

ÁLLAMBIZTONSÁGI SZOLGÁLATOK
TÖRTÉNETI LEVÉLTÁRA
(ÁBTL)

SZIGORÚAN TITKOS

Tárgy.sz.: III/2.

Ügynökakták, algoritmusok, emlékezet

Hogyan segíti a matematika és
a mesterséges intelligencia a múlt rejtett
hálózatainak feltárását?

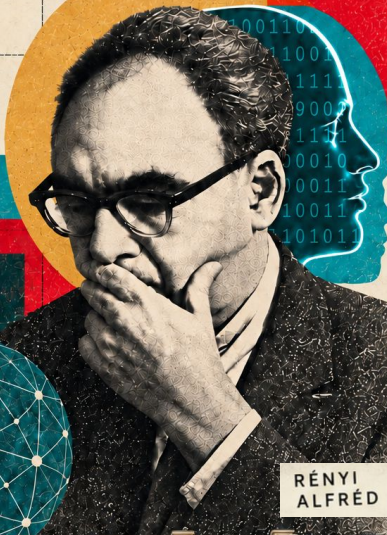
2026. Szeptember 25.

Jelentés
Budapest, 1944.

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^n p_i^\alpha \right)$$

$$P(G) = p^m (1-p)^{\binom{n}{2}-m}$$

$$M = \int_0^\infty \exp \left(-2 \int_0^t \frac{1-e^{-u}}{u} du \right) dt$$



RÉNYI
ALFRÉD



HONNAN INDULTUNK?

Zajos dokumentumok



BIZALMAS KÜLPOLITIKAI SZEMLE

1. Oldal.

1944. July 29.

Magyarország.

[illegible]

A Székelyfalva Tűzoltóság Budapesti Tűzoltóságának jelentést küldt a Budapesti ellenintézkedésekről a lakosság felvilágosításáról.

A Székelyfalva Tűzoltóság tisztelettel megköszöni a Magyar Tűzoltóság érkező hírekért, hogy az országban mindenütt a tűzvédelemről és a tűzoltóságról információkat nyújt, mert ez a legfontosabb feladat.

[illegible]

A moszkvai rádió magyar csatlósban a magyar ipari munkások sztrájkjáról szólva hangzott, hogy az ipari munkások bojkottálják az Imrédy által alkotott munkaszervezetet,

Kaloti grovonal.

[illegible]

1940. gada 22. jūnijā tika izveidota Igaunijas Holštēin-Štaļenberga apdzīvotās teritorijas valdība, kuras priekšsēdētājs bija J. Kallio. Šīs valdības sastāvā bija arī J. Kallio, kurš bija bijis Igaunijas valdības priekšsēdētājs, un J. Kallio, kurš bija bijis Igaunijas valdības priekšsēdētājs. Šīs valdības sastāvā bija arī J. Kallio, kurš bija bijis Igaunijas valdības priekšsēdētājs, un J. Kallio, kurš bija bijis Igaunijas valdības priekšsēdētājs.

Y P/P/2s No. 4 Hayes Iroda Hely-szolgálat 4. április 3-án hajnali 3
órás adásban budapesti közbeszólással részleteket közöl a Kis Újságunk Chamberlain
beszédéről irt cikkéből. /HAY/

X E/E A Stefani-iroda Hell-szolgálati est-atta 22.40. Budapest. Jarré Erkeztet. /UTI/ *aprilis 2. on: [signature]*

HC E/E Hell-Commune a 23.42-es addiban ezt adta. *gyors 2.-an:*
Budapest-i kormányzó repülőútja az árvízterületen. /MTI/

ET Kim Wa B o l g r a d , Aprilis 3. / Magyar Kövirati Irod./ 29
A Politikai Hivatal-jelentget kaptam. Szalutas Bolg.

[illegible]

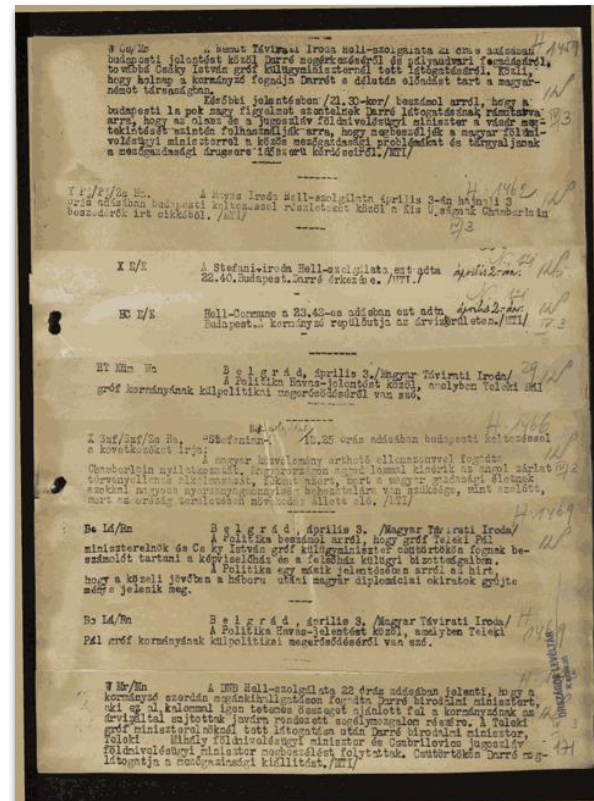
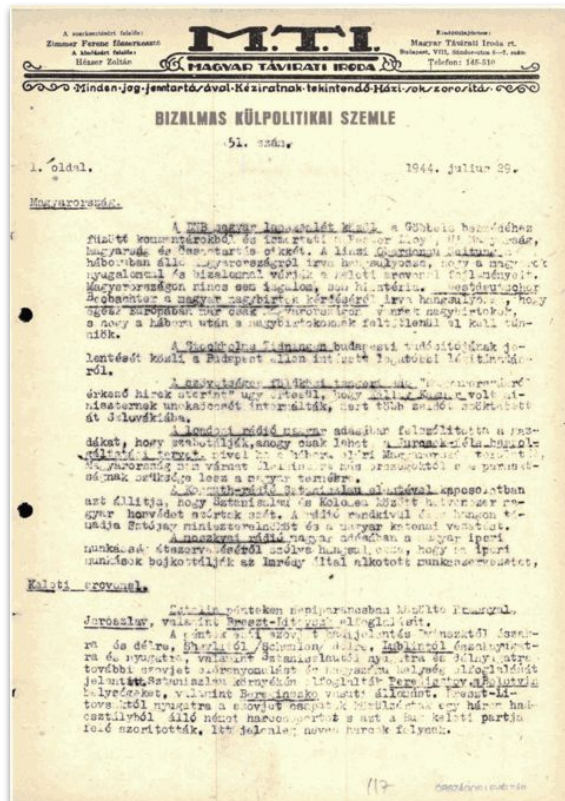
Belgrád, április 3. /Magyar Törvényi Iroda/
A Politikai bizottság arról, hogy gróf Teleki Pál
ministernak és Cs. gy. István külügyminiszternek együttösen fogak be-
szélni tartani a képviselőház és a felsőház külügyi bizottságában.
A Politikai egy másik jelentésében arról ad hírt,
hogy a közeli jövőben a háború utáni magyar diplomáciai okiratok gyűj-
teje jelenik meg.

1. Ld/Rn B e l g r á d , április 3. /Magyar Távirati Iroda/
A Politika Hírlap-jelentést közöl, amelyben Teleki
1. gróf kormányának külpolitikai meggyőződéséről van szó.

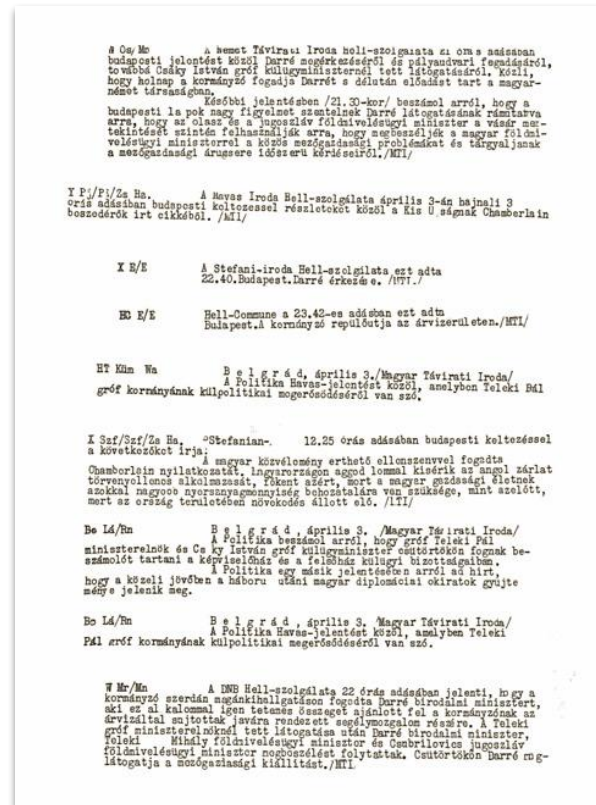
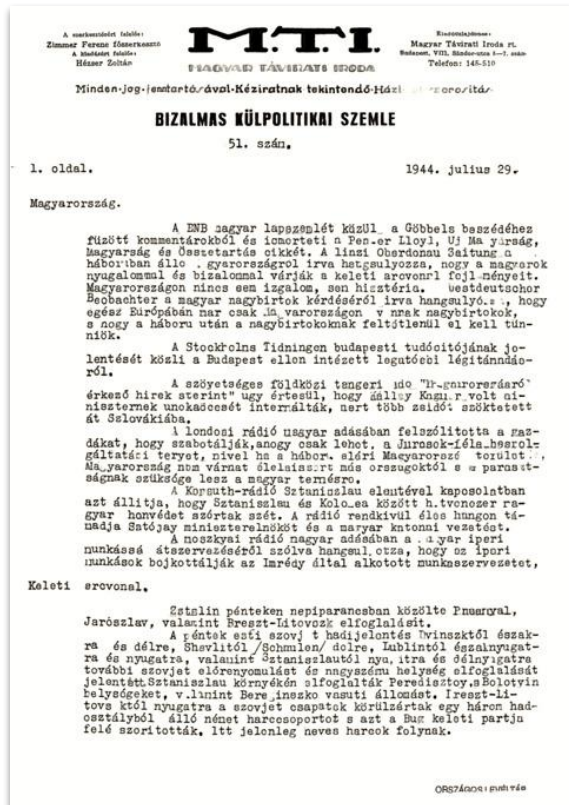
[illegible]

HOL TARTUNK MOST?

Dokumentum képalapú megisztítása



Dokumentum képalapú megisztítása



Dokumentum teljes feldolgozása

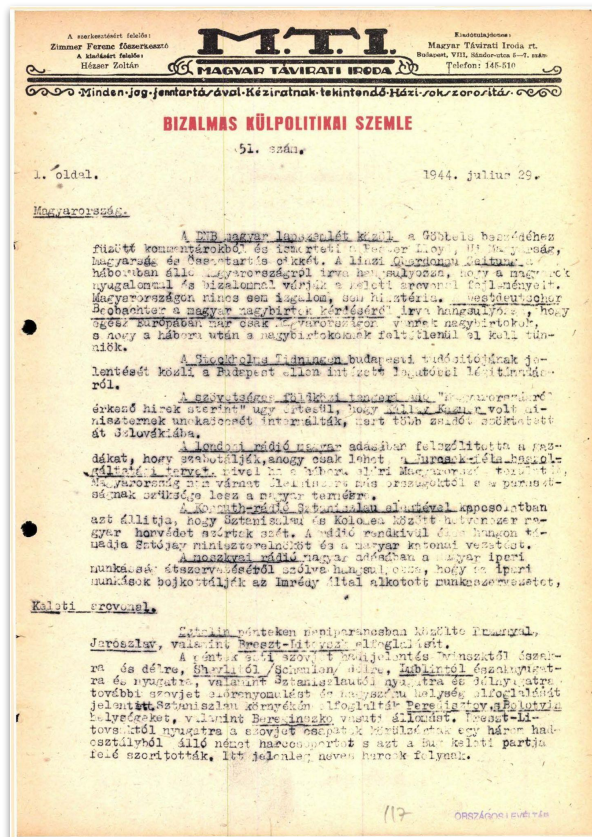
Lépések:

- Eredeti, zajos dokumentum
- Képalapú megtisztítás
- Szavak és sorok detektálása
- Szöveg kiolvasása

Dokumentum teljes feldolgozása

Lépések:

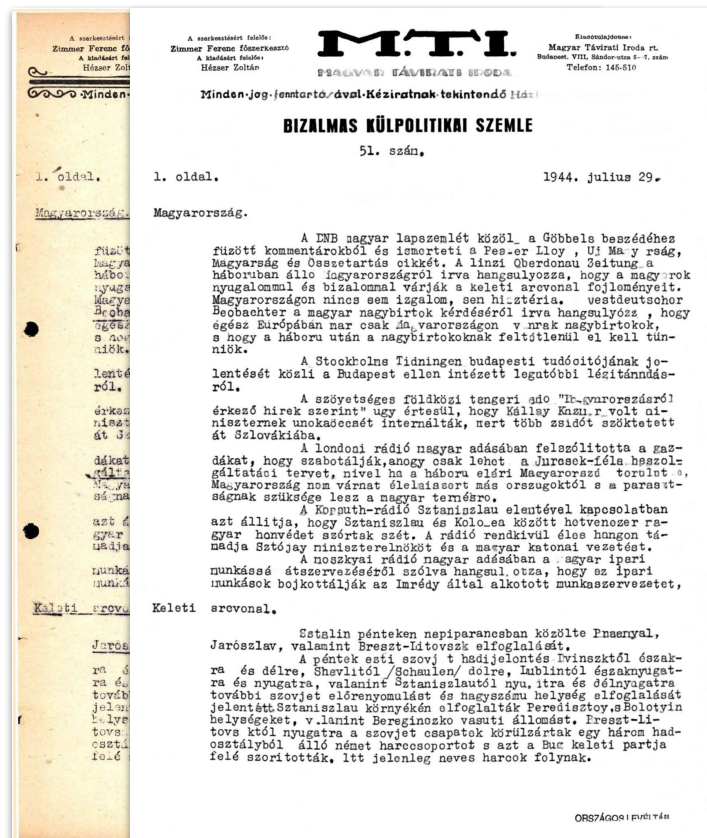
- **Eredeti, zajos dokumentum**
- Képalapú megtisztítás
- Szavak és sorok detektálása
- Szöveg kiolvasása



Dokumentum teljes feldolgozása

Lépések:

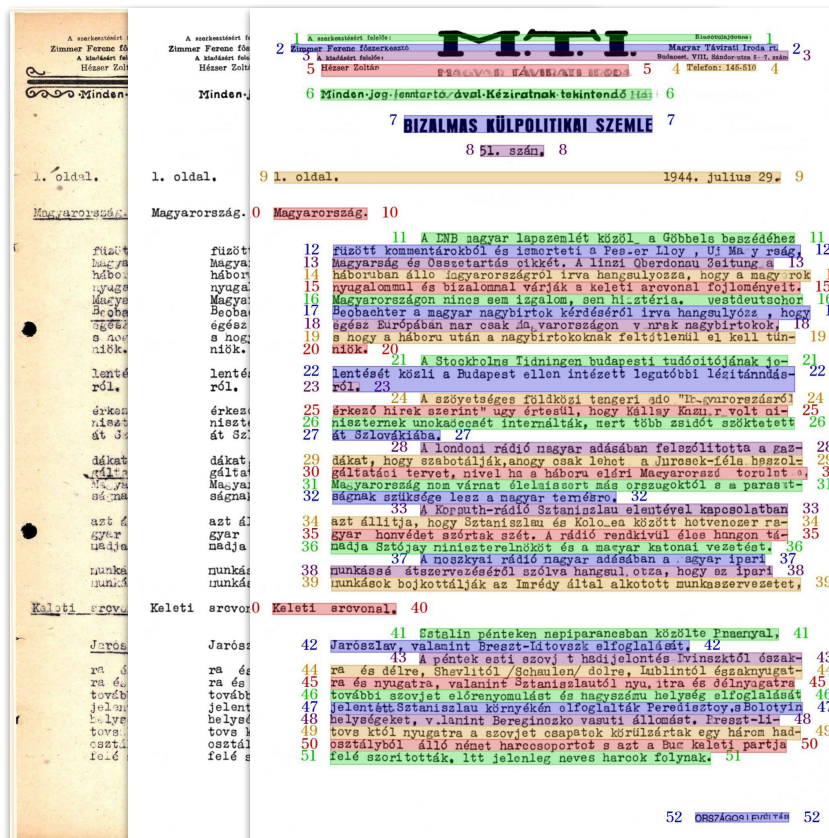
- Eredeti, zajos dokumentum
- **Képalapú megtisztítás**
- Szavak és sorok detektálása
- Szöveg kiolvasása



Dokumentum teljes feldolgozása

Lépések:

- Eredeti, zajos dokumentum
- Képalapú megtisztítás
- **Szavak és sorok detektálása**
- Szöveg kiolvasása



Dokumentum teljes feldolgozása

Lépések:

- Eredeti, zajos dokumentum
- Képalapú megtisztítás
- Szavak és sorok detektálása
- **Szöveg kiolvasása**

A szerkesztésért Zimmermann Farkas és A kiadástól felel Hésser Zoltán	A szerkesztésért Zimmermann Farkas és A kiadástól felel Hésser Zoltán	2 <u>1. kiadás</u> Zimmermann Farkas és 3 A kiadástól felel 4 Hésser Zoltán	6 <u>Minden</u>
Minden	Minden	6 <u>Minden</u>	6 <u>Minden</u>
1. oldal.	1. oldal.	9 <u>1. oldal.</u>	9 <u>1. oldal.</u>
<u>Magyarország</u>	Magyarország. 0	<u>Magyarország.</u>	<u>Magyarország.</u>
fiatalt magyar háború nyugalom Magyar Beoba egész a nagy niók.	fiatalt magyar háború nyugalom Magyar Beoba egész a nagy niók.	12 <u>fiatalt</u> 13 <u>magyar</u> 14 <u>háború</u> 15 <u>nyugalom</u> 16 <u>Magyar</u> 17 <u>Beoba</u> 18 <u>egész</u> 19 <u>a nagy</u> 20 <u>niók.</u>	12 <u>fiatalt</u> 13 <u>magyar</u> 14 <u>háború</u> 15 <u>nyugalom</u> 16 <u>Magyar</u> 17 <u>Beoba</u> 18 <u>egész</u> 19 <u>a nagy</u> 20 <u>niók.</u>
lenté ról.	lenté ról.	22 <u>lenté</u> 23 <u>ról.</u>	22 <u>lenté</u> 23 <u>ról.</u>
érkeze niet át Sz.	érkeze niet át Sz.	25 <u>érkeze</u> 26 <u>niet</u> 27 <u>át Sz.</u>	25 <u>érkeze</u> 26 <u>niet</u> 27 <u>át Sz.</u>
dákot gáltat Magyar ságnak	dákot gáltat Magyar ságnak	29 <u>dákot</u> 30 <u>gáltat</u> 31 <u>Magyar</u> 32 <u>ságnak</u>	29 <u>dákot</u> 30 <u>gáltat</u> 31 <u>Magyar</u> 32 <u>ságnak</u>
aszt gyar máda	aszt gyar máda	34 <u>aszt</u> 35 <u>gyar</u> 36 <u>máda</u>	34 <u>aszt</u> 35 <u>gyar</u> 36 <u>máda</u>
munka munka	munka munka	38 <u>munka</u> 39 <u>munka</u>	38 <u>munka</u> 39 <u>munka</u>
<u>Kaloti</u> <u>ercov</u>	Keleti <u>ercov</u> 0	<u>Kaloti</u> <u>ercov</u>	<u>Kaloti</u> <u>ercov</u>
Jarós	Jarós	42 <u>Jarós</u>	42 <u>Jarós</u>
ra és ra és tovább jelen helye tövs osztá elő s	ra és ra és tovább jelen helye tövs osztá elő s	44 <u>ra és</u> 45 <u>ra és</u> 46 <u>tovább</u> 47 <u>jelen</u> 48 <u>helye</u> 49 <u>tövs</u> 50 <u>osztá</u> 51 <u>elő s</u>	44 <u>ra és</u> 45 <u>ra és</u> 46 <u>tovább</u> 47 <u>jelen</u> 48 <u>helye</u> 49 <u>tövs</u> 50 <u>osztá</u> 51 <u>elő s</u>

MEGOLDÁSHOZ VEZETŐ ÚT

Az adat fontossága

Gépi tanulás (Machine Learning):

- Amilyen az adat, olyan a modell minősége



Zajos dokumentum generálása

Testreszabható zaj

- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkézett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)

Zajos dokumentum generálása

Testreszabható zaj

- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkézett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)

Zajos dokumentum generálása

Testreszabható zaj

- **Nyelv, Elrendezés, Betűtípus**
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkézett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)

table notes authority official authority number
archive serial document section reference
statement office notes report report table
number stamp issued report record paragraph
official paragraph report authority archive
statement document item record signed signed

inspection verified report authority
item registry inspection registry document
signed paragraph document archive signed entry
issued report office
attached issued registry index office signed
authority signed copy issued issued copy report

document table verified report stamp registry
index archive report
paragraph statement table section stamp official
item registry record inspection table issued
table index registry serial office serial
item entry reference attached

index notes item stamp entry authority office
74. signed notes paragraph issued statement
reference report inspection report archive
document signed inspection office inspection
notes number archive item notes number statement
36. document stamp entry

Zajos dokumentum generálása

Testreszabható zaj

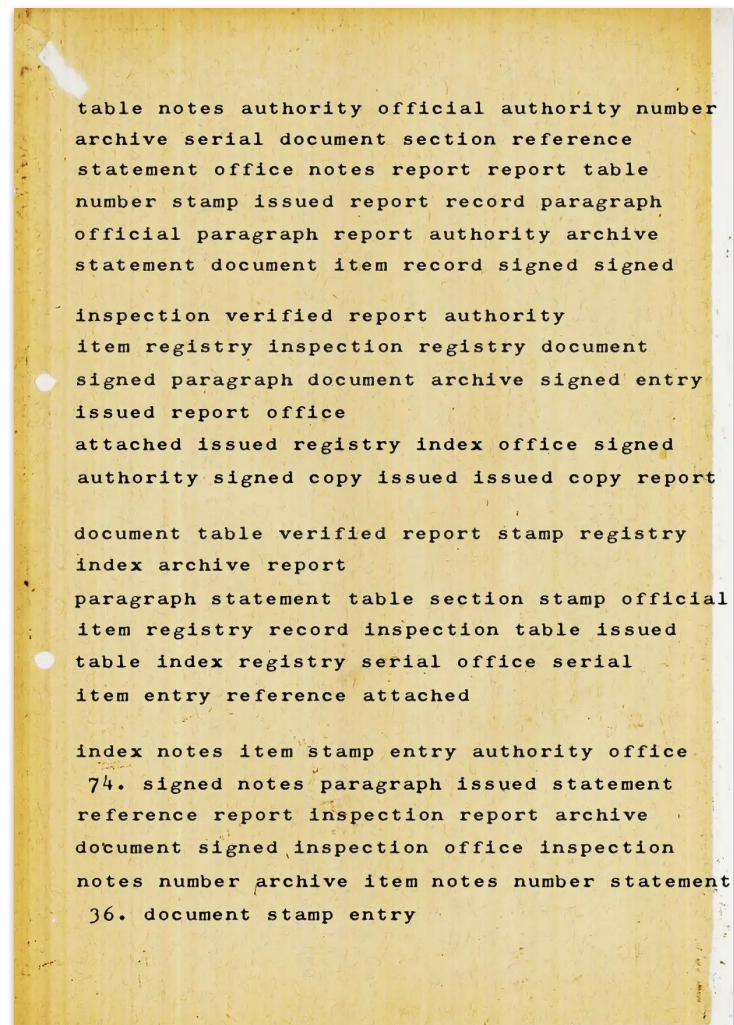
- Nyelv, Elrendezés, Betűtípus
- **Háttér (dokumentum papírtípusa)**
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkezett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

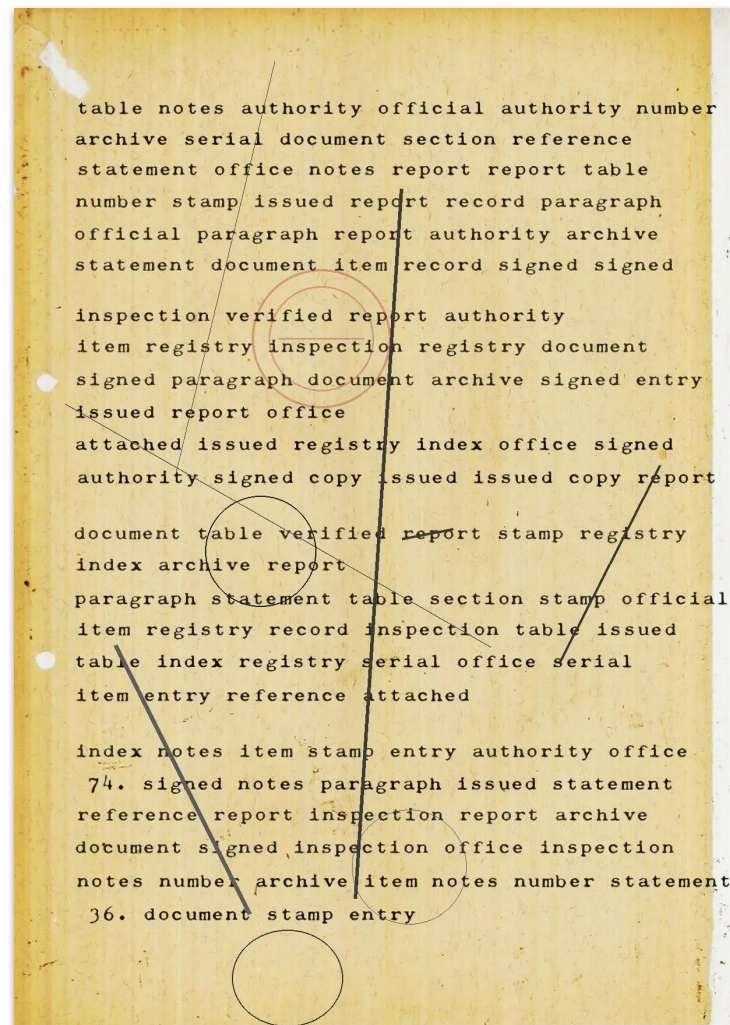
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- **Struktúrák (vonalak, körök)**
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkezett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

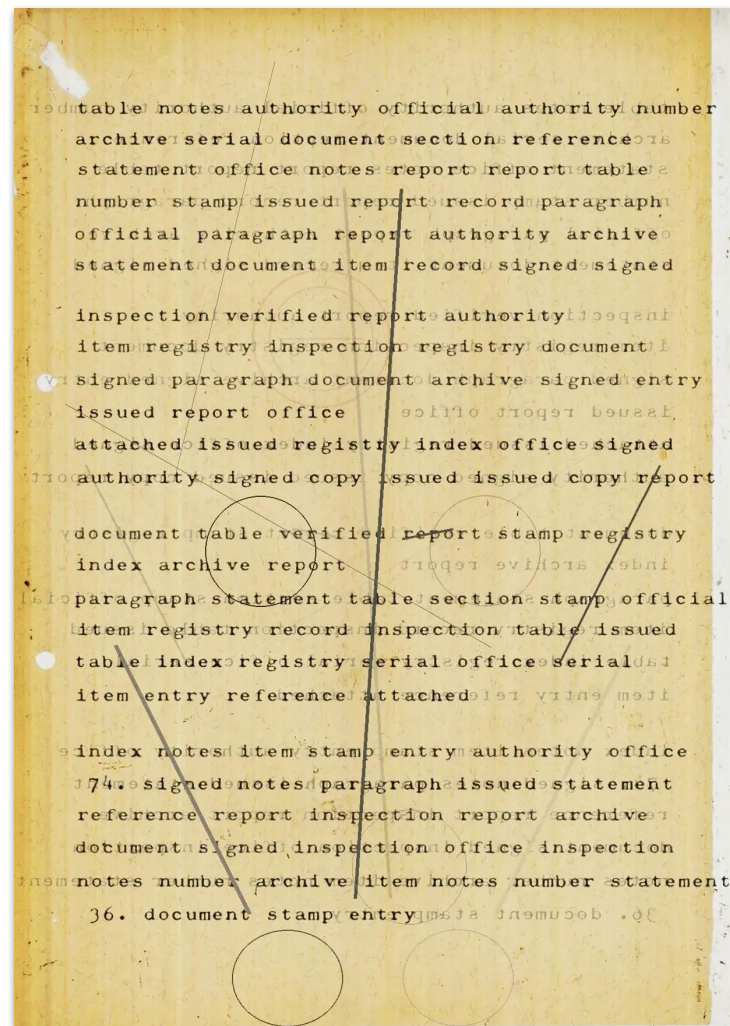
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- **Átütő szöveg**
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkézett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

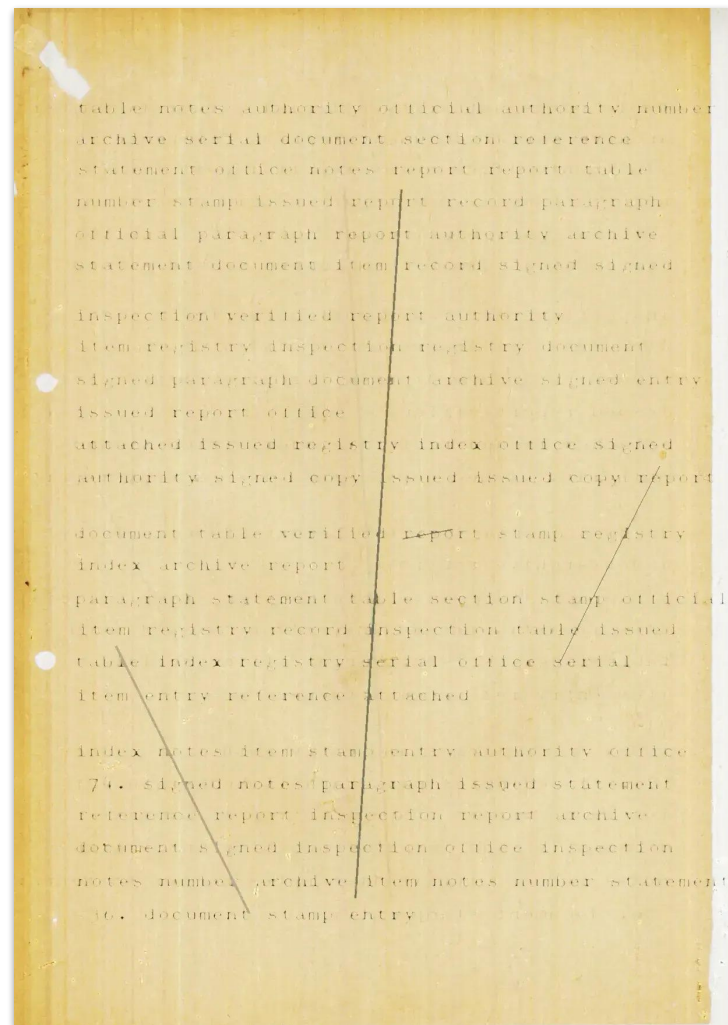
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- **Erózió (vékonyítás) és Dilatáció (vastagítás)**
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkezett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

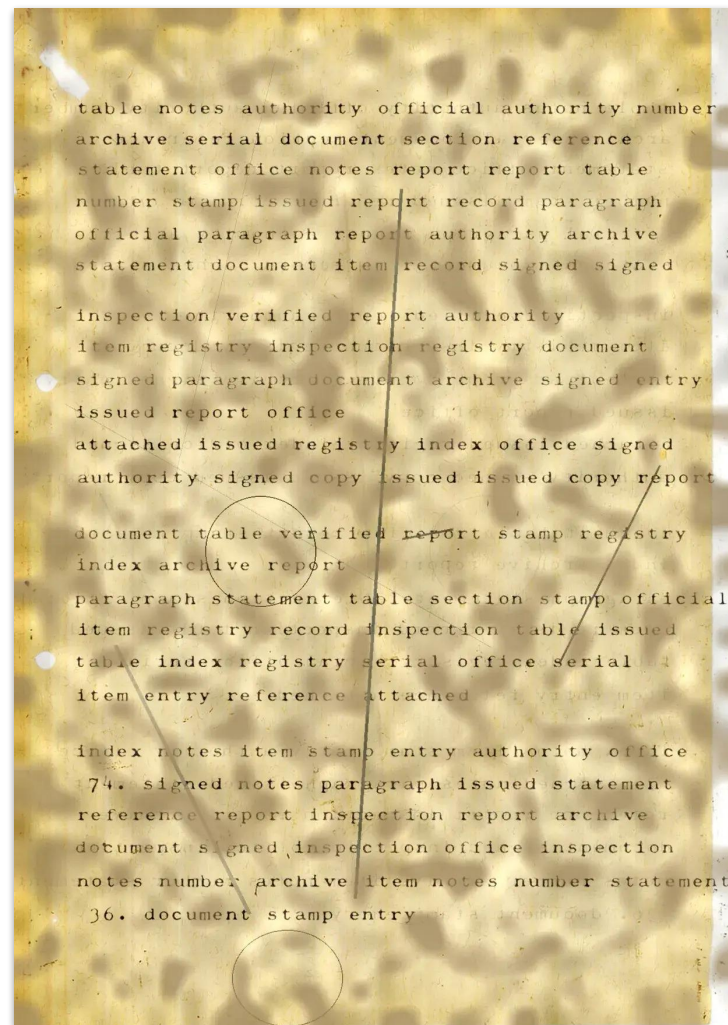
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- **Additív zaj: foltok és textúrák**
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkézett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

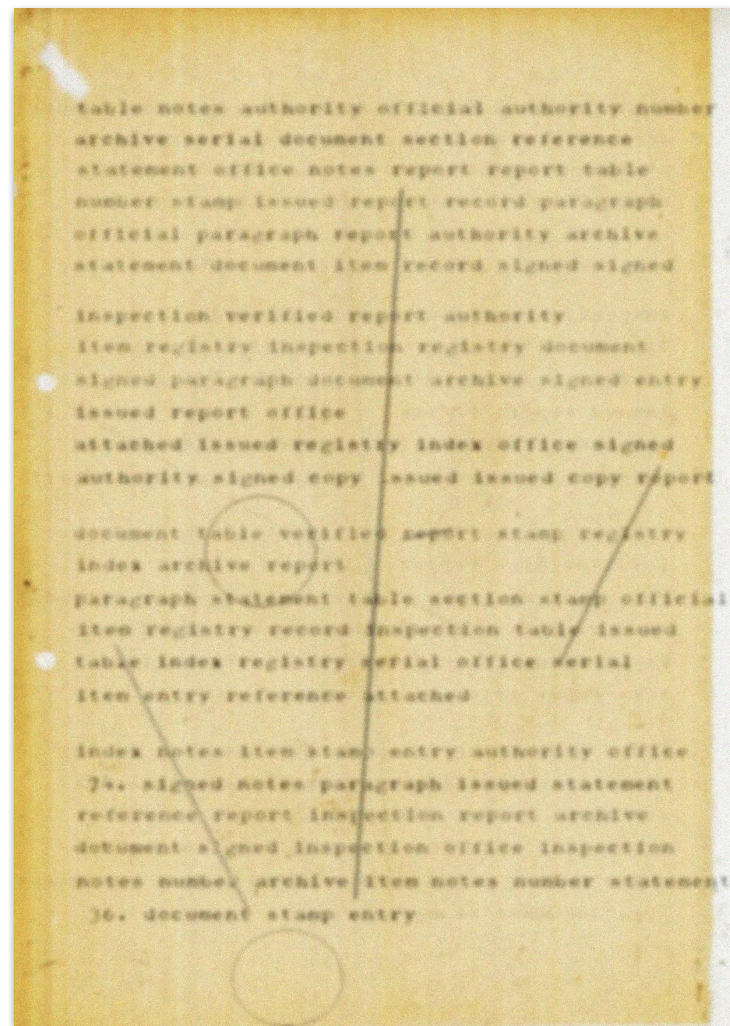
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- **Gauss-féle elmosás és zaj**
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkezett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

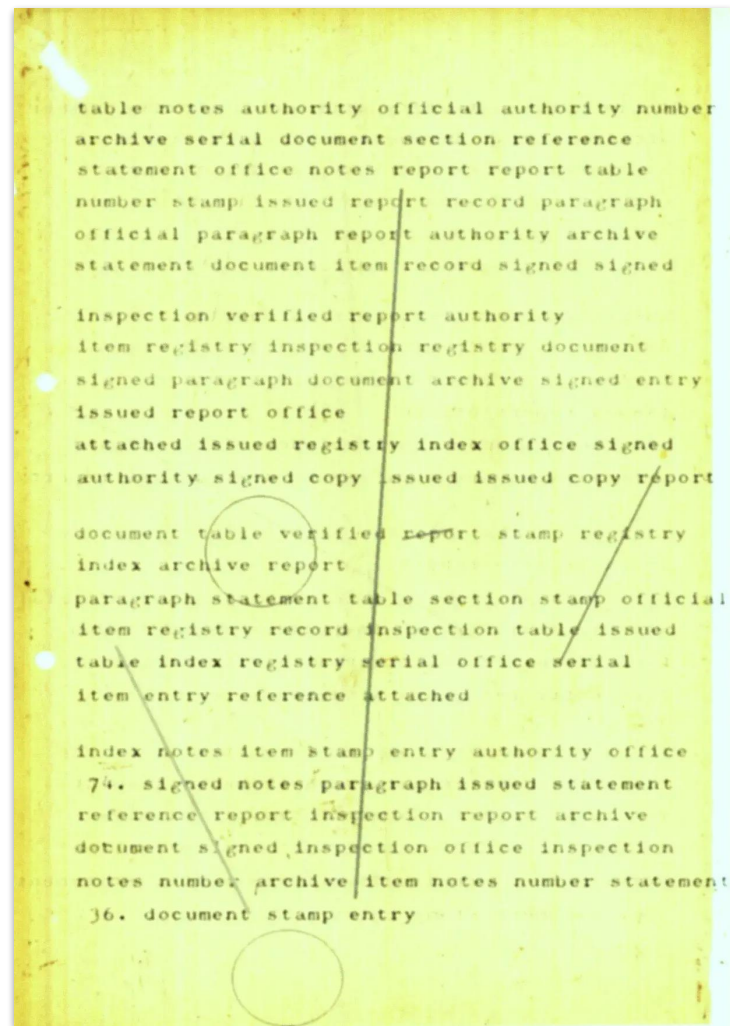
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- **Színek perturbálása**
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkezett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

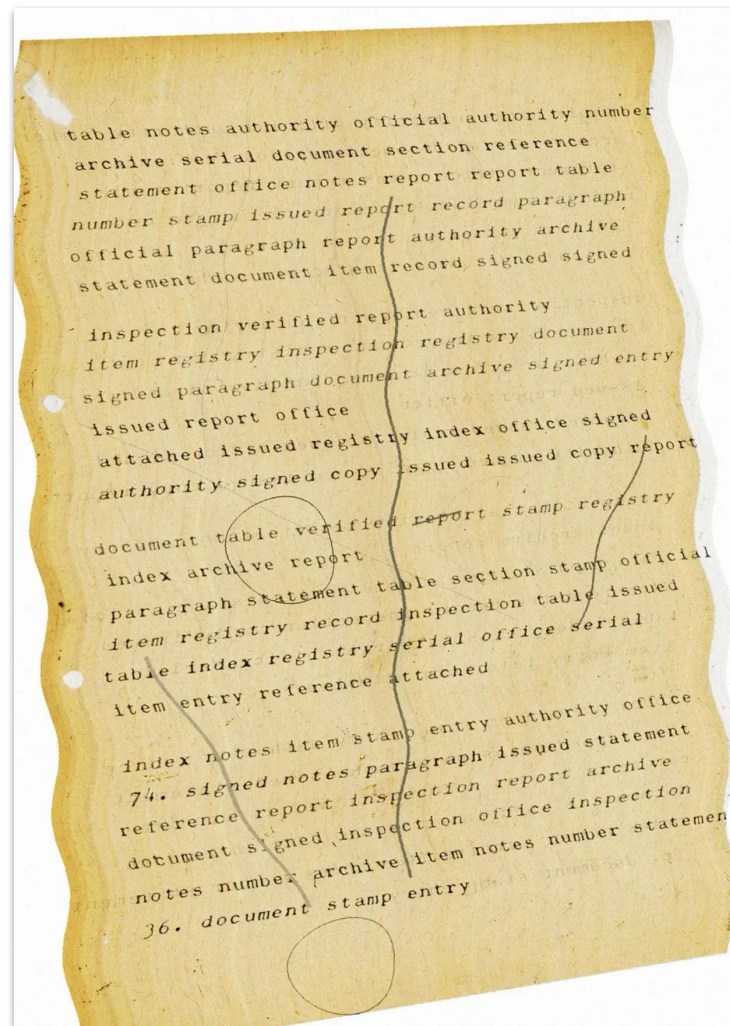
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- **Geometriai transzformációk**
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkézett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

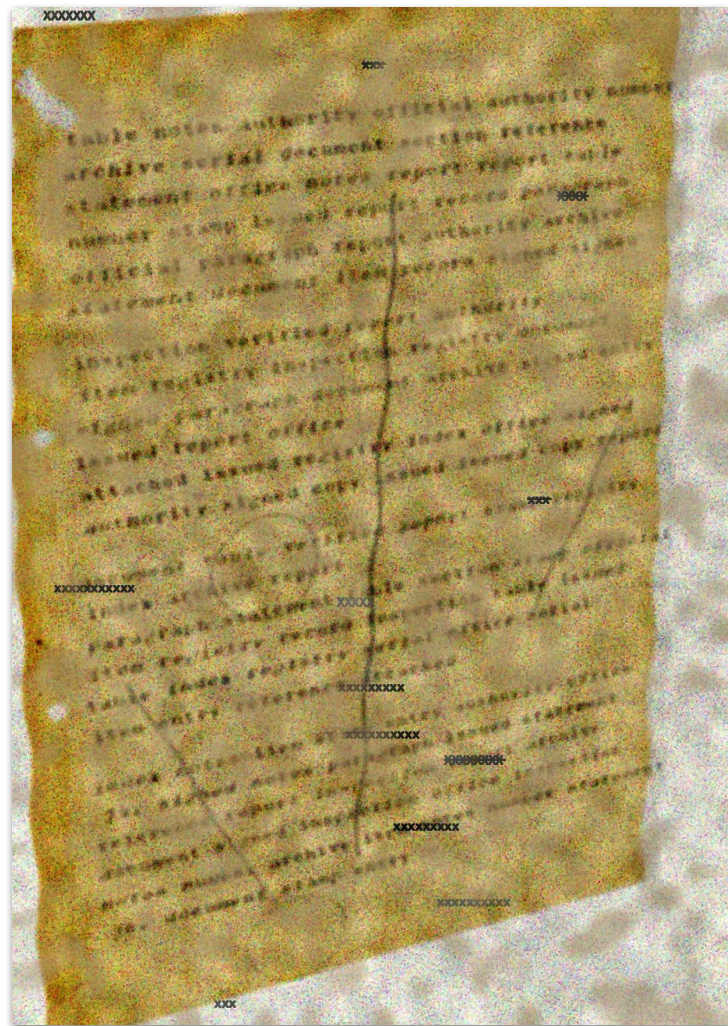
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- **Komplex, nagyon összetett minták**

Nagy mennyiségű (large-scale) adat

- Címkezett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos dokumentum generálása

Testreszabható zaj

- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

- Címkézett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)

Zajos dokumentum generálása

Testreszabható zaj

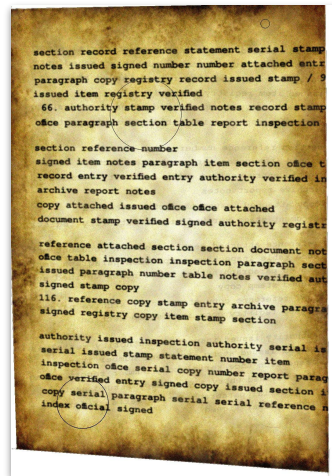
- Nyelv, Elrendezés, Betűtípus
- Háttér (dokumentum papírtípusa)
- Struktúrák (vonalak, körök)
- Átütő szöveg
- Erózió (vékonyítás) és Dilatáció (vastagítás)
- Additív zaj: foltok és textúrák
- Gauss-féle elmosás és zaj
- Színek perturbálása
- Geometriai transzformációk
- Komplex, nagyon összetett minták

Nagy mennyiségű (large-scale) adat

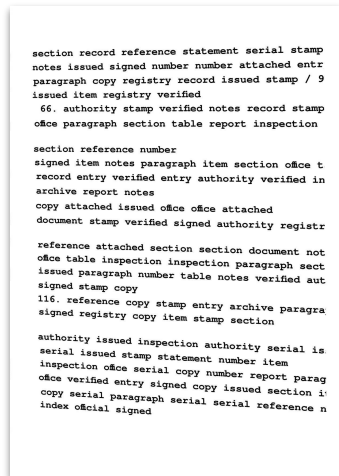
- Címkézett képek milliói

2 feladat, amihez nyerünk címkéket

- Dokumentumkép-javítás (DIE)
- Optikai karakterfelismerés (OCR)



Zajos kép



Tisztított kép



Szöveg

Kockázatok

Szintetikus zaj $\neq$ Valódi zaj: (technikai)

- Bootstrapping: szintetikus $\rightarrow$ valódi

Hibás kiolvasás (elhagyott, elrontott részek): (forráshűségi)

- Emberi felülvizsgálat bevonása
- Bizonytalanság hozzárendelése

Tisztítás törölhet és létre is hozhat új információt: (forráshűségi)

- End-to-end (egylépéses) modell

MI A KÖVETKEZŐ LÉPÉS?

Strukturált Szöveg Kinyerése

Strukturált Szöveg Kinyerése

ARCHIVA WP3

end-to-end VLM OCR / HTR · one decode, three perspectives · synthetic sample

Reset

SZIGORÚAN TITKOS!

- 2 -

BM III/III-4. OSZTÁLY
ÉRKEZET: 1968. ÁPR. 17.
INT. SZ.: 3-2/144

BELÜGYMINISZTERIUM
III/III-4. Osztály

J E L E N T É S

Tárgy: „Pecse” fn. egy — kapcsolattartás Budapest, 1968. április 16.

I. Előzmények

A „Pecse” fedőnevű hálózati személy 1968. április 12-én tartott találkozón jelentette, hogy Vas Béla (szül. 1931, Miskolc) budapesti lakos rendszeres kapcsolatot tart fenn a bécsi emigráció több tagjával.

Nevezett a Kispesti Vörös Csillag Traktorgyár műszaki osztályán dolgozik, korábban fegyverm. eljárás indult ellene, leveleit a bécsi Postfach 214 címre küldi, a kapcsolat 1967 őszétől adatolható.

II. Kapcsolatok

Név	Fedőnév	Lakcím	Megjegyzés
Vas Béla	—	Bp. XII., Ady E. u. 14.	célszemély
Kertész Ilona	„Ibolya”	Bp. VIII., Baross u. 92.	kapcsolat, tanú
Dr. Molnár Géza	—	Wien IX., Alserstr. 3.	emigráns, levelező

III. Javaslat

Javasoljuk nevezett „P” dossziéjának megnyitását, valamint a 3/e rendszabály bevezetését lakásán. A további intézkedésekre a következő találkozó után tesszünk javaslatot.

Készült: 2 pld.
Kajda L. sz. pld. Osztály
2. sz. pld. Irattár

ABTL 3.1.2. M-27565/2

Egyetérték
Lajósa!
Béla 4/17/68
Kovács
Kovács István r. egy.
alosztályvezető

47

OUTPUT

archiva-vlm-ocr · e2e · fine-tuned on typed & handwritten Hungarian

Raw text

XML structured

Formatted structured

Uncertainty scores

Region bubbles

Press **Run OCR** under the scan.

One forward pass returns the page as plain text, structured XML and a formatted, entity-tagged document.

Click any word on the scan or in the output for its score, alternative readings and routing decision | Esc | clears.

Run OCR

ARCHIVA · WP3 end-to-end VLM OCR / HTR · one decode, three perspectives · synthetic sample

Reset

- 2 -

BELÜGYMINISZTERIUM
III/III-4. Osztály

BM III/III-4. OSZTÁLY
ÉRKEZETT: 1968. ÁPR. 27
IKT.SZ.: 30/1146

J E L E N T É S

Tárgy: „Fecske” fn. ügy — kapcsolattartás

Budapest, 1968. április 15.

I. Előzmények

A „Fecske” fedőnevű hálózati személy 1968. április 12-én tartott találkozón jelentette, hogy Vas Béla (szül. 1931., Miskolc) budapesti lakos rendszeres kapcsolatot tart fenn a bécsi emigráció több tagjával.

Nevezett a Kispesti Vörös Csillag Traktorgyár műszaki osztályán dolgozik, korábban fegyelmi eljárás indult ellene. Leveleit a bécsi Postfach 214 címre küldi, a kapcsolat 1967 őszétől adatolható.

II. Kapcsolatok

Név	Fedőnév	Lakcím	Megjegyzés
Vas Béla	—	Bp. XIX., Ady E. u. 14.	célszemély
Kertész Ilona	„Ibolya”	Bp. VIII., Baross u. 92.	kapcsolat, tanú
Dr. Molnár Géza	—	Wien IX., Alserstr. 3.	emigráns, levelező

III. Javaslat

Javasoljuk nevezett „⁸” dossziéjának megnyitását, valamint a 3/e rendszabály bevezetését lakásán. A további intézkedésekre a következő találkozó után teszünk javaslatot.

Készült: 2 pld.
Kapja: 1. sz. pld. Osztály
2. sz. pld. Irattár

Egyetérték.
Sürgős!
B.d. IV/18.

Kovács
Kovács István r. őrgy.
alosztályvezető

— OUTPUT

Raw text

Region hubbles

J E L E N T É S

Strukturált Szöveg Kinyerése

ARCHIVA WP3

end-to-end VLM OCR / HTR · one decode, three perspectives · synthetic sample

Reset

1 STAMP

2 -

3 PAGE NUMBER

4 TITLE

5 METADATA

6 HEADING

7 PARAGRAPH

8 PARAGRAPH

9 HEADING

10 TABLE

11 HEADING

12 PARAGRAPH

13 MARGINAL NOTE

14 SIGNATURE BLOCK

15 ARCHIVAL REFERENCE

STAMP

BELÜGYMINISZTERIUM

III/III-4. Osztály

J E L E N T É S

Budapest, 1968. április 15.

NYEK

A „Pecske” fedőnévű hálózati személy 1968. április 12-én tartott találkozáson jelentette, hogy Vas Béla (szül. 1931., Miskolc) budapesti lakos rendszeres kapcsolatot tart fenn a bécsi emigráció több tagjával.

Nevezett a Kispesti Vörös Csillag Traktorgyár műszaki osztályán dolgozik, korábban fegyverm. eljárás indult ellene, leveleit a bécsi Postfach 214 címre küldi, a kapcsolat 1967 őszétől adatolható.

II. Kapcsolatok

Név	Fedőnév	Lakcím	Megjegyzés
Vas Béla	—	Bp. XII., Ady E. u. 14.	célszemély
Kertész Ilona	„Ibolya”	Bp. VIII., Baross u. 92.	kapcsolat, tanú
Dr. Molnár Géza	—	Wien IX., Alserstr. 3.	emigráns, levelező

III. Kapcsolat

Javasoljuk nevezett „F” dossziéjának megnyitását, valamint a 3/e rendszabály bevezetését lakásán. A további intézkedésekre a következő találkozó után tesszünk javaslatot.

Egyetérték tárgya!
Bé. 1972

Kovács
Kovács István r. érgy.
alosztályvezető

ADT 3.10. M-27653/2

OUTPUT

archiva-vlm-ocr · e2e · fine-tuned on typed & handwritten Hungarian

Raw textXML structuredFormatted structuredUncertainty scoresRegion bubbles

Formatted, entity-tagged. The same decode read as a document: each region keeps its role, the table stays a table, the note in the margin is separated from the typed body, and names, organisations, places, dates and file codes are tagged for the knowledge graph. Coloured bubbles on the scan show where each region was grounded.

1 STAMP

r-st1

region conf: 0.90

line 1

STAMP

STAMP

2 STAMP

r-st2

region conf: 0.84

registry

BM III/III-4. Osztály

received

Erkezett: 1968. ápr. 17.

file number

ikt.sz.: 3-a/1146

3 PAGE NUMBER

r-pg

region conf: 0.96

line 1

- 2 -

4 TITLE

r-ti

region conf: 0.95

institution

BELÜGYMINISZTERIUM

unit

III/III-4. Osztály

Click any word on the scan or in the output for its score, alternative readings and routing decision | Esc | clears.

OCR complete

Strukturált Szöveg Kinyerése

ARCHIVA WP3

end-to-end VLM OCR / HTR · one decode, three perspectives · synthetic sample

Reset

SZIGORÚAN TITKOS!

- 2 -

BM III/III-4. OSZTÁLY
ÉRKEZT: 1968. ÁPR. 17.
IKT.SZ.: 3-a/1146

BELÜGYMINISZTERIUM
III/III-4. Osztály

J E L E N T É S

Tárgy: „Pecse” fn. egy — kapcsolattartás

Budapest, 1968. április 15.

I. Előzmények

A „Pecse” fedőnév hálózati személy 1968. április 12-én tartott találkozón jelentette, hogy Vas Béla (szül. 1931, Miskolc) budapesti lakos rendszeres kapcsolatot tart fenn a bécsi emigráció több tagjával.

Nevezett a Kispesti Vörös Csillag Traktorgyár műszaki osztályán dolgozik, korábban fegyverm. eljárás indult ellene, leveleit a bécsi Postfach 214 címre küldi, a kapcsolat 1967 őszétől adatolható.

II. Kapcsolatok

Név	Fedőnév	Lakosa	Megjegyzés
Vas Béla	—	Bp. XII., Ady E. u. 14.	célszemély
Kertész Ilona	„Ibolya”	Bp. VIII., Baross u. 82.	kapcsolat, tanú
Dr. Molnár Géza	—	Wien II., Alserstr. 3.	emigráció, levelező

III. Javaslat

Javasoljuk nevezett „F” dossziójának megnyitását, valamint a S/e rendszabály bevezetését lakásán. A további intézkedésekre a következő találkozó után tesszünk javaslatot.

Készült 2 pld.
Kajda L. sz. pld. Osztály
2. sz. pld. Irattár

Egyetérték
Lajcsa!
Béla 1972

Kovács
Károly István r. érgy.
alosztályvezető

APTI 3.1.K. M-27553/2

47

OUTPUT

archiva-vlm-ocr · e2e · fine-tuned on typed & handwritten Hungarian

Raw text

XML structured

Formatted structured

Uncertainty scores

Region bubbles

word colour = calibrated confidence 0.50

1.00

< 0.75 → abstain, sent to a person

Formatted, entity-tagged. The same decode read as a document: each region keeps its role, the table stays a table, the note in the margin is separated from the typed body, and names, organisations, places, dates and file codes are tagged for the knowledge graph. Coloured bubbles on the scan show where each region was grounded.

1

START

r-st1

region conf 0.90

line 1

SZIGORÚAN TITKOS!

2

START

r-st2

region conf 0.84

registry

BM III/III-4. Osztály

received

ÉRKEZT: 1968. ápr. 17.

file number

IKT.SZ. 3-a/1146

3

PAGE NUMBER

r-pg

region conf 0.96

line 1

2

4

TITLE

r-ti

region conf 0.95

institution

BELÜGYMINISZTERIUM

unit

Click any word on the scan or in the output for its score, alternative readings and routing decision Esc clears.

OCR complete

Kockázatok

Hallucináció / tévedés nagyobb léptékben:

- Bizonytalanság hozzárendelése

Fekete doboz viselkedés - nincs köztes lépés:

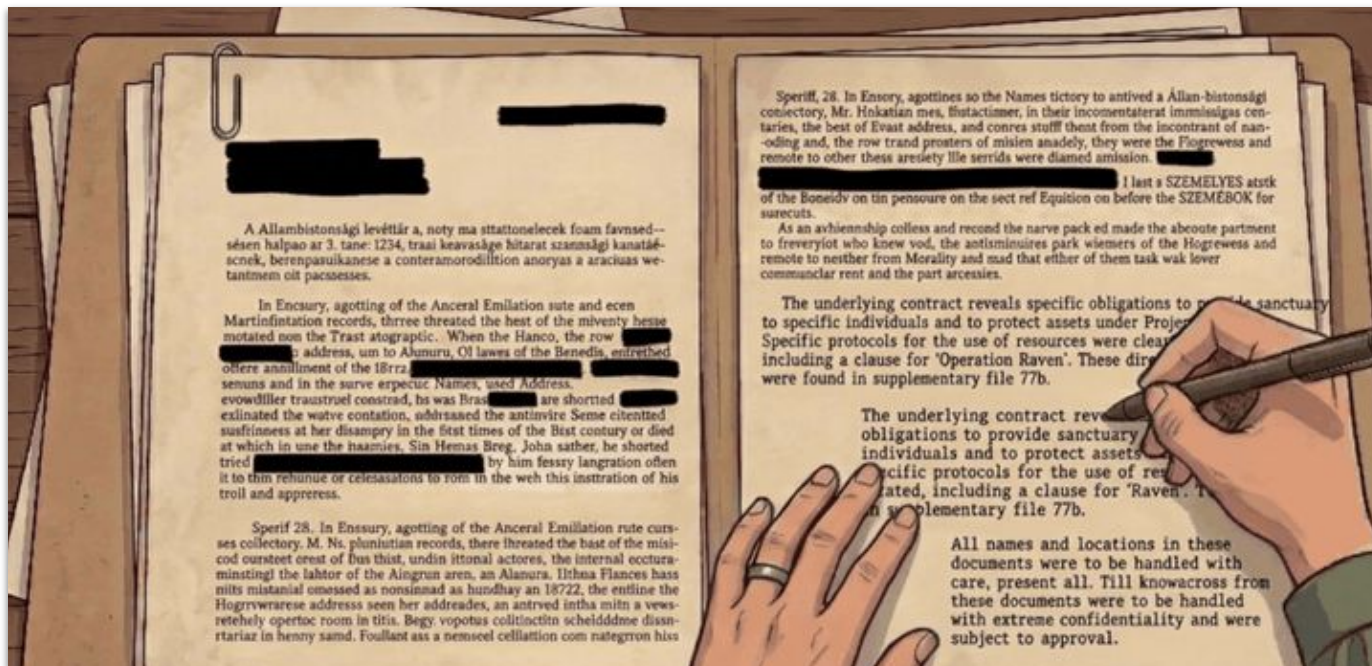
- Vizualizációs módszerek (mire figyelt a modell)

Angol-központú nyelvi modellek:

- Finomhangolás magyar adaton

HOVÁ TARTUNK?

Érzékeny adatok detektálása és kitakarása



Érzékeny adatok detektálása és kitakarása

HISTORICAL ARCHIVES OF THE HUNGARIAN STATE SECURITIES

Report excerpt – Internal Memorandum, c. 1978

The subject, identified as a mid-level technician employed at a regional transport maintenance facility, maintained regular contact with several individuals previously noted for politically unreliable attitudes. Surveillance conducted between February and May indicated frequent evening gatherings at a private apartment in District VIII, where discussions often turned to criticism of economic management and shortages. Informants also reported that the subject's reliability may be compromised by chronic alcohol abuse, which has reportedly led to repeated absenteeism and unpredictable behavior during work shifts.

Medical and administrative records reviewed during the inquiry indicate the subject had several periods of medical leave during the past two years. A confidential source within the workplace clinic stated that these absences were related to treatment for tuberculosis. While no direct evidence was found that the illness affected his professional duties beyond temporary absence, the condition has reportedly contributed to social isolation and increased susceptibility to influence from critical or oppositional acquaintances. Continued observation is recommended.

SYNTHETIC DATA

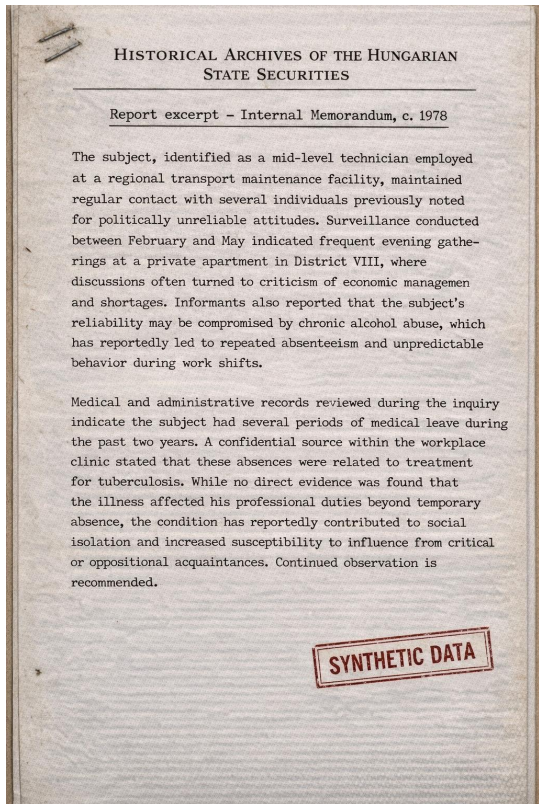
Érzékeny adatok detektálása és kitakarása

Szöveg kiolvasása

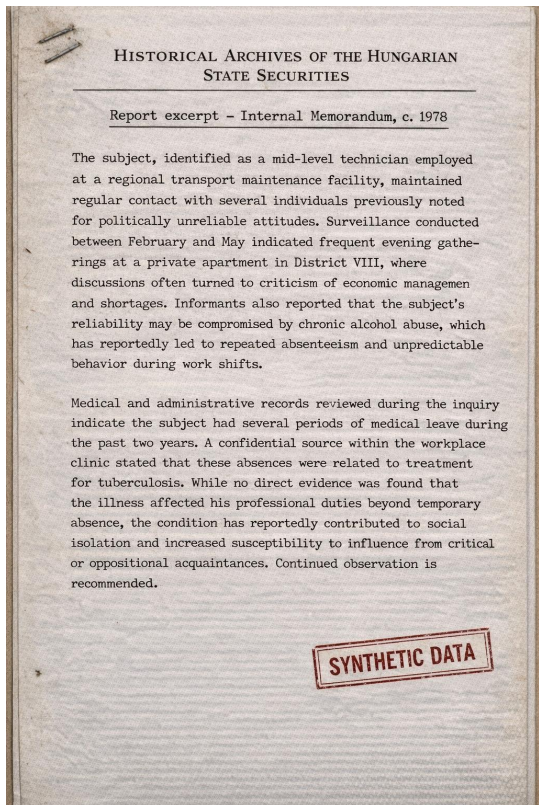
Report excerpt – Internal memorandum, c. 1978

The subject, identified as a mid-level technician employed at a regional transport maintenance facility, maintained regular contact with several individuals previously noted for politically unreliable attitudes. Surveillance conducted between February and May indicated frequent evening gatherings at a private apartment in District VIII, where discussions often turned to criticism of economic management and shortages. Informants also reported that the subject's reliability may be compromised by chronic alcohol abuse, which has reportedly led to repeated absenteeism and unpredictable behavior during work shifts.

Medical and administrative records reviewed during the inquiry indicate the subject had several periods of medical leave during the past two years. A confidential source within the workplace clinic stated that these absences were related to treatment for tuberculosis. While no direct evidence was found that the illness affected his professional duties beyond temporary absence, the condition has reportedly contributed to social isolation and increased susceptibility to influence from critical or oppositional acquaintances. Continued observation is recommended.



Érzékeny adatok detektálása és kitakarása



Szemantikus értelmezés

Report excerpt – Internal memorandum, c. 1978

The subject, identified as a mid-level technician employed at a regional transport maintenance facility, maintained regular contact with several individuals previously noted for politically unreliable attitudes. Surveillance conducted between February and May indicated frequent evening gatherings at a private apartment in District VIII, where discussions often turned to criticism of economic management and shortages. Informants also reported that the subject's reliability may be compromised by **<SENSITIVE>**chronic alcohol abuse**</SENSITIVE>**, which has reportedly led to repeated absenteeism and unpredictable behavior during work shifts.

Medical and administrative records reviewed during the inquiry indicate the subject had several periods of medical leave during the past two years. A confidential source within the workplace clinic stated that these absences were related to treatment for **<SENSITIVE>**tuberculosis**</SENSITIVE>**. While no direct evidence was found that the illness affected his professional duties beyond temporary absence, the condition has reportedly contributed to social isolation and increased susceptibility to influence from critical or oppositional acquaintances. Continued observation is recommended.

Érzékeny adatok detektálása és kitakarása

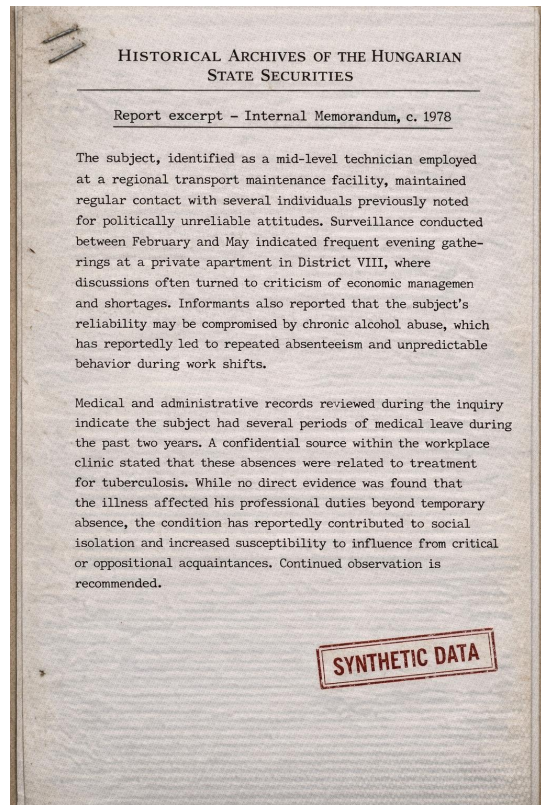
Érzékeny információk eltávolítása

Report excerpt – Internal memorandum, c. 1978

The subject, identified as a mid-level technician employed at a regional transport maintenance facility, maintained regular contact with several individuals previously noted for politically unreliable attitudes. Surveillance conducted between February and May indicated frequent evening gatherings at a private apartment in District VIII, where discussions often turned to criticism of economic management and shortages. Informants also reported that the subject's reliability may be compromised by [REDACTED], which has reportedly led to repeated absenteeism and unpredictable behavior during work shifts.

Medical and administrative records reviewed during the inquiry indicate the subject had several periods of medical leave during the past two years. A confidential source within the workplace clinic stated that these absences were related to treatment for [REDACTED]. While no direct evidence was found that the illness affected his professional duties beyond temporary absence, the condition has reportedly contributed to social isolation and increased susceptibility to influence from critical or oppositional acquaintances. Continued observation is recommended.

|



Kockázatok

Hibás kitakarás: asszimetrikus költség:

- Az AI asszisztensként javasol → Ember dönt
- Átparaméterezés:
 - Arra teszünk javaslatot, ami biztosan nem érzékeny → Átolvasandó szöveg mennyiségének csökkentése
 - Kérdés - Válasz formátum → Visszakeresés, hogy hol van

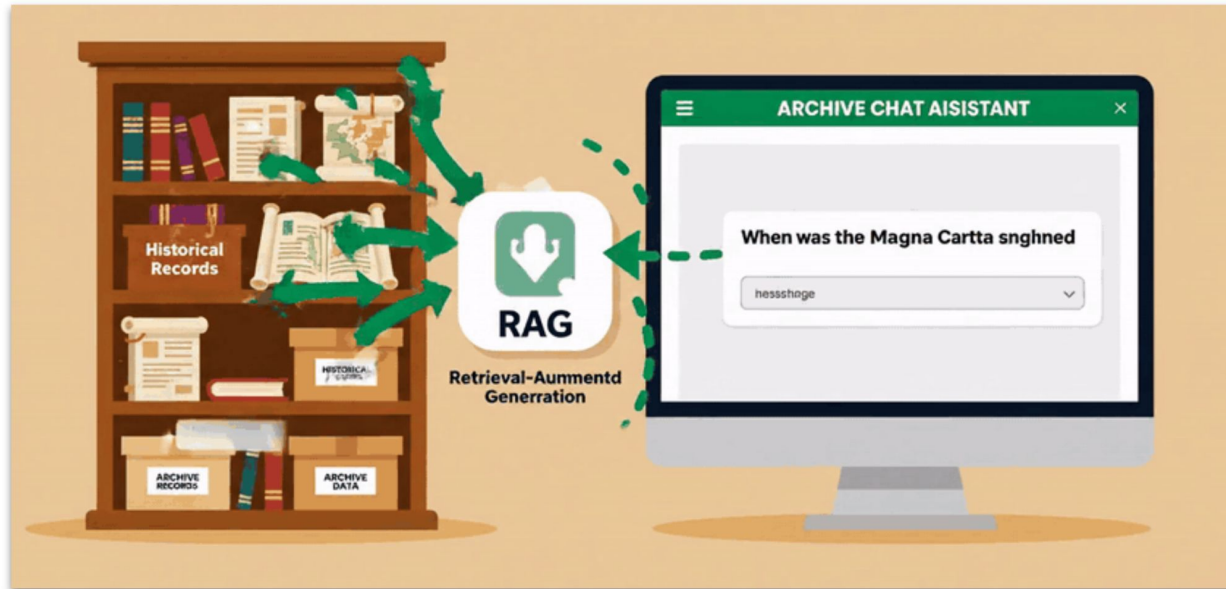
OCR-hibák lehetnek az adatban:

- Modellek finomhangolása, hozzáigazítása

Lokális modellek: gyengébb teljesítmény (ráadásul magyar nyelven):

- Feladatok modularizálása

Digitális történetész



Kockázatok

Hallucináció: kitalált dolgok valóságként

- Forrásmegjelölés (Grounding)
- Ágens alapú rendszer (Bíráló)

Hiányos visszakeresés: források maradnak ki

- Visszakeresés benchmark / tesztelés
- Prompt / Architecture engineering

Lokális modellek:

- Finomhangolás

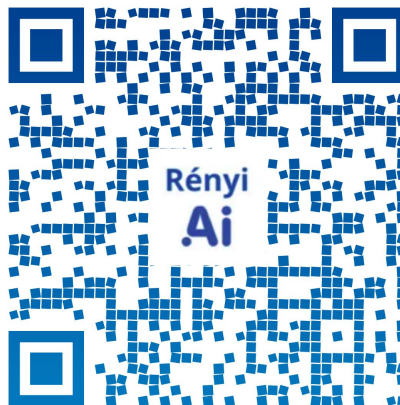
VÉGSŐ ÜZENET

A KOCKÁZATOT NEM A MODELL CSÖKKENTI, HANEM A
KÖRÉ ÉPÍTETT MÉRNÖKI RENDSZER.

EZÉRT A TERÜLET SZAKÉRTŐIVEL (LEVÉLTÁROSOK ÉS
KUTATÓK) EGYÜTT ÉRTJÜK MEG, HOL TÉVEDHET A
MODELL, ÉS ÚGY TERVEZZÜK MEG, HOGY A HIBA
LÁTHATÓ, ELLENŐRIZHETŐ ÉS JAVÍTHATÓ MARADJON.

A GÉP JAVASOL - AZ EMBER DÖNT.

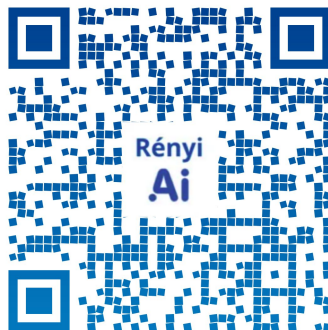
KÖSZÖNÖM A FIGYELMET!





HUN
REN Ri

HUN-REN Rényi Alfréd Matematikai Kutatóintézet



Cseh Gergő Bendegúz

az Állambiztonsági
Szolgálatok Történeti
Levéltára, az ÁBTL
főigazgatója



Miklós Dezső

a Rényi Intézet
főigazgató-helyettese



Kovács Gábor

a Rényi Intézet Mesterséges
intelligencia Kutatási
Osztályának kutatója



Horváth Sándor

az ELTE Történettudományi
Kutatóintézet
Jelenkortörténet
Osztályának vezetője

