M Ě S T S K Á Č Á S T P R A H A 4

RADA MĚSTSKÉ ČÁSTI PRAHA 4

usnesení 24. jednání ze dne 4. 11. 2026
číslo 24R-797/2026

k uzavření dílčí objednávky na základě rámcové dohody na poskytování cloudových služeb umělé inteligence č. 2026/1679/OIKT ze dne 8.9.2026 se společností WHIRR CREW s.r.o. Rybná 716/24, 110 00 Praha 1, IČO: 03369145

Rada městské části Praha 4

I. r o z h o d l a
uzavřít dílčí objednávku na základě rámcové dohody na poskytování cloudových služeb umělé inteligence č. 2026/1679/OIKT ze dne 8.9.2026 se společností WHIRR CREW s.r.o. Rybná 716/24, 110 00 Praha 1, IČO: 03369145., dle přílohy č. 1 tohoto usnesení

II. z m o c ň u j e
místostarostu MČ Praha 4 Filipa Váchu
k uzavření dílčí objednávky dle bodu I. tohoto usnesení

III. u k l á d á
vedoucímu odboru informačních a komunikačních technologií
předložit dílčí objednávku dle bodu I. tohoto usnesení k podpisu

T: 6. 11. 2026

Zmocnění přijímám:

Předkladatel: Filip Vácha, místostarosta MČ Praha 4

P ř í l o h a

k usnesení Rady městské části Praha 4 č. 24R-797/2026, ze dne 4. 11. 2026

DÍLČÍ OBJEDNÁVKA č.
k rámcové dohodě č. SML/2026/1679/OIKT - cloudové služby umělé inteligence

Objednatel	Městská část Praha 4, Antala Staška 2059/80b, 140 46 Praha 4, IČO 00063584
Poskytovatel	WHIRR CREW s.r.o., Rybná 716/24, Staré Město, 110 00 Praha 1 IČO 03369145, DIČ CZ03369145
Kontaktní osoba objednatele	Mgr. Martin Fíla, vedoucí odboru IKT xxx
Kontaktní osoba poskytovatele	Vladimír Kříž, jednatel xxx
Datum objednávky	_____
Požadovaný termín plnění	Od data uveřejnění v registru na 12 měsíců. První 3 měsíce pilot, následně rozšíření podle výsledků.

1. Předmět dílčí objednávky

Objednáváme zřízení prostředí pro interní AI asistenty zaměstnanců ÚMČ Praha 4 a poskytování služeb AI inference. Součástí je pět agendových asistentů, kontrola limitů a měsíční výkazy spotřeby. První 3 měsíce proběhne pilot s předpokládanou skupinou 30 aktivních uživatelů. Předpokládaný počet 150 uživatelů je pouze odhad a může být kdykoliv zvýšen podle požadavků úřadu. Navýšení počtu uživatelů samo o sobě nemění maximální cenu objednávky ani sjednané limity čerpání. Podrobný rozsah a ceny obsahuje příloha č. 1, která je nedílnou součástí této objednávky.

Oblast služby	☒ Generativní/textová AI ☒ Dokumenty a znalostní zdroje ☒ Multimodální AI ☒ Hlasové služby ☒ Integrace/konzultace Multimodální a hlasové služby pouze podle schválené konfigurace.
Konkrétní model	Claude Haiku 5.5; Claude Sonnet 5.5 - inference v EU. Související modely D1/E1/F1/F2 podle schváleného katalogu a rozsahu, se zpracováním dat výhradně v EU.
Výrobce/provozovatel modelu	Výrobce generativních modelů: Anthropic. Inference: Amazon Bedrock v EU. Poskytovatel integrační služby: WHIRR CREW s.r.o.
Způsob přístupu	☒ uživatelské rozhraní ☒ API - programové rozhraní Open WebUI objednatele; LiteLLM pro směrování, limity a měření spotřeby.

Rozsah plnění	Implementace, konfigurace a úvodní školení: nejvýše 185,6 hodiny / 290 000 Kč bez DPH. AI inference a související spotřební služby: společný limit 405 000 Kč bez DPH za 12 měsíců. Bez licenčního poplatku za jednotlivého uživatele.
Jednotková cena bez DPH	Implementace a konzultace G1/G2: 1 562,50 Kč/h. Ceník modelů, cache a služeb D1/E1/F1/F2 je uveden v příloze č. 1. Účtují se poměrná množství.
Maximální cena objednávky bez DPH	695 000 Kč bez DPH DPH 21 %: 145 950 Kč Celkem včetně DPH: 840 950 Kč
Místo zpracování dat	Open WebUI, zdrojová databáze dokumentů a znalostní báze pro vyhledávání (RAG) zůstávají v prostředí objednatele. Pro externí inferenci se předává konkrétní dotaz a nezbytné relevantní výňatky, nikoli celá znalostní báze. AI inference probíhá výhradně v Evropské unii podle písemně schválené konfigurace.
Doba uchování dat	Dokumenty a historii v Open WebUI spravuje objednatel. Pro inference se navrhuje vypnuté obsahové logování a krátkodobá cache dle modelu. Přesné retenční doby a výmaz budou písemně potvrzeny před zahájením zpracování dat.
Další zpracovatelé	Amazon Web Services EMEA SARL a schválené subjekty AWS pro zvolenou konfiguraci se zpracováním dat výhradně v EU. Úplný seznam, regiony EU a datové podmínky budou písemně potvrzeny před zahájením čerpání příslušných služeb.
Zpracování osobních údajů	☒NE ☐ ANO - samostatná smlouva o zpracování osobních údajů je povinnou součástí této objednávky. Do jejího uzavření pouze veřejné nebo anonymizované podklady.

2. Bezpečnostní a datové podmínky

- Data objednatele nesmějí být použita k trénování, dotrénování nebo jinému zdokonalování modelu.
- Data objednatele nesmějí být zpřístupněna jiným zákazníkům ani použita pro jiné účely než poskytnutí této objednané služby.
- Poskytovatel nesmí bez předchozího písemného souhlasu objednatele změnit konkrétní model, provozovatele modelu, místo zpracování nebo další zpracovatele uvedené v této objednávce.
- Po uplynutí sjednané doby uchování budou data odstraněna, pokud právní předpis nevyžaduje jejich další uchování.

3. Akceptace a fakturace

Implementace se akceptuje písemným potvrzením funkční integrace a předáním sjednaných výstupů objednateli. Průběžné práce a spotřeba se dokládají měsíčním výkazem. Fakturace proběhne měsíčně podle doložených hodin a skutečně odebraných služeb, nejvýše do limitu implementace 290 000 Kč a spotřebních služeb 405 000 Kč bez DPH. Nevyčerpaná část se neúčtuje. Splatnost faktury je 30 dnů od doručení; faktura obsahuje číslo rámcové dohody a této dílčí objednávky. Výkaz rozliší implementační hodiny, model, běžné vstupní tokeny, čtení a zápisy cache, výstupní tokeny a případné další jednotky. Poskytovatel uvede průběžné čerpání a zbývající limit; před dosažením limitu upozorní objednatele a další čerpání omezí podle dohodnutého postupu.

4. GDPR – pouze při označení ANO

Je-li výše označeno „Zpracování osobních údajů: ANO“, poskytovatel nesmí zahájit zpracování osobních údajů před podpisem samostatné smlouvy o zpracování osobních údajů vztahující se k této objednávce. V případě rozporu mezi touto objednávkou a smlouvou o zpracování osobních údajů mají v otázkách ochrany osobních údajů přednost přísnější požadavky.

Za objednatele: _____ Datum: _____

Potvrzení poskytovatele: _____ Datum: _____

Příloha č. 1 Rozsah plnění a jednotkové ceny

Nedílná součást dílčí objednávky k rámcové dohodě č. SML/2026/1679/OIKT. Podklad: cenová nabídka č. WC-P4-INF-2026-10-26 ze dne 26. 10. 2026. Ceny jsou uvedeny v Kč; není-li uvedeno jinak, bez DPH.

1 Rozpočet prvního roku

Část plnění	Bez DPH	DPH 21 %	S DPH
Implementace, konfigurace a úvodní školení	290 000 Kč	60 900 Kč	350 900 Kč
AI inference a související spotřební služby	405 000 Kč	85 050 Kč	490 050 Kč
Maximální cena objednávky	695 000 Kč	145 950 Kč	840 950 Kč

Limit 405 000 Kč je společný maximální rozpočet skutečně odebraných modelů a souvisejících spotřebních služeb. Nevyčerpaná část se neúčtuje. Konkrétní měsíční limity budou nastaveny při zahájení pilotu.

2 Generativní modely

Ceny za 10 milionů tokenů příslušného typu. Vstup bez cache, čtení z cache a zápis do cache se vykazují odděleně. Stejný token se neúčtuje současně jako běžný vstup a zápis či čtení cache. Množství se účtuje poměrně, bez zaokrouhlování na celé bloky.

Model a režim	Vstup bez cache	Čtení z cache	Zápis do cache	Výstup
Claude Haiku 5.5 Anthropic / EU	55 Kč	5,50 Kč	68,75 Kč	275 Kč
Claude Sonnet 5.5 Anthropic / EU	1 050 Kč	55 Kč	1 050 Kč	5 250 Kč

Ceny zápisu do cache platí pro standardní dobu cache Anthropic 5 minut. Jiné retenční režimy, dlouhý kontext a zrychlené tarify se sjednají před použitím. Ceny nepřekračují maxima odpovídající kategorie A/B v modelovém koši; reasoning se započítává do výstupních tokenů.
Používány budou pouze modely a inferenční konfigurace se zpracováním dat výhradně v Evropské unii. Konkrétní model, provozovatel a inferenční konfigurace podléhají předchozímu písemnému schválení objednatele podle čl. 4 rámcové dohody. Přepínání včetně případných záložních konfigurací probíhá pouze v rámci schváleného zpracování v EU.

3 Související spotřební služby

Služba z katalogu	Účtovací jednotka	Cena bez DPH
D1 Vektorizace dokumentů (embeddings)	10 mil. vstupních tokenů	630 Kč
E1 OCR a vytěžování dokumentů	10 000 stran	6 300 Kč
F1 Přepis české řeči na text	10 000 minut	5 900 Kč
F2 Převod českého textu na řeč	10 mil. znaků	5 050 Kč

Služby D1/E1/F1/F2 budou aktivovány jen v dohodnutém rozsahu, se zpracováním dat výhradně v Evropské unii; jejich čerpání se započítává do společného limitu 405 000 Kč. Výrobce, model a datové podmínky vycházejí ze schváleného katalogu.

Příloha č. 1 Rozsah implementace a provozu

4 Rozpis implementace a konfigurace

Práce	Kód	Hodiny	Cena bez DPH
Architektura a tok dat	G1	16	25 000 Kč
Příprava serverového prostředí, nasazení Open WebUI, LiteLLM a napojení Bedrock	G1	49,6	77 500 Kč
Přístupy, limity a měření spotřeby	G1	16	25 000 Kč
Pět základních agendových asistentů	G1	40	62 500 Kč
Schválené dokumenty, základní RAG a OCR	G1	24	37 500 Kč
Testování, provoz a obnova	G1	16	25 000 Kč
Dokumentace a úvodní školení	G2	16	25 000 Kč
Upřesnění agend a koordinace	G2	8	12 500 Kč
Celkem při sazbě 1 562,50 Kč/h		185,6	290 000 Kč

Příprava serverového prostředí zahrnuje konfiguraci nasazení, síťového propojení a ověření přístupů i připravenosti k provozu. Jde o plánovaný rozsah prací. Účtuje se podle doložených hodin, nejvýše do limitu implementace. Přesné zadání pěti asistentů a vzorové výstupy se upřesní s garanty agend.

5 Pět agendových asistentů

- OSOC: záznamy a podklady ve Wordu.
- OVZ/OPR: porovnání smluv a návrh stanoviska.
- OŽ a zákon 106: podklady, lhůty a návrhy odpovědí.
- Porady a komunikace: zápisy, texty a překlady.
- Veřejné metodiky: vyhledávání ve schválených zdrojích s odkazy.

Asistenti připravují podklady pro další práci zaměstnanců úřadu. Skutečná spotřeba závisí na použitých modelech, délce dokumentů a četnosti dotazů.

6 Rozsah a provozní podmínky

Open WebUI a dokumenty budou na serverech objednatele. Poskytovatel zajistí přístup k modelům přes své LiteLLM a Amazon Bedrock s inferencí výhradně v EU, směrování, kontrolu limitů a měsíční výkazy spotřeby. Cena se odvíjí od provedených prací a skutečného využití, bez licenčního poplatku za jednotlivého uživatele.
Zdrojová databáze dokumentů a lokální data, nad nimiž probíhá vyhledávání (znalostní báze RAG včetně lokálního vyhledávacího indexu), zůstávají uloženy v prostředí objednatele. Tato databáze se nepřenáší ani nereplikuje do prostředí poskytovatele nebo externího modelu. Vyhledávání v ní probíhá lokálně. Pro vytvoření odpovědi se prostřednictvím LiteLLM a Amazon Bedrock předává konkrétní uživatelský dotaz a pouze nezbytné relevantní výňatky z podkladů ke zpracování výhradně v Evropské unii, podle písemně schválené konfigurace a datových podmínek. Tyto podklady se využívají pro vyhledávání a odpovědi, nikoli pro trénování nebo dotrénování modelů.
Objednávka zahrnuje základní práci se schválenými dokumenty a vzory. Propojení s E-spis, Agendio či ASPI a generování obrázků vyžadují upřesnění rozhraní, modelů a ceny v samostatně odsouhlaseném rozsahu.

7 Datové podmínky před zahájením provozu

Před zahájením zpracování skutečných osobních údajů musí být uzavřena samostatná smlouva o zpracování osobních údajů vztahující se k této objednávce. Určí účel, kategorie dat, schválené modely, místa zpracování, další zpracovatele a uchování dat. Do té doby probíhá ověřování na veřejných nebo anonymizovaných podkladech.