

Monitorovacia správa - priebežná

(Táto príloha sa do systému e-grant nevkladá, slúži ako pomôcka na prípravu Monitorovacej správy. Monitorovacia správa sa predkladá prostredníctvom formuláru v systéme e-grant.

V prípade projektov podporených v rámci výziev:

- 09I05-03-V02 - Podpora výskumných projektov zameraných na digitalizáciu ekonomiky v TRL úrovniach 1-3,
- 09I04-03-V02 - Podpora výskumných projektov zameraných na dekarbonizáciu ekonomiky v TRL úrovniach 1-3,
- 09I03-03-V03 - Veľké projekty pre excelentných výskumníkov,

Prijímateľ vyplňa MS v systéme e-grant v slovenskom aj v anglickom jazyku, pričom v oboch jazykoch uvedie rovnaké informácie.

Príloha č. 2 PpP verzia 2.6

Kód projektu:	09I01-03-V04-00100	Kód výzvy:	napr. 09I01-03-V04
Poradové číslo monitorovacej správy:	00001	Komponent:	K09 Efektívnejšie riadenie a posilnenie financovania výskumu, vývoja a inovácií
Aktuálne monitorovacie obdobie:	1.1.2025-30.9.2025	Investícia:	1. Podpora medzinárodnej spolupráce a zapájania sa do projektov Horizont Európa a Európsky inovačný a technologický inštitút (EIT – European Institute of Innovation and Technology)
Jednoznačná identifikácia projektu:	09I01-03-V04-00100, Morálna citlivosť a ľudské práva pre spracovanie jazykov s obmedzenými zdrojmi; Meno a priezvisko hlavného riešiteľa: Juraj Podroužek, fáza kariéry hlavného riešiteľa: R3		

1. Popis činností uskutočnených v rámci projektu počas monitorovacieho obdobia

"

Riešenie projektu za prvé obdobie prebiehalo podľa plánu a v súlade s harmonogramom projektu, ako to dokumentujú ďalšie sekcie správy. V rámci riešenia boli realizované aktivity súvisiace s výskumom metód pre posúdenie dopadov systémov AI na ľudské práva a to hlavne prvej časti výskumnej úlohy, ktorá je venovaná rešeršu dostupných FRIA metodológií a posúdeniu ich súladu s AIA (úloha 3.1) a tiež výskumom metód efektívneho dotrénovania veľkých jazykových modelov pre jazyky s obmedzenými zdrojmi a ich aplikovanie za účelom dotrénovania vybraného základného jazykového modelu pre riešenie konkrétnych úloh v slovenčine a to predovšetkým prvej časti výskumnej úlohy ktorá pozostáva z analýzy súčasného stavu poznania, výberu vhodnej

2. Míľniky

Štruktúra kapitoly pre projekty rozdelené na pracovné balíky

V tejto časti sa uvádzajú informácie o stave dosahovania všetkých míľnikov uvedených v prílohe č. 2 ZoPPM, a to nasledovne:

Poradové číslo a názov míľnika: Uvádzajú sa samostatne všetky míľniky v poradí, ako sú uvedené v prílohe č. 2 ZoPPM.

Typ a číslo monitorovacej správy: Uvádza sa "priebežná monitorovacia správa" a číslo tej monitorovacej správy, ktorá pokrýva monitorovacie obdobie, v ktorom bol míľnik úplne dosiahnutý (t. j. niektorá z prechádzajúcich alebo aktuálna priebežná monitorovacia správa), inak toto pole zostáva prázdne.

Popis dosiahnutia míľnika: Popis sa uvádza iba v prípade, ak bol míľnik dosiahnutý (úplne alebo čiastočne) počas aktuálneho monitorovacieho obdobia, inak ostáva pole prázdne. Uvedie sa stručný popis, ako bol míľnik dosiahnutý a názov dokumentu, ktorým prijímateľ preukazuje úplné dosiahnutie míľnika. Prijímateľ zároveň uvádza, či je daný dokument prílohou tejto monitorovacej správy alebo ŽoP. V prípade, že je prílohou ŽoP, je potrebné uviesť typ a číslo ŽoP, prílohou ktorej daný dokument je. V prípade, že míľnik nebol ku dňu ukončenia aktuálneho monitorovacieho obdobia dosiahnutý úplne, uvedie sa stav jeho dosiahnutia ku dňu ukončenia aktuálneho monitorovacieho obdobia.

Poradové číslo a názov míľnika	Typ a číslo monitorovacej správy	Názov pracovného balíka/pracovných balíkov	Popis dosiahnutia míľnika
--------------------------------	----------------------------------	--	---------------------------

M1 Dokončený prehľad FRIA metodológií a posúdenie ich súladu s AIA. Zhromaždená a spracovaná dátová množina a dotrénovaná prvá verzia jazykového modelu. Pripravená analýza existujúcich konceptualizácií a metód merania morálnej citlivosti.	priebežná, 0001	NV3	Dátová množina je dostupná na online úložisku; Dotrénovaný model je dostupný online. Výsledky prehľadu FRIA a analýzy existujúcich konceptov morálnej citlivosti sú dostupné v rámci priebežnej správy.
M2 Dokončená FRIA pre AutoML systémy aj s aplikáciou na modely projektu ALFIE. Dotrénovaná a vyhodnotená finálna verzia jazykového modelu. Otestovaný koncept morálnej citlivosti na vybraných metodikách pre posúdenie dôveryhodnej AI.			Dátová množina je dostupná na online úložisku; Dotrénovaný model je dostupný online. Výsledky prehľadu FRIA a analýzy existujúcich konceptov morálnej citlivosti sú dostupné v rámci priebežnej správy.

3. Výstupy projektu			
Štruktúra kapitoly pre projekty rozdelené na pracovné balíky			
V tejto časti sa uvádzajú informácie o stave dosahovania všetkých výstupov projektu uvedených v prílohe č. 2 ZoPPM, a to nasledovne: Poradové číslo a názov výstupu: Uvádzajú sa samostatne všetky výstupy v poradí, ako sú uvedené v prílohe č. 2 ZoPPM. Typ a číslo monitorovacej správy: Uvádza sa "priebežná monitorovacia správa" a číslo tej monitorovacej správy, ktorá pokrýva monitorovacie obdobie, v ktorom bol výstup dosiahnutý (t. j. niektorá z prechádzajúcich alebo aktuálna priebežná monitorovacia správa), inak toto pole zostáva prázdne. Popis dosiahnutia výstupu Popis sa uvádza iba v prípade, ak bol výstup dosiahnutý (úplne alebo čiastočne) počas aktuálneho monitorovacieho obdobia, inak ostáva pole prázdne. Uvedie sa stručný popis, ako bol výstup dosiahnutý a názov dokumentu, ktorým prijímateľ preukazuje úplné dosiahnutie výstupu projektu. Prijímateľ zároveň uvádza, či je daný dokument prílohou tejto monitorovacej správy alebo ŽoP. V prípade, že je prílohou ŽoP, je potrebné uviesť typ a číslo ŽoP, prílohou ktorej daný dokument je. Dokument preukazujúci úplné dosiahnutie výstupu musí obsahovať aj splnenie požiadaviek na povinnú publicitu v zmysle VZP. V prípade, že výstup projektu nebol ku dňu ukončenia aktuálneho monitorovacieho obdobia dosiahnutý, uvedie sa jeho stav ku dňu ukončenia aktuálneho monitorovacieho obdobia.			
Poradové číslo a názov výstupu	Typ a číslo monitorovacej správy	Názov pracovného balíka/pracovných balíkov	Popis dosiahnutia výstupu

V1.1 Priebežná správa o implementácii a dosiahnutých výsledkoch	priebežná monitorovacia správa I	NV1	Odovzdanie priebežnej monitorovacej správy I vrátane príloh opisujúcich doteraz dosiahnuté výsledky, a to konkrétne "Prehľad FRIA metodológií a posúdenie ich súladu s AIA" (Príloha A) a "Analýza existujúcich konceptualizácií a metód morálnej citlivosti" (Príloha B)"
V1.2 Záverečná správa o dosiahnutých výsledkoch	Záverečná monitorovacia správa.	NV1	
V2.1 Správa o výsledkoch komunikácie, diseminácie a využitia výsledkov projektu	Záverečná monitorovacia správa.	NV2	
V3.1 Výskumná správa o metódach posúdenia dopadov systémov AI na ľudské práva	Záverečná monitorovacia správa.	NV2	
V3.2 Výskumná správa o metódach efektívneho dotrénovania veľkých jazykových modelov pre jazyky s obmedzenými zdrojmi a ich aplikovaní za účelom dotrénovania vybraného základného jazykového modelu pre riešenie konkrétnych úloh v	priebežná monitorovacia správa I	NV3	Odovzdanie priebežnej monitorovacej správy I vrátane prílohy opisujúcej doteraz dosiahnuté výsledky, konkrétne "Príloha C: Správa o modelovaní a dátach", dátovej sady zverejnenej prostredníctvom platformy Hugging Face (https://huggingface.co/datasets/kinit/synthetic-queries-and-ml-instructions) a modelu
V3.3 Výskumná správa o modelovaní morálnej citlivosti tvorcov systémov AI	Záverečná monitorovacia správa.	NV3	

4. Identifikácia problémov a rizík, prijaté opatrenia na ich odstránenie

V tejto časti sa uvádzajú identifikované problémy, ktoré nastali počas monitorovacieho obdobia, za ktoré sa správa predkladá, v kontexte plánovaných cieľov projektu/výstupov a míľnikov. Prijímateľ zároveň popíše opatrenia, ktoré prijal v súvislosti so vznikom problémov. V tejto časti prijímateľ identifikuje riziká, ktoré by mohli nastať počas nasledujúceho monitorovacieho obdobia a popíše opatrenia, ktoré plánuje prijať na ich elimináciu. **Prijímateľ uvedie aj prípadné zmeny v spôsobe, postupe a metodike realizácie projektu, ktoré nastali v porovnaní s plánovanými v pôvodnom Opise projektu (súčasť ŽoPPM) počas monitorovacieho obdobia, ktorými zabezpečil dosiahnutie stanovených cieľov a výstupov projektu.**

V rámci Žiadosti boli identifikované relevantné riziká na celú dobu riešenia projektu a boli navrhnuté opatrenia na ich minimalizáciu. Môžeme konštatovať, že v rámci prvej polovice riešenia projektu sa Žiadne z rizík nenaplnilo a riešenie projektu prebieha podľa pôvodného plánu. Zároveň konštatujeme, že doteraz neboli identifikované ani Žiadne nové riziká, s ktorými by pôvodná projektová Žiadosť nepočítala. Za najrelevantnejšie riziko nasledujúceho obdobia považujeme možný nedostatok participantov a príležitostí na verifikáciu metód merania morálnej citlivosti (pravdepodobnosť stredná, závažnosť vysoká). Na jeho minimalizáciu realizujeme včasné a cieľené oslovovanie cez existujúce aj nové partnerské siete.

5. Popis plánovaných činností v nasledujúcom monitorovacom období

V tejto časti sa uvádza popis plánovaných činností pre nasledujúce monitorovacie obdobie. V prípade indikovaného omeškania sa osobitne uvedú plánované činnosti, ktorými sa zabezpečí realizácia projektu v požadovanom termíne.

V nasledujúcom monitorovacom období budú prebiehať predovšetkým nasledujúce činnosti:

1) FRIA pre AutoML/ALFIE

Vykonáme posúdenie dopadov na základné práva pre vybrané AutoML modely pilotované v rámci pôvodného HE projektu ALFIE.

2) Empirická adaptácia a pilot merania morálnej citlivosti tvorcov AI

Na báze vytvoreného rámca 7 komponentov a mapovania metód navrhujeme indikátory pre prostredie vývoja AI a zrealizujeme pilotnú kombinovanú štúdiu (rozhovory, focus skupiny, dotazník) s kontinuálnou revíziou nástroja. Výstupom bude prvá verzia adaptovaného meracieho rámca a plán rozšírenej validácie.

3) NLP

Rozšírenie vytvorenej dátovej sady o dátové body simulujúce komplexné scenáre. Dátové body v prvej verzii dátovej sady nereprezentujú konverzáciu používateľa a chatbota, ale iba jednu ucelenú požiadavku používateľa. Cieľom rozšírenia dátovej sady bude pridať scenáre, kedy používateľ poskytne informácie, ktoré majú byť jazykovým modelom extrahované, vo viacerých na seba nadväzujúcich odpovediach, simulujúc konverzáciu s chatbotom.

Dotrénovanie ďalších veľkých jazykových modelov na rozšírenej dátovej sade.

Porovnanie dotrénovaného modelu s multilinguálnymi modelmi priamo v kontexte AutoML platformy vyvíjanej v rámci projektu ALFIE.

Okrem toho sa bude pokračovať v manažmente projektu na exekutívnej dennej ako aj strednodobej strategickej báze. S pribúdajúcimi výsledkami sa rozšíria aktivity zameraná na potenciálnu exploitáciu riešenia a zabezpečenie jeho udržateľnosti. Členovia tímu sa zúčastnia medzinárodných aj národných fór, na ktorých budú osobne prezentovať výsledky projektu. Budeme naďalej pokračovať v propagácii projektu a dosiahnutých výsledkov, a to ako pre širokú verejnosť, tak aj v rámci vedeckej komunity a stakeholderov.

6. Prílohy

V tejto časti sa uvádzajú prílohy, ktorými prijímateľ preukazuje skutočnosti uvedené v častiach 2 a 3.

Príloha A: Prehľad FRIA metodológií a posúdenie ich súladu s AIA

Príloha B: Analýza existujúcich konceptualizácií a metód merania morálnej citlivosti

Príloha C: Správa o modelovaní a dátach

7. Vyhlásenie za osobu zodpovednú za projekt

Ja, dolu podpísaná osoba zodpovedná za projekt čestne vyhlasujem, že údaje uvedené v tejto monitorovacej správe a všetkých jej prílohách sú úplné a pravdivé.

Meno a priezvisko osoby zodpovednej za projekt/časť projektu:

Dátum podpisu a podpis osoby zodpovednej za projekt/časť projektu:

8. Vyhlásenie za prijímateľa

Ja, dolu podpísaný/á štatutárny orgán / osoba oprávnená konať za prijímateľa čestne vyhlasujem, že údaje uvedené v tejto monitorovacej správe a všetkých jej prílohách sú úplné a pravdivé.

Meno a priezvisko štatutárneho zástupcu/

Dátum podpisu a podpis štatutárneho