

Workshop riešiteľov projektu APVV DIGINARCH-AI

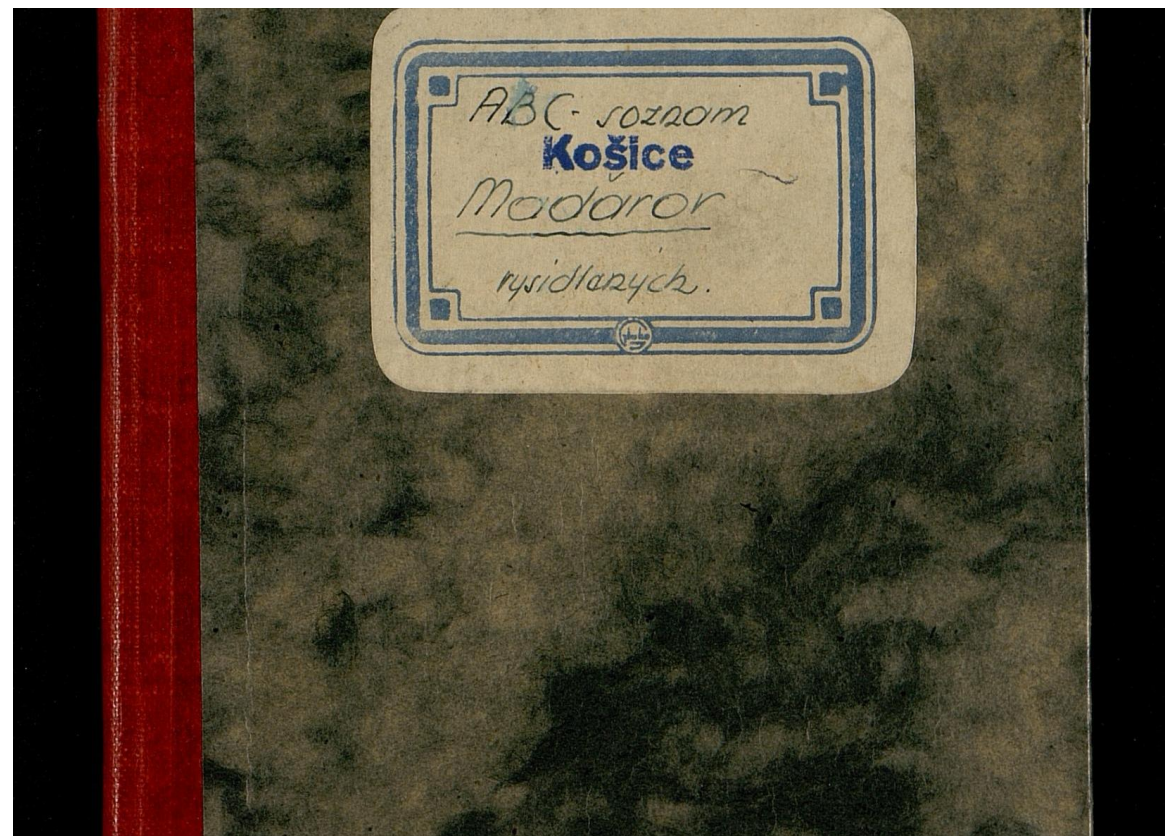


AGENTÚRA
NA PODPORU
VÝSKUMU A VÝVOJA

Imrich Nagy



10.09.2026



Ciel' stretnutia

Projektový stav

- ročná správa za rok 2025
- plnenie míľnikov
- publikácie a vykazovanie

Odborné bloky

- príprava dokumentov
- digitalizácia
- sumarizačný model
- transkripčný model

Stav po ročnej správe 2025

Ročná správa

Rada APVV hodnotila ročnú správu za rok 2025 pozitívne. Kontrola sa sústreďuje na vecné plnenie, publikácie a rozpočet.

- publikované výstupy priebežne posielať v PDF
- prezentácie z konferencií posielať v PDF
- výstupy sústrediť v repozitári UMB

Upozornenie v hodnotení

Dôsledne dodržiavať čerpanie pridelených finančných prostriedkov.

Prehľad pracovných balíkov a aktuálnych úloh

PB	Oblasť	Úlohy
PB1	Príprava dokumentov	Košice splnené, Banská Bystrica pripravená na prevoz
PB2	Model sumarizácie	návrh posunutý k integrovanému AI nástroju
PB3	Historický výskum	Úvodný informačný materiál vypracovaný, výskum, štúdie a konferencie pokračujú
PB4	Fyzický stav dokumentov	vstupná kontrola splnená za ŠA Košice
PB5	Digitalizácia	fond ŠA Košice kompletne zdigitalizovaný
PB6	Transkripčný model	Sady GroundTruth sa postupne dopĺňajú; testovanie v Transkribuse vykonané;

Vývoj modelu sumarizácie obsahu digitalizovaných dokumentov

Sumarizačný model

- návrh modelu pre obsah digitalizátov
- pracovná databáza z transkribovaného obsahu
- integrovaný nástroj s viacerými modulmi

Expert z praxe: Mgr. Peter Babic, IBM Slovakia

4

články prijaté
alebo vydané

2

články v
neskorom štádiu
posudzovania

3

články v príprave

2

články zaslané
do redakcií

Historický výskum

Historický výskum

- heuristika literatúry a prameňov
- doplňujúci výskum vo vybraných archívoch
- obsahová analýza dokumentov
- verifikácia výstupov zo sumarizačného modelu

- SYRNÝ, Marek: Riešenie tzv. maďarskej otázky v odboji a po II. sv. vojne Osídľovací úrad pre Slovensko a Fond národnej obnovy.
- viaceré konferenčné príspevky odovzdané

Ošetrovanie a výskum fyzického stavu dokumentov

- mechanická očista a sterilizácia fondu
- kontrola kyslosti a pevnosti papiera
- výskum ochrany materiálu a predĺženia životnosti

Vstupná kontrola fyzického stavu dokumentov – splnené za ŠA Košice

Digitalizácia fondov

Fond ŠA Košice

Kompletne zdigitalizovaný fond zahŕňa obsah 204 archívnych krabíc. Tím prevzal 408 skomprimovaných priečinkov cez zabezpečený univerzitný cloud.

204

krabíc

408

priečinkov

177 131

súborov na formát

AI modely a transkripčia

Transkripčný model

- GroundTruth datasety sú pripravené a ďalej sa dopĺňajú
- model bol otestovaný v prostredí Transkribus
- vzniká aplikácia na vizualizáciu digitalizátov a prepisov

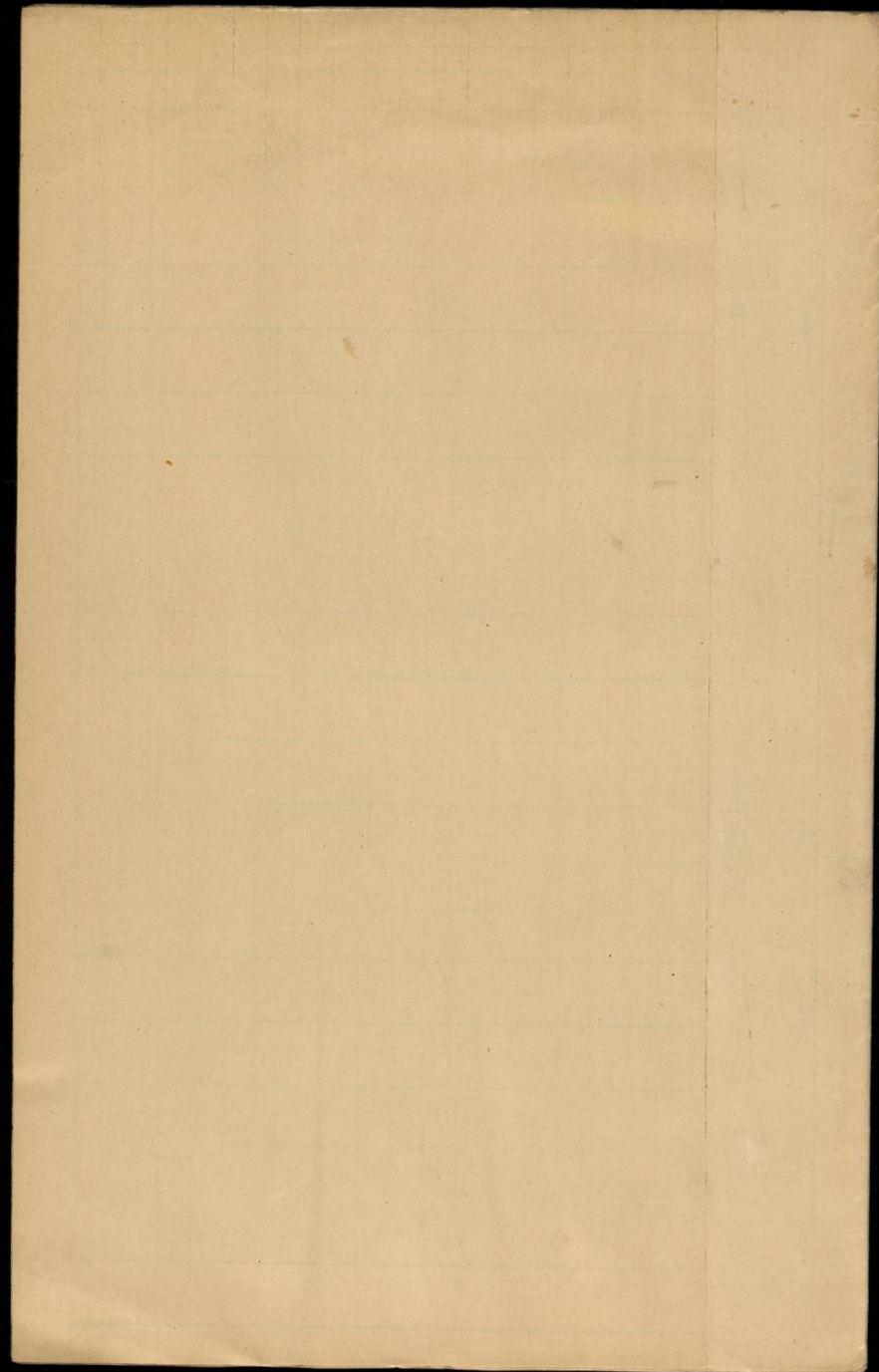
Kvalita vstupov pre transkripciu

Nie všetky digitalizáty majú informačnú hodnotu pre obsahovú analýzu

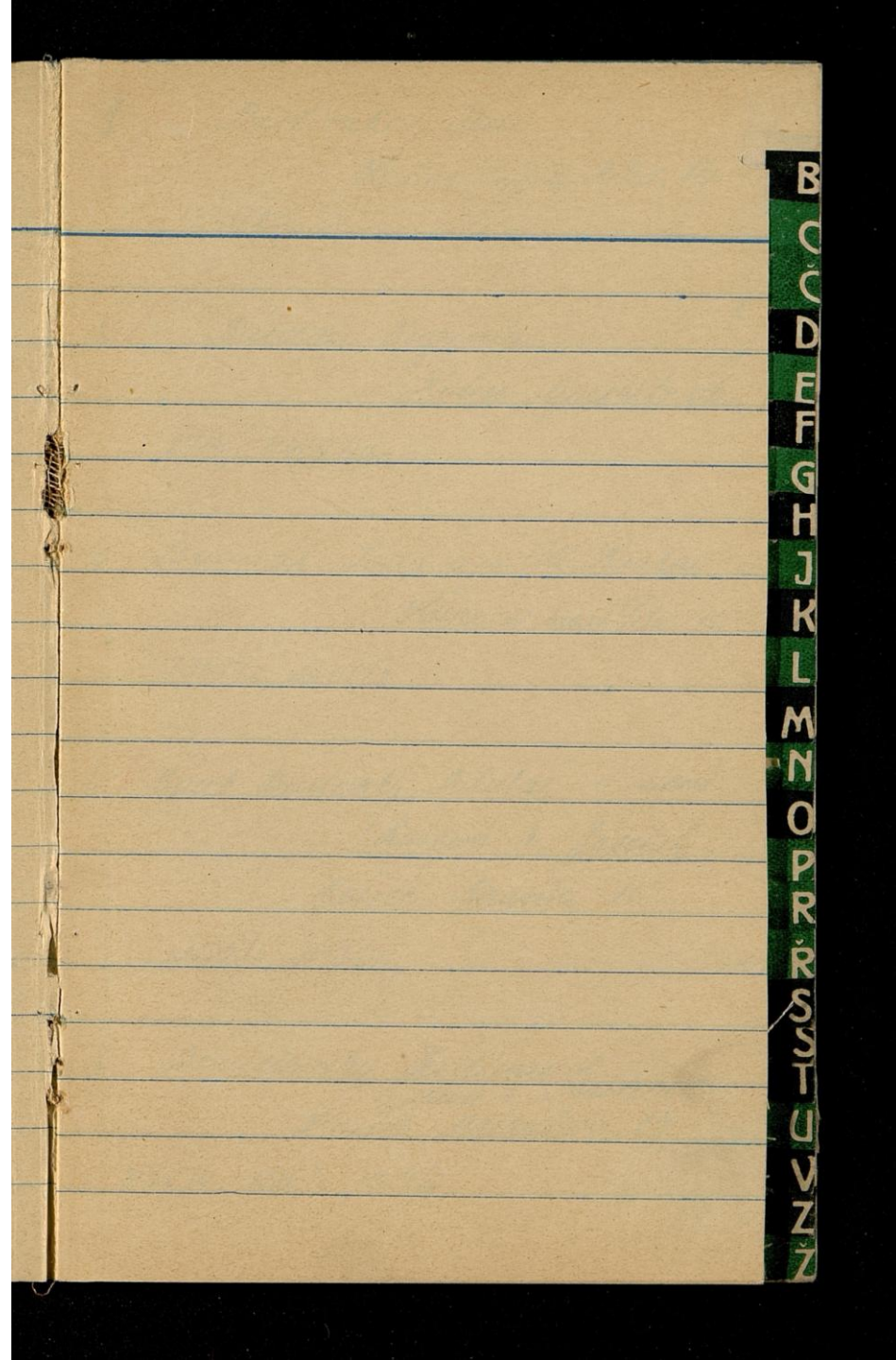
- prázdne strany
- prázdne formuláre
- spisové obaly
- šmírkaky a poštové potvrdenia
- mnohopočetné kópie rovnakého dokumentu

Prázdne a nerelevantné položky sa manuálne vyradujú z transkripcie. Duplicitné dokumenty treba softvérovo kontrolovať, aby neskresľovali štatistické vyhodnotenie.

SK_1540_11913_76_0272



SK_1540_11913_186-3_0008



SK_1540_11913_77_0899

[illegible]

SK_1540_11913_93_0011

Číslo:/194.....

PPPR - pracovná skupina pre pozemkovú reformu

v

Zasiela sa Vám s dožiadanim, aby ste v smysle ustanovenia § 2, odst. 1 zák. č. 148/1947 Sb. oznámili Vaše stanovisko.

V dňa 194.....

.....
prednosta OOÚ.

Číslo:/194.....

Oblasťná osídľovacia úradovňa

v

Zasielam späť s tým, že s vydaním výmeru súhlasíme — nesúhlasíme.

Dôvody:

V dňa 194.....

.....
prednosta pracovnej skupiny PPPR.

SK_1540_11913_93_0788

OCENENIE — ÉRTÉKELÉS:		Kész	Pengő
a) Kategória:			
b) Valuta:			
Konečná hodnota budovy —			
Az épület értéke:			
c) Poznámky pri ocenení — Megjegyzések az értékelésnél:			

Jégy

SK_1540_11913_87_0009

PREZÍDIUM OSÍDLOVACIEHO ÚRADU
PRE SLOVENSKO

Lep 34/ia-11

DOPORUČENÉ

Vec: _____

Výplata sumy Kčs _____ h _____

podľa účtu fy _____

Výdavok tento je finančným zákonom odôvodnený.
Z administratívneho stanoviska niet proti výplate
námietok.

Učtárni O. Ú.

k zisteniu úveru na pol _____ § _____

riad.-mimoriad. rozpočtu, upraveniu a pripojeniu
poukazovacej doložky.

Bratislava, dňa _____

Za predsedu: _____

P. ex. sp. 11

19. 11. 1951

21. V. 1951 *la*

Prezidiu O. Ú.

Kap _____ tit _____ § _____ pol _____

Ročný úver Kčs _____ h _____

12 roč. úveru Kčs _____ h _____

Dosiaľ vyčerpané Kčs _____ h _____

Zbýva voľných Kčs _____ h _____

Upravené:

Riadna - mimoriadna - potrebná vec:

Kap _____ tit _____ § _____ pol _____

Brutto Kčs _____ h _____

Srážky spolu Kčs _____ h _____

Netto k výplate Kčs _____ h _____

slovom _____

Bratislava, dňa _____

Prednosta účtárne: _____

Učtárni O. Ú.

k prevzatiu poukazu a realizovaniu upravenej čiast-
ky Pošt. sporiteľňou v Bratislave podľa pripojeného
platobného rozkazu.

Bratislava, dňa _____

Ref.: _____ Za predsedu: _____

Poukaz prevzatý.

Predpísané

Kniha _____ list _____ pol _____

UČTÁREŇ O. Ú.

Bratislava, dňa _____

Ref.: _____ Prednosta účtárne: _____

SK_1540_11913_81_0056

Alex. Verbóczy.

L

2000

42
100
142

825.8 = 100 -
20.

655.20
13 10
131
669.6.

98 x 220
196
4960
21.560

215.60 x 24
43120
86240
119440

3154 x 1690
18924
283860
5330260

5881 x 1440
23524
23524
8468640

21160 -
5144
16386

21560
4310
17250

19.190
1410
3.610
361
3.249

15 x 220 x 15
1100
3300
165
3135

160 x 4.5
640
800
1200
92
648

160
1320
1920

SK_1540_11913_83_0008

Doručenka ODBOR FINANČNÝ
rady Krajského národného výboru
v Košiciach

1. Prijímateľ' nebol v byte (prevádzovní, obchodnej miestnosti, kancelárii, pracovnom mieste) zastihnutý.
Spis bol preto doručený dospelému príslušníkovi rodiny*) — uložený dňa _____ u tunajšieho poštového úradu a daňovník bol o tom písomným oznámením upovedomený, keďže vyzve, aby v tejto miestnosti bol dňa _____ hod. prítomný, nevyhovet.')

Lehota 30-dňová pre uloženie spisu u tunajšieho úradu uplynula dňa _____

2. Prijímateľ' odprel prijať spis. Spis bol preto dňa _____ 195 _____ uložený u tunajšieho pošt. úradu*) zanechaný na mieste doručenia.')

3. Poštová plná moc pre _____

P. _____ zo dňa _____

Dňa _____

Denná pečiatka dodávacieho poštového úradu _____

Podpis doručovateľa _____

Vedúci úradu _____

*) Čo sa nehodi prečiarknite

(pre ktorý úrad?)

Adresát (prijímateľ):

P. T. _____

Štefan Szabó,

in. 2452/1957. Klín nad Bodrogom č. d. 23.

Číslo: Schm. _____ okr. Kráľovský Chlmec.

Potvrdzujem svojim vlastnoručným podpisom, že som horeuvedenú zásielku dnes dostal(a).

V _____ dňa X. 3. 195 _____

Doručené doručovateľom: _____

Podpis _____

Na doručke uvedte dátum doručenia, podpíšte a odovzdajte ju doručovateľovi. — Skl. č. 46

Transport 16-11-D.

S o z n a m p r e s i d l e n c o v .

Por. čís.	Meno presídlenca /hlava rodiny/	Počet osob	Adresa:	Odsídlený podľa	
1.	Sokoly Ondrej 1914, Alžbeta 1916, Viliam 1941, Ladislav 1946	5	Slanec 272	5	
2.	Varga Jan 1894, Mária 1903, Ján 1920, Ondrej 1923, Stefan 1933	5	" 73	5	
3.	Gyöngyösi Ondrej 1890	1	" 46	5	
4.	Ivan František 1908, Ondrej 1874, Júlia 1911, Ján 1928, Mikuláš 1932	5	" 59	5	
5.	Ivan Ondrej 1898, Júlia 1908, Ondrej 1921, Júlia 1923	4	" 57	5	
6.	Harda František 1910, Zofia 1916, František 1942	3	" 273	5	
7.	Možer František 1914	1	" 9	5	
8.	Peko Jozef 1879, Katarína 1888, Ondrej 1915	3	" 38	5	
9.	Peko Michal 1894, Alžbeta 1895, František 1922, Jolana 1926	4	" 76	5	
10.	Peko Ondrej 1911, Alžbeta 1914, Mikuláš 1937, Kata 1935	4	" 44	5	
11.	Šváby Dezider 1903	1	" 5	5	
12.	Varga Jan 1894, Mária 1903, Ján 1920, Ondrej 1923, Stefan 1933, Albert 1945	6	" 73	5	
13.	Peko Ján 1896, Zuzana 1900, Ján 1924, Alžbeta 1927, Zuzana 1935	5	" 50	5	
14.	Barna Ondrej 1890, Anna 1887, Michal 1925	3	Gerčel	1-4	
15.	Jager /Kostya/ Ľudovít 1875, Alžbeta 1885, Ľudovít 1919, Irena 1917	4	Sl. Nové Mesto 22	1-4	
15	jednotiek	53	osob		

Transport 16-11-D.

S o z n a m p r e s i d l e n c o v .

Por. čís.	Meno presídlenca /hlava rodiny/	Počet osob	Adresa:	Odsídlený podľa	
1.	Sokoly Ondrej 1914, Alžbeta 1916, Viliam 1941, Ladislav 1946	5	Slanec 272	5	
2.	Varga Jan 1894, Mária 1903, Ján 1920, Ondrej 1923, Stefan 1933	5	" 73	5	
3.	Gyöngyösi Ondrej 1890	1	" 46	5	
4.	Ivan František 1908, Ondrej 1874, Júlia 1911, Ján 1928, Mikuláš 1932	5	" 59	5	
5.	Ivan Ondrej 1898, Júlia 1908, Ondrej 1921, Júlia 1923	4	" 57	5	
6.	Harda František 1910, Zofia 1916, František 1942	3	" 273	5	
7.	Možer František 1914	1	" 9	5	
8.	Peko Jozef 1879, Katarína 1888, Ondrej 1915	3	" 38	5	
9.	Peko Michal 1894, Alžbeta 1895, František 1922, Jolana 1926	4	" 76	5	
10.	Peko Ondrej 1911, Alžbeta 1914, Mikuláš 1937, Kata 1935	4	" 44	5	
11.	Šváby Dezider 1903	1	" 5	5	
12.	Varga Jan 1894, Mária 1903, Ján 1920, Ondrej 1923, Stefan 1933, Albert 1945	6	" 73	5	
13.	Peko Ján 1896, Zuzana 1900, Ján 1924, Alžbeta 1927, Zuzana 1935	5	" 50	5	
14.	Barna Ondrej 1890, Anna 1887, Michal 1925	3	Gerčel	1-4	
15.	Jager /Kostya/ Ľudovít 1875, Alžbeta 1885, Ľudovít 1919, Irena 1917	4	Sl. Nové Mesto 22	1-4	
15	jednotiek	53	osob		

Ochrana dát pri práci s AI

Základná ochrana spracovávaných dokumentov zostáva zachovaná aj pri platenom API prístupe k modelom Gemini a OpenAI.

Transkribus

európska platforma pre dokumenty, servery v Rakúsku

Gemini API

pri platenom API sa obsah nepoužíva na zlepšovanie produktov

OpenAI API

API dáta sa štandardne nepoužívajú na tréning modelov

Praktický význam: dokument posielame službe na spracovanie, nie na verejné zverejnenie ani na bežné trénovanie modelov.

Porovnateľná základná ochrana

Nástroje majú odlišný právny režim, ale podobnú ochrannú logiku

Otázka	Transkribus	Gemini API, platené	OpenAI API, platené
Kto spracúva dáta	READ-COOP v prostredí Transkribus	Google podľa DPA pre platené služby	OpenAI ako sprostredkovateľ podľa DPA
Použitie na tréning	dáta slúžia na poskytovanie služby	prompty, súbory a odpovede sa nepoužívajú na zlepšovanie produktov	API dáta sa nepoužívajú na tréning, pokiaľ to zákazník nepovolí
Uchovávanie	bežné projekty ostávajú v účte, vybrané API spracovania majú krátku retenciu	bezpečnostné logy a niektoré funkcie majú obmedzenú retenciu	štandardné API logy najviac 30 dní, pri vybraných režimoch menej
Ľudský prístup	prístup obmedzený na oprávnené osoby	pri platenom API slúži prístup najmä na bezpečnosť a zákonné účely	prístup majú len oprávnené osoby na podporu, bezpečnosť a právny súlad

Podmienky bezpečného používania

Čo dodržať v projekte

- používať len platené API účty, nie bezplatné testovacie režimy
- mať akceptovaný DPA rámec
- posilať len dáta potrebné na konkrétnu úlohu
- nepoužívať webové doplnky a grounding pri citlivých dokumentoch
- mazať súbory, cache a uložené konverzácie, keď už nie sú potrebné

Záver

Platené API funguje ako externý odborný nástroj: dokument mu dočasne zveríme na spracovanie. Poskytovateľ ho podľa svojich pravidiel nemá používať ako bežný tréningový materiál a prístup k nemu má byť obmedzený.

Oficiálne dokumenty k ochrane dát

Transkribus

<https://legal.transkribus.org/privacy>

<https://legal.transkribus.org/terms>

<https://help.transkribus.org/is-my-data-safe>

Google Gemini API

<https://ai.google.dev/gemini-api/terms>

<https://ai.google.dev/gemini-api/docs/zdr>

<https://cloud.google.com/terms/data-processing-addendum>

OpenAI API

<https://developers.openai.com/api/docs/guides/your-data>

<https://openai.com/policies/data-processing-addendum/>

<https://openai.com/enterprise-privacy/>

Úlohy na najbližšie obdobie

Míľniky podľa pracovných balíkov a spoločné povinnosti tímov

Balík	Termín	Úlohy
PB1	splnené	príprava na digitalizáciu dokumentov
PB2	31.12.2026 31.01.2027 31.08.2027	ukončený vývoj sumarizačného modelu konferencia IEEE s publikovaným výstupom funkčný sumarizačný model
PB3	31.12.2026 31.03.2027 31.08.2027	štúdia k historickému kontextu štúdia o manažmente povojnových presunov štúdia o Slovenskom osídľovacom úrade
PB4	priebežne	pokračovať vo výskume fyzického stavu dokumentov
PB5	30.06.2027	vyhodnotiť ďalšiu fázu digitalizácie archívnych fondov
PB6	30.06.2027 31.12.2027	vyhodnotiť prvú fázu automatickej transkripcie štúdia o univerzálnom modeli pre historické strojopisy
PB7	31.12.2027	pripraviť návrh komunikačného modulu a modelu NLP

Spoločné úlohy

Do konca finančného roka 2026 dočerpať rozpočet v maximálnej možnej miere. Zmeny treba vopred komunikovať s projektovou manažérkou a po uzávierke poslať podklady o čerpaní.

Do 10.01.2027 zaslať

- slovný popis činností a výskumu
- bibliografické údaje výstupov
- PDF publikácií a prezentácií
- odôvodnenie výdavkových položiek