pracovního dokumentu nese pouze Odborná skupina na vysoké úrovni pro umělou inteligenci (dále jen „skupina AI HLEG“). Přestože se zaměstnanci útvarů Komise na přípravě těchto pokynů podíleli, názory vyjádřené v tomto dokumentu odrážejí stanovisko skupiny AI HLEG a nelze je za žádných okolností považovat za oficiální stanovisko Evropské komise. Tento návrh představuje první výstup skupiny AI HLEG. Konečné znění bude Komisi předloženo v březnu 2019. Konečné znění druhého výstupu skupiny AI HLEG (*Politika v oblasti UI a investiční doporučení*) bude předloženo v polovině roku 2019.

Více informací o Odborné skupině na vysoké úrovni pro umělou inteligenci najdete na příslušné internetové stránce (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

Politiku opakovaného použití dokumentů Evropské komise upravuje rozhodnutí 2011/833/EU (Úř. věst. L 330, 14.12.2011, s. 39). K veškerému použití nebo reprodukci fotografií či jiného materiálu, k nimž EU nemá autorské právo, je nutné získat svolení přímo od držitelů autorských práv.

## **SHRNUTÍ**

Tento pracovní dokument představuje návrh etických pokynů pro UI, které vypracovala Odborná skupina Evropské komise na vysoké úrovni pro umělou inteligenci (skupina AI HLEG). Konečné znění pokynů má být zveřejněno v březnu 2019.

Umělá inteligence patří v současné době k silám, které zásadním způsobem transformují naši společnost. Zároveň nám nabízí jedinečnou příležitost ke zvýšení prosperity a růstu Evropy. Musíme usilovat o to, abychom ji nepromarnili. V posledním desetiletí se díky dostupnosti velkého množství digitálních dat, výkonné počítačové architektuře a dalšímu vývoji technologií UI (např. strojovém učení) podařilo v této oblasti dosáhnout zásadního pokroku. Významný rozvoj využití umělé inteligence u autonomních vozidel, ve zdravotní péči, domácích/pečovatelských službách zajišťovaných roboty, ve vzdělávání nebo oblasti kybernetické bezpečnosti umožňuje zvyšovat kvalitu našeho každodenního života. Umělá inteligence hraje také klíčovou roli při řešení mnoha hlavních celosvětových výzev, jako je zajištění zdraví a dobrých životních podmínek celosvětové populace, změna klimatu, prosazování důvěryhodných právních a demokratických systémů a další úkoly, které byly formulovány jako cíle OSN v oblasti udržitelného rozvoje.

Má tedy obrovský potenciál přinést výhody pro jednotlivce i celou společnost. Zároveň jsou však s UI spojena určitá rizika, která je třeba řádně řídit. Vzhledem k tomu, že výhody, které UI přináší, nad těmito riziky obecně převažují, musíme postupovat tak, aby byly **výhody maximalizovány a rizika minimalizována**. Proto je třeba **zaujmout ve vztahu k UI přístup zaměřený na člověka**, tedy mít stále na paměti, že rozvoj a využívání umělé inteligence by nemělo být samoučelné, ale mělo by mít za cíl zlepšovat životní podmínky lidí a zvyšovat prosperitu společnosti jako takové. **Musíme stavět na důvěryhodnosti a spolehlivosti technologií UI**. Jen tak budou moci lidé výhody, které jim UI nabízí, využívat s jistotou a v plné míře.

Důvěryhodná UI musí mít **dvě složky**: 1) měla by dodržovat základní práva, příslušné právní předpisy a základní zásady a hodnoty, a mít tak „**etický účel**“, a 2) měla by být **technicky odolná a spolehlivá**, jelikož i dobře míněná, avšak technologicky nedostatečně zvládnutá UI může způsobit neúmyslnou újmu.

Tyto pokyny proto stanoví **rámec pro důvěryhodnost UI**:

- **Kapitola I** se věnuje **zajištění etického účelu UI** stanovením základních práv, zásad a hodnot, které by tato technologie měla dodržovat.
- Z těchto zásad vychází **Kapitola II**, která obsahuje **pokyny pro realizaci důvěryhodné UI** a věnuje se jak etickému účelu, tak technické spolehlivosti UI. Obsahuje seznam požadavků, které by měla spolehlivá technologie UI splňovat, a předkládá přehled technických a netechnických metod, jež lze při praktické realizaci použít.
- **Kapitola III** se věnuje **praktickému kontextu** požadavků a uvádí konkrétní, i když neúplný hodnotící seznam, podle něho je možné důvěryhodnost technologií UI posuzovat. Tento seznam je pak přizpůsoben konkrétním případům použití.

Na rozdíl od jiných dokumentů na téma etické umělé inteligence se tyto pokyny nesnaží nabídnout jen další soupis základních hodnot a zásad pro UI, ale spíše poskytovat postup pro konkrétní realizaci a integraci těchto principů do systémů UI. Pokyny lze rozdělit do třech úrovní abstrakce: nejabstraktnější, uvedené v první kapitole zabývající se základními právy, zásadami a hodnotami, po mnohem konkrétnější uvedené v třetí kapitole spolu s hodnotícím seznamem.

Pokyny jsou určeny všem **zúčastněným stranám, jež technologie UI vyvíjejí, uvádějí do provozu nebo používají**. Jedná se o podniky, organizace, veřejné orgány, instituce, jednotlivce a jiné subjekty. Konečná verze těchto pokynů bude obsahovat mechanismus, díky němuž se zúčastněné strany budou moci k

pokynům vyjádřit (na dobrovolné bázi).

Zde je důležité zdůraznit, že pokyny nemají nahrazovat příslušné právní předpisy či jiné související politické dokumenty (tato otázka bude řešena v druhém výstupu skupiny AI HLEG – dokumentu Politika UI a investiční doporučení – jež má být předložen v květnu 2019), ani nemají vydání takových dokumentů bránit. Na pokyny musíme navíc nahlížet jako na otevřený dokument, který bude potřeba pravidelně aktualizovat, aby se zajistilo, že vždy odráží naši stávající znalost UI a nejnovější vývoj v této oblasti. Výchozím bodem pro diskusi na téma „**Důvěryhodná UI vytvořená v Evropě**“ by proto měl být tento dokument.

Evropa může svůj etický přístup propagovat pouze v případě, kdy bude v této oblasti sama konkurenceschopná na celosvětové úrovni. Nicméně **pro dosažení odpovědné konkurenceschopnosti je nezbytný právě etický přístup k UI**, protože vytvoří důvěru uživatelů a usnadní širší využívání UI. Cílem těchto pokynů není utlumit inovace UI v Evropě, ale namísto toho využívat etické roviny jako inspirace pro vytvoření jedinečné značky UI – značky, která usiluje o zajištění ochrany a výhod pro jednotlivce i společnost jako takovou. Právě tak může Evropa zaujmout vedoucí postavení v rozvoji špičkové, bezpečné a etické umělé inteligence. Předpokladem plného využití potenciálu UI v zájmu evropských občanů je zajištění důvěryhodnosti této technologie.

Tyto pokyny mají v neposlední řadě **rozvíjet diskusi** na téma etického rámce pro UI **na globální úrovni**.

## **POKYNY**

Jednotlivé kapitoly pokynů nabízejí postupy, jak zajistit důvěryhodnost UI. Jsou určeny všem relevantním zúčastněným stranám, které technologie UI vyvíjejí, zavádějí do praxe nebo používají. Shrnutí najdete níže:

### **Kapitola I: Základní pokyny pro zajištění etického účelu UI:**

- Zajistit, aby byla **UI zaměřená na člověka**: Technologie UI by měly být vyvíjeny, zaváděny a používány za **etickým účelem** a vycházet z dodržování základních práv, zohledňovat společenské hodnoty a etické principy *prospěšnosti, neškodlivosti, lidské autonomie, spravedlnosti a vysvětlitelnosti*. To je nezbytným předpokladem zajištění **důvěryhodnosti UI**.
- Vycházet při hodnocení možných budoucích účinků UI na člověka a společnost jako takovou ze základních práv, etických principů a hodnot. Věnovat **mimořádnou pozornost** situacím, které se týkají **zranitelnějších skupin obyvatel**, jako jsou děti, osoby se zdravotním postižením či menšiny, nebo situacím, ve kterých existuje **asymetrie distribuce moci, postavení nebo dostupnosti informací**, například u vztahů mezi zaměstnavateli a zaměstnanci nebo podniky a spotřebiteli.
- Uvědomit si a zohlednit skutečnost, že přes veškeré výhody, které jednotlivcům i společnosti obecně přináší, může mít UI i negativní dopad. To znamená být ostražití zejména v oblastech kritického významu.

### **Kapitola II: Základní pokyny pro zajištění důvěryhodnosti UI:**

- Začlenit níže uvedených **požadavků na důvěryhodnost UI do vývoje těchto technologií od samého začátku**: zodpovědnost, správa dat, řešení zohledňující potřeby všech uživatelů, správa autonomie UI (lidský dohled), nediskriminace, zajištění autonomie člověka, ochrana soukromí, spolehlivost a odolnost, bezpečnost, transparentnost.
- Zvážit použití technických i netechnických metod, které zajistí realizaci těchto požadavků v rámci systému UI. Zohledňovat tyto požadavky při sestavování týmu lidí, kteří budou na tomto systému pracovat, dále při vytváření systému jako takového, jeho testovacího prostředí a možnostech využití.

- **Poskytovat zúčastněným stranám** (zákazníkům, zaměstnancům atd.) proaktivně **jasné informace** o schopnostech, kapacitě a omezeních systému UI a umožnit jim tak vytvořit si realistická očekávání. Zajistit **sledovatelnost** systému UI je v této souvislosti naprostou prioritou.
- Učinit z konceptu důvěryhodnosti UI **součástí kultury organizace** a poskytovat zúčastněným stranám informace o tom, jak je tento koncept zohledňován při vývoji a využívání systémů UI. Důvěryhodnost UI mohou organizace rovněž začlenit do svých etických zásad a kodexů chování.
- Zajistit účast a **zapojení zúčastněných stran** na vytváření systému UI a jeho dalším rozvoji. Sestavovat pracovní týmy věnující se vývoji, testování a provádění produktu tak, aby byly co **nejrozmanitější**.
- Usilovat o **usnadnění kontrolovatelnosti** systémů UI, zvláště v kritických kontextech a situacích. Navrhnout systém v co největší míře tak, aby bylo možné vysledovat vazbu jednotlivých rozhodnutí na konkrétní data, vstupy a předem dané modely. Definovat **explikační metody** systému UI.
- Zajistit specifický postup pro **správu odpovědnosti**.
- Napláňovat **odbornou přípravu a vzdělávání** a zajistit, aby byli vedoucí pracovníci, vývojáři, uživatelé i zaměstnavatelé vyškoleni v otázkách důvěryhodnosti UI a orientovali se v nich.
- Je třeba mít na paměti, že mezi různými cíli může docházet k zásadním konfliktům (např. transparentnost může vést ke zneužití, zjištění a odstranění podjatosti může být v rozporu s ochranou soukromí). Je proto důležité o takovýchto kompromisních situacích informovat a evidovat je.
- Rozvíjet výzkum a inovace s cílem optimalizovat plnění požadavků na důvěryhodnost UI.

### **Kapitola III: Základní pokyny pro posuzování důvěryhodnosti UI:**

- Vytvořit **hodnotící seznam** pro důvěryhodnou UI, který se využije při vývoji, realizaci a využívání UI, a přizpůsobit ho konkrétnímu využití daného systému.
- Hodnotící seznam nemůže být **nikdy vyčerpávající**, jelikož důvěryhodnost UI není možné zjišťovat pomocí zaškrtování políček. Jde o kontinuální proces určování požadavků, vyhodnocování řešení a optimalizace výsledků během celého životního cyklu systému UI.

Tyto pokyny jsou součástí vize takového přístupu k UI, který bude zaměřený na člověka a který umožní Evropě stanout na světové špičce inovací v oblasti etické, bezpečné a moderní UI. Cílem je pomoci vytvořit značku **důvěryhodné UI vytvořené v Evropě**, s jejíž pomocí se evropským občanům zajistí lepší životní podmínky.