

Co nesmí do AI. Kontrola dřív, než stisknete odeslat.

Účetní vloží do chatu celou smlouvu kvůli shrnutí. Shrnutí je dobré. Rodné číslo a číslo účtu dvou lidí ale skončily u služby, se kterou firma nemá smlouvu, a v rozsahu, který shrnutí vůbec nepotřebovalo. Tenhle systém text před vložením projede, najde v něm, co tam nepatří, a vydá kopii se značkami místo identifikátorů. S tou se dá pracovat dál. Rodné číslo, IČO, IBAN, číslo účtu a kartu neodhaduje: ověřuje je kontrolní číslicí.

Co dostáváte.

BALÍČEK

Hotový systém, ne seznam rad

Jeden soubor `udaje.py` obsahuje třináct detektorů, redakci, nález a strážce pro AI nástroj. Bez závislostí, bez API klíče, bez sítě. Nic neodchází z vašeho počítače, což je u tohoto úkolu podstatné.

DENNÍ POUŽITÍ

Jeden příkaz před vložením

```
cat soubor | python3 udaje/udaje.py text
```

vypíše text s nahrazenými identifikátory. To, co vypadne, vložíte do chatu. Funguje to i bez instalace a je to ta část, kterou budete používat nejčastěji.

STRÁŽCE

Hranice, která není jen doporučení

Hook zastaví AI nástroj, když si chce přečíst soubor s ověřeným rodným číslem, číslem účtu nebo klíčem, a pošle ho k redigované kopii. Je to jediná část balíčku, která tomu opravdu zabrání; zbytek stojí na disciplíně.

■ VĚTA, NA KTERÉ TO STOJÍ

Nálezy se maskují i před vámi. Celou hodnotu systém nevypíše nikdy, ani do vlastního nálezu.

Nástroj, který najde rodná čísla a pak je vypíše do reportu, problém jen přesunul. Tenhle v nálezu uvádí typ, místo a maskovaný úryvek. Ověřuje to příkaz kontrola: hledá skutečné hodnoty ve všech svých výstupech.

MECHANISMUS

Čtyři kroky a jedna věc, kterou uděláte dnes.

Systém má dvě poloviny. Jedna je na text, který právě chcete někam vložit. Druhá je na složku s podklady, kterou projedete jednou a pak ji hlídáte.

- 1 Text na vstupu.** Protáhnete text příkazem a dostanete ho zpět s nahrazenými identifikátory. Na chybový výstup se vypíše, co bylo nahrazeno, co zůstalo k vašemu rozhodnutí a co se nenahrazuje, protože se to nemá vkládat vůbec.
- 2 Složka s podklady.** Příkaz projde soubory, u každého nálezu zapíše typ, soubor, řádek, sloupec a maskovaný úryvek, a vedle toho vyrobí redigované kopie se stejnou strukturou složek.
- 3 Strážce.** Hook v AI nástroji zastaví čtení i zápis souboru, ve kterém jsou ověřené osobní nebo přístupové údaje, a odkáže na redigovanou kopii. Vrací návratový kód 2, takže nástroj nemůže pokračovat.
- 4 Pravidlo do firemních pravidel.** Z toho, co se u vás skutečně našlo, vypíše oddíl „Co nevkládáme do AI“. Vložíte ho do pravidel z balíčku Pravidla pro práci s AI, nebo kamkoli, kde pravidla máte.

CO NAJDE JEDEN BĚH NA MODELOVÉ FIRMĚ (SKUTEČNÝ VÝSTUP)

```
$ python3 udaje/udaje.py projdi podklady
```

Běh 1: prohlédnuto 11 souborů, nálezů 63.

Číslo bankovního účtu: 2 (ověřeno)

IBAN: 1 (ověřeno)

Číslo platební karty: 1 (ověřeno)

Heslo, token nebo klíč: 2 (ověřeno)

Rodné číslo: 5 (ověřeno)

E-mailová adresa: 9 (pravděpodobné)

IČO: 5 (pravděpodobné)

Mzdový nebo exekuční údaj: 5 (pravděpodobné)

Označení důvěrnosti: 10 (pravděpodobné)

Zdravotní údaj: 7 (pravděpodobné)

Adresa s PSČ: 2 (nejisté)

Možné jméno osoby: 7 (nejisté)

Telefonní číslo: 7 (nejisté)

Redigované kopie k vložení do AI: redigovano/b1 (8 souborů)

Nález: nalezy/n1.md

nekontrolováno: skeny/predavaci_protokol.pdf - binární soubor (obrázek, PDF, archiv); text z něj tenhle skript nečte

Jeden soubor podrobně: `python3 udaje/udaje.py --ukazka projdi udaje/ukazka/smlouvy`

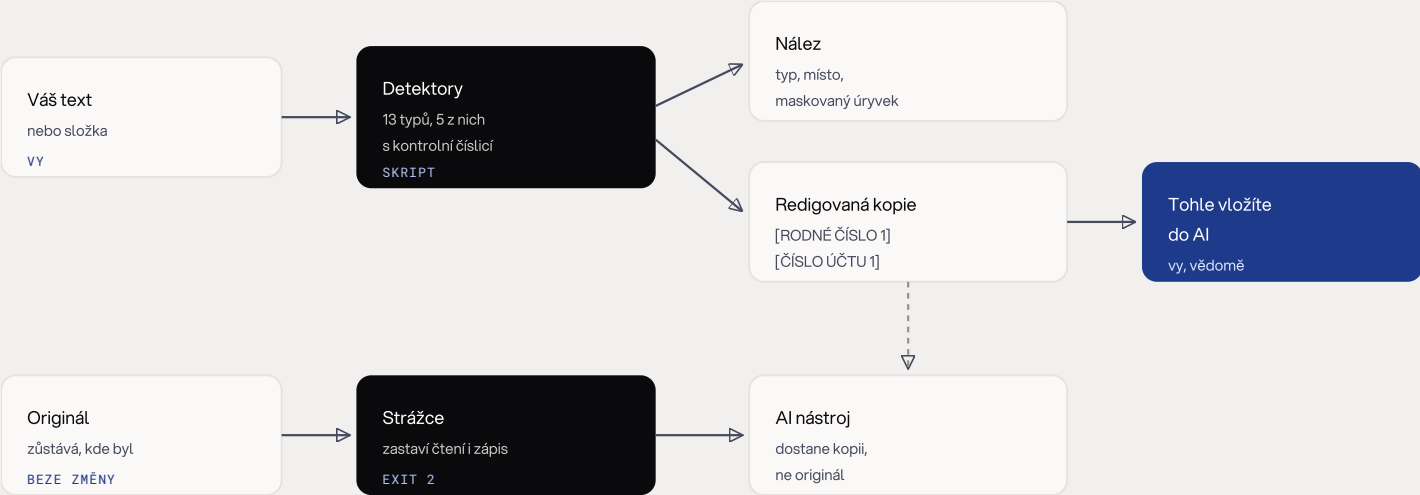
PROČ SE OVĚŘUJE KONTROLNÍ ČÍSLICE

Protože bez ní je takový nástroj k ničemu. Devítimístné a desetimístné číslo je v podkladech firmy na každé druhé řádce: číslo zakázky, kód výrobku, variabilní symbol. Kdyby to systém hlásil jako rodné číslo, přestanete nálezy číst do třetího dne. Rodné číslo se tu ověřuje dělitelností jedenácti i smysluplností data, IČO váhovým součtem, karta Luhnovým algoritmem, IBAN modulo 97 a české číslo účtu dvěma váhovými součty.

SCHÉMA

Z originálu se nikam nejde. Do AI jde kopie.

Celý smysl je v tom, že se mezi váš soubor a model dostane jeden krok navíc. Ten krok trvá vteřinu a dělá rozdíl mezi „vložíli jsme shrnutí smlouvy“ a „poslali jsme rodná čísla dvou lidí službě, se kterou nemáme smlouvu“.



CO SE NAJDE	NAHRADÍ SE V KOPII?	PROČ
Rodné číslo, číslo účtu, IBAN, karta, klíč, e-mail	ano, vždy	jsou to identifikátory; úkol bez nich funguje dál
Jméno, adresa, telefon	jen s přepínačem	poznávají se podle tvaru, takže rozhodnutí patří člověku
Zdravotní a mzdový údaj, označení důvěrnosti	ne, jen se hlásí	vyškrtnutím slova se text pro AI vhodným nestane; tenhle úkol se nedělá
IČO	ne	veřejný údaj z obchodního rejstříku

TŘI SKUPINY NÁLEZŮ A PROČ SE LIŠÍ

Nahrazují se: rodné číslo, číslo účtu, IBAN, karta, přístupové údaje, e-mail. Jsou to identifikátory, bez kterých úkol funguje dál.

Nahradí se s přepínačem: jména, adresy, telefony. Poznávají se jen podle tvaru, takže rozhodnutí necháváme na vás.

Jen se hlásí: zdravotní a mzdové údaje, označení důvěrnosti, IČO. U prvních tří proto, že vyškrtnutím slova „astma“ se text pro AI vhodným nestane; ty se nevkládají. IČO je naopak veřejné.

CO SE STANE S ORIGINÁLEM

Nic. Skript ho otevře ke čtení, spočítá otisk a zapíše si ho do `prohlednuto.tsv`, aby bylo vidět, co bylo prohlednuto a co se od té doby změnilo. Redigovaná kopie jde do vlastní složky běhu, takže se nikdy nestane, že by vám kopie přepsala originál.

Jeden příkaz, který použijete od zítřka.

Tohle je celá podstata balíčku a funguje to bez instalace, bez nastavení a bez AI nástroje. Vezmete text, který byste jinak vložili do chatu, proženete ho a vložíte to, co vypadne.

SKUTEČNÝ PRŮCHOD, VČETNĚ HLÁŠENÍ

\$ cat poznamka.txt | python3 udaje/udaje.py text

VSTUP:

Nájemnice Jana Veselá, r. č. 785123/1003, byt 304, platí na 19-2000145399/0800.

Kontakt jana.vesela@email.cz, 603 112 487. V srpnu byla na neschopence (astma).

VÝSTUP (to vložíte do AI):

Nájemnice Jana Veselá, r. č. [RODNÉ ČÍSLO 1], byt 304, platí na [ČÍSLO ÚČTU 1].

Kontakt [E-MAIL 1], 603 112 487. V srpnu byla na neschopence (astma).

HLÁŠENÍ:

Nahrazeno: Rodné číslo 1x, Číslo bankovního účtu 1x, E-mailová adresa 1x

Ponecháno (rozhodněte sami): Možné jméno osoby 1x, Telefonní číslo 1x.

Přepínač --včetně-nejistych je nahradí taky.

Nalezeno, ale nenahrazeno: Zdravotní údaj 1x. Tyhle údaje se vyškrtnutím slova nespraví; zdravotní a mzdové do AI nekládejte vůbec, IČO firmy je naopak veřejné.

CO CHCETE PO AI	CO Z TEXTU VYPUSTIT	CO NECHAT
Shrnout smlouvu	rodné číslo, číslo účtu, adresu	ustanovení, lhůty, částky, typ smlouvy
Odpovědět nespokojenému zákazníkovi	rodné číslo, číslo objednávky s údaji	popis problému; jméno jen když píšete odpověď jemu
Udělat zápis z porady	mzdy, zdravotní důvody absence, hodnocení lidí jmenovitě	úkoly, termíny, rozhodnutí
Projít životopisy	všechno; tenhle úkol je vysoce rizikové použití	nic, dokud nemáte ošetřený dohled a smlouvu o zpracování

VŠIMNĚTE SI TŘÍ VĚCÍ

1.

Text dál dává smysl: víte, že jde o jednoho nájemníka a jeden účet.

2.

Telefon a jméno zůstaly, protože se poznávají jen podle tvaru, a systém to říká nahlas.

3.

Slovo „astma“ zůstalo taky, ale s poznámkou, že tenhle text do běžné AI nepatří vůbec; to není věc redakce, ale rozhodnutí o úkolu.

JAK Z TOHO UDĚLAT ZVYK

Dejte si to na klávesovou zkratku, ať to není příkaz, ale reflex. Na macOS přes Shortcuts nebo Automator (akce Spustit shellový skript), na Windows přes PowerShell:

Get-Clipboard | python udaje\udaje.py text | Set-Clipboard .

Zkopírujete text, zmáčknete zkratku, vložíte. Tím je hotovo devadesát procent práce.

DETEKTORY

Třináct typů a u každého, jak jistě to ví.

Jistota není ozdoba. Rozhoduje o tom, jestli se nález nahradí sám, jestli se vás systém zeptá, a jestli kvůli němu strážce zastaví práci AI nástroje.

TYP	JAK SE POZNÁ	JISTOTA	CO SE S NÍM STANE
Rodné číslo	tvar + dělitelnost 11 + platné datum	ověřeno	nahradí se
Číslo platební karty	13 až 19 číslic + Luhn	ověřeno	nahradí se
IBAN	tvar + modulo 97	ověřeno	nahradí se
Číslo bankovního účtu	český tvar + dva váhové součty	ověřeno	nahradí se
Heslo, token, klíč	tvary sk-, ghp_, AKIA, Alza, PRIVATE KEY, heslo=	ověřeno	nahradí se
E-mailová adresa	tvar adresy	pravděpodobné	nahradí se
IČO	osm číslic + váhový součet	pravděpodobné	nenahrazuje se, je veřejné
Zdravotní údaj	slovník (i bez diakritiky)	pravděpodobné	jen se hlásí, GDPR čl. 9
Mzdový a exekuční údaj	slovník	pravděpodobné	jen se hlásí
Označení důvěrnosti	slovník (NDA, obchodní tajemství)	pravděpodobné	jen se hlásí
Telefonní číslo	tvar devíti číslic	nejisté	hlásí, nahradí s přepínačem
Adresa s PSČ	ulice, číslo, PSČ	nejisté	hlásí, nahradí s přepínačem
Možné jméno osoby	seznam běžných křestních jmen	nejisté	hlásí, nahradí s přepínačem

ZAČÁTEK VÝPISU DETEKTORŮ (SKUTEČNÝ VÝSTUP)

```
$ python3 udaje/udaje.py vzorky
```

Detektory (13) a to, co umí a neumí:

- Rodné číslo (rodne-cislo), jistota ověřeno, ověřuje kontrolní číslici, v kopii se nahrazuje
GDPR čl. 5 odst. 1 písm. c) (minimalizace); nakládání s rodným číslem navíc omezuje § 13c zák. č. 133/2000 Sb.
Rodné číslo z textu vypusťte. Žádný běžný úkol pro AI ho nepotřebuje; k rozlišení osob v dokumentu stačí značka.
- Číslo platební karty (platebni-karta), jistota ověřeno, ověřuje kontrolní číslici, v kopii se nahrazuje
GDPR čl. 32 (zabezpečení zpracování); pravidla kartových asociací ukládání čísla karty v čitelné podobě zakazují
Číslo karty nekládejte nikam, ani do AI. Pokud už vloženo bylo, nechte kartu zablokovat.
- IBAN (iban), jistota ověřeno, ověřuje kontrolní číslici, v kopii se nahrazuje
GDPR čl. 5 odst. 1 písm. c)

CO DETEKTORY NEPOZNAJÍ, A JE TO V NÁLEZU NAPSANÉ

Nemoc opsanou bez obvyklého slova („od jara nevyjde ani do schodů“). Jméno bez typického českého tvaru a cizince. Text v obrázku, PDF, archivu nebo v tabulce v binárním formátu: ty se neprohližejí a každý takový soubor je v nálezu v oddílu Nekomontrolováno. Údaj rozdělený mezi dvě buňky. Příkaz `vzorky` tenhle seznam vypíše celý. Proto nález nikdy neříká, že je text čistý, a proto by to po něm neměl říkat ani váš asistent.

NÁLEZ

Typ, místo, maskovaný úryvek. Hodnota nikdy.

Nález je soubor, který můžete beze strachu poslat kolegovi, otevřít na schůzce nebo nechat v repozitáři. Je v něm všechno, co potřebujete k nápravě, a nic, co by někomu pomohlo.

SKUTEČNÝ VÝSTUP: UDAJE/NALEZY/N1.MD (PŘEHLED A DVA SOUBORY)

Přehled

```
| Typ | Počet | Jistota | Předpis |
|---|---|---|---|
| Číslo bankovního účtu | 2 | ověřeno | GDPR čl. 5 odst. 1 písm. c) |
| IBAN | 1 | ověřeno | GDPR čl. 5 odst. 1 písm. c) |
| Číslo platební karty | 1 | ověřeno | GDPR čl. 32 (zabezpečení zpracování); pravidla kartových asociací ukládání čísla karty v čitelné podobě zakazují |
| Heslo, token nebo klíč | 2 | ověřeno | GDPR čl. 32 (zabezpečení zpracování); vaše vnitřní pravidlo |
| Rodné číslo | 5 | ověřeno | GDPR čl. 5 odst. 1 písm. c) (minimalizace); nakládání s rodným číslem navíc omezuje § 13c zák. č. 133/2000 Sb. |
| E-mailová adresa | 9 | pravděpodobné | GDPR čl. 5 odst. 1 písm. c) |
| IČO | 5 | pravděpodobné | veřejný údaj z obchodního rejstříku; osobním údajem je jen u podnikající fyzické osoby |
| Mzdový nebo exekuční údaj | 5 | pravděpodobné | GDPR čl. 5 odst. 1 písm. c) a čl. 32 |
| Označení důvěrnosti | 10 | pravděpodobné | nejde o osobní údaj; § 504 zák. č. 89/2012 Sb. (obchodní tajemství) a vaše smlouvy o mlčenlivosti |
| Zdravotní údaj | 7 | pravděpodobné | GDPR čl. 9 odst. 1 (zvláštní kategorie); pokuta až 20 mil. EUR nebo až 4 % obrátu (čl. 83 odst. 5) |
| Adresa s PSČ | 2 | nejisté | GDPR čl. 5 odst. 1 písm. c) |
| Možné jméno osoby | 7 | nejisté | GDPR čl. 5 odst. 1 písm. c) |
| Telefonní číslo | 7 | nejisté | GDPR čl. 5 odst. 1 písm. c) |
```

README-UKAZKA.md

```
- **Číslo bankovního účtu** (ověřeno), řádek 15, sloupec 14: `19-xxxxxxxx/0800`
- **Označení důvěrnosti** (pravděpodobné), řádek 23, sloupec 40: `xxxxxxxx txxxxxxxx`
- **Označení důvěrnosti** (pravděpodobné), řádek 23, sloupec 60: `dxxxxxx`
- **Označení důvěrnosti** (pravděpodobné), řádek 23, sloupec 69: `mxxxxxxxx`
- **Označení důvěrnosti** (pravděpodobné), řádek 23, sloupec 82: `Nxx`
- **Mzdový nebo exekuční údaj** (pravděpodobné), řádek 24, sloupec 63: `exxxxxx`
- **Zdravotní údaj** (pravděpodobné), řádek 25, sloupec 33: `axxxx`
- **Zdravotní údaj** (pravděpodobné), řádek 25, sloupec 40: `axxxxxx`
- **Zdravotní údaj** (pravděpodobné), řádek 25, sloupec 49: `lxxxxxx zxxxxx`
- **Zdravotní údaj** (pravděpodobné), řádek 25, sloupec 66: `nxxxxxxxx`
```

crm/export_najemnici.csv

```
- **E-mailová adresa** (pravděpodobné), řádek 2, sloupec 35: `jxxxxxxxxx@email.cz`
- **IČO** (pravděpodobné), řádek 3, sloupec 37: `99xxxx16`
- **E-mailová adresa** (pravděpodobné), řádek 3, sloupec 46: `uxxxxx@kavarnaumostu.cz`
- **IČO** (pravděpodobné), řádek 4, sloupec 35: `99xxxx94`
- **E-mailová adresa** (pravděpodobné), řádek 4, sloupec 44: `kxxxxxxxx@navrhy-interiery.cz`
- **IČO** (pravděpodobné), řádek 5, sloupec 47: `99xxxx08`
- **E-mailová adresa** (pravděpodobné), řádek 5, sloupec 56: `fxxxxxxxx@termoservis-hl.cz`
```

CO S NÁLEZEM UDĚLAT

Projít shora dolů a u každého řádku se zeptat, jestli ten údaj v souboru musí být. Překvapivě často nemusí: exporty z CRM nosí sloupce, které nikdo nepoužívá, a poznámky k platbám obsahují čísla karet, která tam zapsal někdo před třemi lety. Co nemusí být v souboru, nemusí se ani maskovat.

JAK ČÍST MASKOVANÝ ÚRYVEK

U čísel zůstávají krajní znaky, abyste nález našli v originálu, ale prostředek chybí. U e-mailu zůstává doména, protože ta je obvykle to podstatné (je to zákazník, nebo kolega?). U klíče zůstává jen prefix, podle kterého se pozná služba. Celou hodnotu si otevřete v originálu, na vlastní odpovědnost a bez toho, aby prošla přes jakýkoli report.

REDIGOVANÁ KOPIE

Smlouva, která dál dává smysl, ale nemá v sobě lidi.

Tohle je ten rozdíl proti radě „neposílejte do AI osobní údaje“. Tu radu nikdo nedodrží, protože úkol zůstává. Redigovaná kopie úkol zachová: značky jsou číslovány v rámci souboru, takže se dvě osoby nepopletou.

SKUTEČNÝ VÝSTUP: UDAJE/REDIGOVANO/B1/SMLOUVY/NAJEMNI_SMLOUVA.MD (ZAČÁTEK)

Smlouva o nájmu bytu č. 304

****Pronajímatel:**** VLTAVIA INVEST s.r.o., IČO 99990008, se sídlem Jankovcova 1522/53, 170 00 Praha 7, zapsaná v obchodním rejstříku vedeném Městským soudem v Praze.

****Nájemnice:**** Jana Veselá, r. č. [RODNÉ ČÍSLO 1], trvale bytem Tusarova 1246/17, 170 00 Praha 7.

I. Předmět nájmu

Pronajímatel přenechává nájemnici do užívání byt č. 304 o velikosti 2+kk ve 3. nadzemním podlaží domu Jankovcova 1522/53, Praha 7.

II. Nájemné a platby

Nájemné činí 18 400 Kč měsíčně, záloha na služby 3 200 Kč měsíčně. Nájemnice platí nájemné na účet pronajímatele [ČÍSLO ÚČTU 1], variabilní symbol 304, vždy do 5. dne kalendářního měsíce.

Jistota ve výši trojnásobku nájemného byla složena dne 1. 3. 2026 na tentýž účet.

III. Doba nájmu

Nájem se sjednává na dobu určitou od 1. 3. 2026 do 28. 2. 2027. Nedojde-li k písemné výpovědi alespoň tři měsíce před uplynutím sjednané doby, nájem se prodlužuje o dalších dvanáct měsíců.

IV. Kontaktní údaje

Nájemnice: [E-MAIL 1], telefon 603 112 487.

CO V KOPII ZŮSTALO SCHVÁLĚ

IČO pronajímatele, protože je veřejné a bez něj by shrnutí ztratilo smysl. Jméno a adresa, protože jsou v kategorii, kterou systém sám nenahrazuje. A celý zbytek textu: ustanovení, lhůty, částky. Model na nich pracuje stejně dobře jako na originálu.

ČÍSLA ZNAČEK DRŽÍ V RÁMCI SOUBORU

[RODNÉ ČÍSLO 1] je ve smlouvě na dvou místech a je to pokaždé tentýž člověk; model tedy ví, že jde o jednu osobu. V jiném souboru začíná číslování znovu, protože kopie má dávat smysl sama o sobě a protože průběžné číslování napříč soubory by samo o sobě bylo stopou.

KDYŽ CHCETE PRYČ I JMÉNA

Přepínač --vcetne-nejistych nahradí jména, adresy a telefony: z nájemnice se stane [OSOBA 1] a text je použitelný i ve veřejném nástroji. Je to vědomé rozhodnutí, proto to systém nedělá sám.

STRÁŽCE

Jediná část, která opravdu zabrání vložení.

Všechno ostatní v tomhle balíčku stojí na tom, že si vzpomenete. Strážce ne. Je to hook, který se ptá před každým čtením a zápisem souboru, a když v něm najde ověřený osobní nebo přístupový údaj, práci zastaví. AI nástroj dostane návratový kód 2 a tuhle zprávu:

SKUTEČNÁ ZPRÁVA STRÁŽCE (HOOK STRAZCE, EXIT 2)

Co nesmí do AI: soubor smlouvy/najemni_smlouva.md obsahuje Rodné číslo (2x), Číslo bankovního účtu (1x). S originálem nepracujte. Spustíte `python3 udaje/udaje.py projdi smlouvy/najemni\_smlouva.md`, nález ukažete uživateli a dál pracujte s redigovanou kopií ze složky udaje/redigovano/. Hodnoty z toho souboru nikam nevypisujete.

UDÁLOST V AI NÁSTROJI	CO STRÁŽCE UDĚLÁ
Nástroj chce přečíst soubor s rodným číslem	zastaví, pošle k redigované kopii
Nástroj chce přepsat soubor s číslem účtu	zastaví
Nástroj pracuje s čistým souborem	pustí dál, nic nehlásí
Nástroj pracuje s redigovanou kopií	pustí dál (kopie identifikátory nemá)
Soubor je PDF nebo obrázek	pustí dál, ale v nálezů je uvedený jako neprohlédnutý

CO HOOK HLÁSÍ PŘI STARTU SEZENÍ (SKUTEČNÝ VÝSTUP)

```
$ python3 udaje/udaje.py hook start
```

```
## Co nesmí do AI
Běh 1: prohlédnuto 11 souborů, nálezů 63.
Číslo bankovního účtu: 2 (ověřeno)
IBAN: 1 (ověřeno)
Číslo platební karty: 1 (ověřeno)
Heslo, token nebo klíč: 2 (ověřeno)
Rodné číslo: 5 (ověřeno)
E-mailová adresa: 9 (pravděpodobné)
IČO: 5 (pravděpodobné)
Mzdový nebo exekuční údaj: 5 (pravděpodobné)
Označení důvěrnosti: 10 (pravděpodobné)
Zdravotní údaj: 7 (pravděpodobné)
Adresa s PSČ: 2 (nejisté)
Možné jméno osoby: 7 (nejisté)
```

ČEMU TO ZABRÁNÍ A ČEMU NE

Zabrání tomu, aby asistent otevřel soubor, o kterém nevíte, že v něm rodná čísla jsou. To je nejčastější cesta, jak se osobní údaje dostanou do modelu: nikdo to neudělal schválně, jen někdo požádal o shrnutí složky.

Nezabrání tomu, abyste text zkopírovali do prohlížeče rukou. Na to je příkaz na straně 4 a klávesová zkratka. A nezabrání tomu, aby šel ven text z PDF, který neumí přečíst; to ale říká nahlas, místo aby mlčel.

HRANICE

Tři věci, které systém **neudělá ani omylem.**

- Nalezenou hodnotu nevypíše celou.** Ani do nálezu, ani do kopie, ani do deníku, ani na obrazovku. Mapa původních hodnot se nikam neukládá; existuje jen v paměti po dobu běhu.
- Originál nemaže ani nepřepisuje.** Redigovaná kopie je nový soubor ve složce běhu. O každém prohlédnutém originálu se ukládá jen otisk, aby bylo vidět, co se od posledního projetí změnilo.
- Nerozhoduje za vás a nic neodesílá.** Nejisté nálezy hlásí, ale nenahrazuje, dokud to nepovolíte. Skript nepoužívá síť, takže se text nemá jak dostat ven.

OVĚŘENÍ HRANIC, SKUTEČNÝ VÝSTUP

```
$ python3 udaje/udaje.py kontrola
```

```
Hranice 1 (chráněnou hodnotu nevypíše celou): OK - ověřeno proti 26
nalezeným hodnotám (14 identifikátorů) ve 0 výstupech
Hranice 2 (originál nemaže ani nepřepisuje): OK - dokumentů 0, kontrolní
součty sedí; originálů jen čtených 0
Hranice 3 (nerozhoduje za vás, nic neodesílá): OK - žádná síť, nejisté
nálezy jen hlášené
Detektory: 13, z toho s kontrolní číslicí 5
Nastavení: hlídaná složka .
Běhy: 0 (příkaz projdi)
Výsledek: vše v pořádku
```

SAMOTEST: POSLEDNÍCH ŠEST Z 37 KONTROL (SKUTEČNÝ VÝSTUP)

```
$ python3 udaje/udaje.py test
```

```
OK strážce v hlášení neuvede žádnou hodnotu a pošle ke kopii
OK čistý soubor strážce propustí
OK pravidlo do firemních pravidel nese jen typy, žádné identifikátory
OK kontrola hranic čistá
OK změna originálu uživatele není chyba, jen výzva k novému projetí
OK ruční změna nálezu je chyba (dokument se nepřepisuje)
```

JAK SE HRANICE 1 OVĚŘUJE

Ne slovem. Příkaz `kontrola` vezme skutečné hodnoty z hlídané složky (rodná čísla, účty, karty, klíče, e-maily, telefony, adresy, jména) a hledá je ve všech souborech, které skript vyrobil. Když by jediná z nich kdekoli byla, vypíše CHYBU. Totéž dělá samotest, a je to jedna z jeho 37 kontrol.

CO DĚLÁ KONTROLA JEŠTĚ

Hlídá, že dokumenty, které skript zapsal, se od zápisu nezměnily, a odděleně hlásí, kolik vašich originálů se od posledního projetí změnilo. To druhé není chyba, ale pokyn: projedte je znovu.

Tři úrovně. První stojí deset minut.

ÚROVEŇ 1 · PŘÍKAZ A ZKRATKA

Rozbalit balíček a používat `text` před vložením. Nepotřebujete AI nástroj, nastavení ani složku; stačí Python 3. Když to dáte na klávesovou zkratku, stane se z toho zvyk do týdne. Většina užítku je tady.

ÚROVEŇ 2 · HLÍDANÁ SLOŽKA A STRÁŽCE

Nastavíte, která složka se hlídá, zapojíte hook do Claude Code nebo Codexu a projedete ji. Od té chvíle platí, že asistent nesmí otevřít soubor s ověřeným osobním údajem. Příkaz `hlídat` pak sleduje změny; nebo to dělá plánovač (na macOS launchd, na Windows schtasks).

Nálezy se nepřepisují, každý běh má své číslo. Po měsíci je na nich vidět, jestli nálezů ubývá, což je jediné měřítko, které u téhle věci dává smysl.

ÚROVEŇ 3 · KONTROLA NA CESTĚ VEN

Zásada: `udaje.py` zůstává bez sítě. Co potřebuje síť, jsou samostatné skripty mimo tuhle složku, které si text nechají zkontrolovat a teprve pak ho pošlou dál.

Druhá věc úrovně 3 je propojení s pravidly: `pravidlo` vypíše oddíl „Co nekládáme do AI“ z toho, co se u vás skutečně našlo, a ten patří do firemních pravidel. Pravidlo, které vzniklo z vašich vlastních nálezů, lidé berou jinak než obecný zákaz.

CO NASAZENÍ NEVYŘEŠÍ

Lidi, kteří si nemají kam jít pro odpověď. Když zaměstnanec neví, jestli smí něco vložit, a nemá se koho zeptat, vloží to. Proto má smysl mít vedle téhle kontroly jedno jméno a jednu adresu, kam se dá napsat, a odpovídat tam do druhého dne.

CO CHCETE	ÚROVEŇ	CO NA TO POTŘEBUJETE
Přestat vkládat rodná čísla do chatu	1	jeden příkaz, deset minut
Aby to lidé opravdu dělali	1	klávesovou zkratku, ne školení
Aby to asistent nemohl obejít	2	hook strážce v AI nástroji
Vědět, co v podkladech vůbec máme	2	jedno projetí složky
Mít pravidlo, které vzniklo z našich dat	3	příkaz pravidlo

Co tenhle systém neumí a nikdy neuměl.

Nástroj na ochranu dat, u kterého nevíte, kde končí, je nebezpečnější než žádný: dává jistotu, kterou nemá čím krýt. Proto je tahle stránka tady a proto totéž říká i každý nález.

- Nález není potvrzení, že je text čistý.** Je to seznam toho, co detektory našly. Co nenašly, v něm není, a to včetně věcí, které tam jsou.
- Z obrázků, PDF a archivů nechte.** Sken smlouvy projde bez jediného nálezu, protože v něm není text. Skript to u každého takového souboru napíše do oddílu Nekontrolováno; přečtete si ho dřív než přehled.
- Zdravotní a mzdové údaje poznává podle slov.** Opis bez obvyklého slova mu unikne. Tahle kategorie je upozornění, ne síto.
- Jména poznává podle seznamu běžných českých křestních jmen.** Neobvyklé jméno ani cizince nenajde, a naopak občas označí název firmy. Proto je v kategorii, kterou nenahrazuje sám.
- Neřeší, jestli nástroj smí osobní údaje zpracovávat.** To rozhoduje smlouva o zpracování podle GDPR čl. 28 odst. 3, ne tenhle výpis. Soupis nástrojů a jejich zařazení řeší balíček Pravidla pro práci s AI.

CO DETEKTORY NEUMÍ, PŘÍMO Z NÁSTROJE

```
$ python3 udaje/udaje.py vzorky
```

- jméno bez typického českého zakončení; naopak hlásí i názvy firem
- text v obrázku, PDF, archivu nebo v tabulce v binárním formátu
- údaj rozdělený mezi dva soubory nebo dvě buňky

Proto je výstup nález k projití, ne potvrzení, že je text čistý.

CO LIDÉ ČEKAJÍ, ŽE TO NAJDE	NAJDE?	PROČ
Rodné číslo v tabulce	ano	ověřené kontrolní číslicí
Rodné číslo na skenu smlouvy	ne	sken není text; je v Nekontrolováno
Jméno Petr Novák	ano	Petr je v seznamu běžných křestních jmen
Jméno Aiman Haddád	ne	seznam zná běžná česká jména
Diagnózu napsanou jako „bolesti zad od jara“	ne	slovník to nezná
Číslo zakázky 20260042	ne, a je to správně	neprojde kontrolou IČO

PROČ JE SEZNAM OMEZENÍ I V KAŽDÉM NÁLEZU

Protože dokument si přečte jeden člověk jednou, ale nález otevře ten, kdo s podklady pracuje, pokaždé. Kdyby omezení byla jen tady, za měsíc by někdo řekl „systém nic nenašel, je to v pořádku“ nad složkou plnou skenů.

JEDNA VĚC, KTEROU NEDĚLEJTE

Neptejte se chatbota, jestli jsou v textu osobní údaje. Tím, že jste text vložili k té otázce, už opustil váš počítač. Je to nejčastější chyba, kterou u téhle věci lidé dělají, a je v ní zvláštní ironie.

SHRNUŤÍ

Pět vět, které si odnese.

- 1 Rada „nekládejte osobní údaje do AI“ nefunguje, protože úkol zůstává. Funguje redigovaná kopie, se kterou se úkol udělat dá.
- 2 Rodné číslo, IČO, IBAN, číslo účtu a karta se dají ověřit kontrolní číslicí. Nástroj, který to nedělá, zahltí falešnými nálezy a lidé ho přestanou číst.
- 3 Zdravotní údaj se neredaguje, ten se nekládá. Je to zvláštní kategorie podle GDPR čl. 9 a vyškrtnutím jednoho slova se text vhodným nestane.
- 4 Osobní údaje patří jen do nástroje, se kterým má firma smlouvu o zpracování (GDPR čl. 28 odst. 3), a jen v rozsahu, který úkol potřebuje (čl. 5 odst. 1 písm. c)).
- 5 Když se údaj do nástroje dostane, je důležitější to vědět než to zamlčet. Přístupové údaje se mění hned, hlášení se nikomu nevyčítá.

JAK POZNÁTE, ŽE TO FUNGUJE

Podle toho, že nálezů ubývá, ne podle toho, že je nikdo nehlásí.

Každý běh má své číslo a zůstává. Po měsíci se podíváte, jestli jich je méně. Firma, ve které nikdo nikdy nic nenahlásil, nemá čistý provoz; má jen lidi, kteří vědí, že se hlášení nevyplácí.

CO JE V BALÍČKU

`udaje.py` celý mechanismus · `INSTALACE.md` postup pro váš AI nástroj · `NAVOD-CHATGPT.md` ruční verze a instrukce do projektu · `hooks/` čtyři konfigurace · `ukazka/` modelová firma s deseti soubory

Vyzkoušení bez vlastních dat: `python3 udaje/udaje.py ukazka` . Hranice detektorů: `vzorky` . Samotest (37 kontrol): `python3 udaje/udaje.py test` .

CO UDĚLAT ZÍTRA RÁNO, I KDYBYSTE BALÍČEK NEPOUŽILI

1. Zjistěte, na jakých účtech lidé AI používají, a jestli má firma s dodavatelem smlouvu o zpracování. **2.** Vypněte v nastavení trénování na vašich datech; u osobních tarifů bývá zapnuté. **3.** Řekněte lidem, komu hlásit, že se do nástroje dostal údaj, který tam neměl být, a že se to nevyčítá. Tyhle tři věci nestojí nic a pokryjí většinu rizika.

Právní doložka. Tento dokument ani příložený balíček nejsou právní službou ani právním posouzením a nenahrazují konzultaci s advokátem ani posouzení vlivu na ochranu osobních údajů. Nález není potvrzením souladu s GDPR ani tvrzením, že text neobsahuje osobní údaje; vypisuje jen to, co detektory našly, a uvádí, co nebylo prohlédnuto. Právní stav je zachycen k 2. 10. 2026 (GDPR, zákon č. 110/2019 Sb., zákon č. 133/2000 Sb.). Modelová firma VLTAVIA INVEST s.r.o., všechny osoby, čísla účtů, rodná čísla a klíče v ukázkách jsou smyšlené; čísla s kontrolní číslicí jsou sestavená tak, aby prošla kontrolou, a nepatří žádné skutečné osobě ani účtu. © 2026 Český institut pro AI a data, z. ú.