



Budapest ML Forum, 2023



# AI lehetőségek a szerződéskezelésben

2023 Június 8

Rábay Kristóf

Senior Data Scientist, Hiflylabs

# Hiflylabs

Creating business value **in a world full of data**. We engineer data solutions and build apps that make an impact.

## Areas of competence:



Data Platform



Advanced Analytics



Analytics Engineering



Data Visualization



Web Application



Management & Digital Consulting



Mobile



IT & Support

Our primary focus is **Data**

**20+**  
years' experience  
in data industry

**150+**  
staff of data  
experts

**500+**  
data projects  
delivered



# Agenda

1

Szerződéskezelés

**Mik a szerződéskezelés elemei?** Milyen kihívásokat rejt?

2

MI hozzáadott értéke

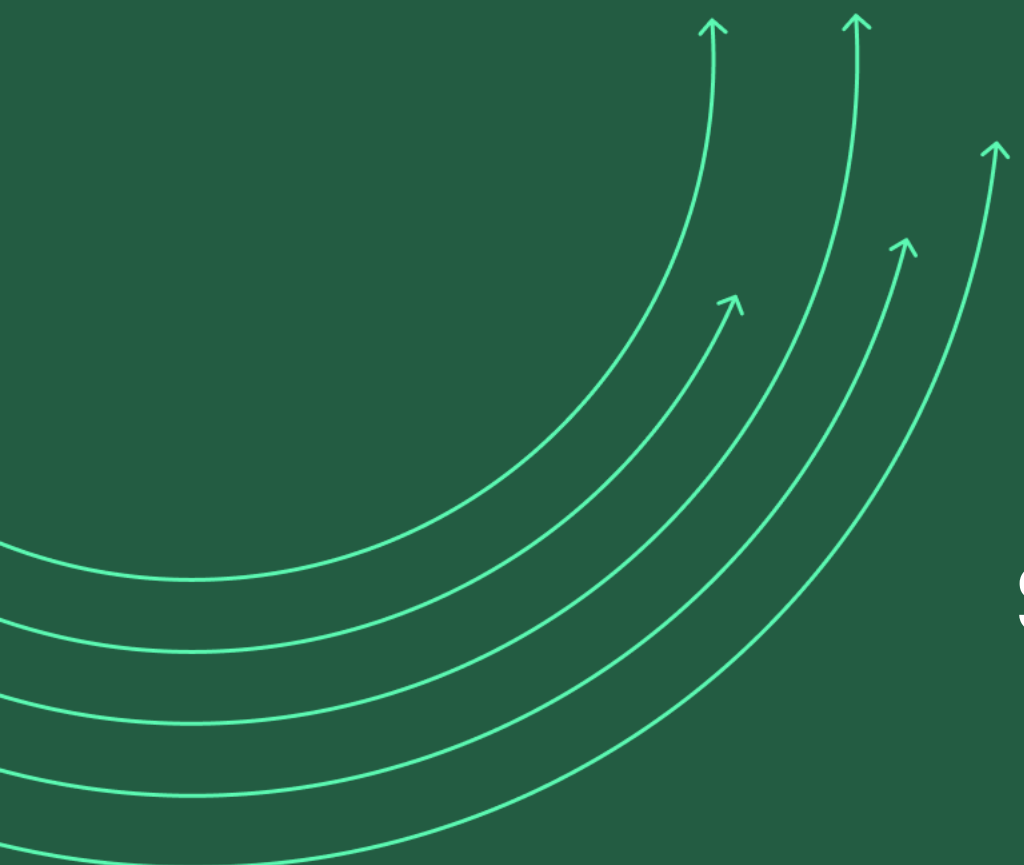
**Hogyan hasznosítható e területen az MI?**

Mi az üzleti hozzáadott értéke, milyen technológiát alkalmaz?

3

Felhasználási területek

**Információkinyerés, osztályozás és Generative AI** példák és demó



# Szerződéskezelés

# Miről szól a szerződéskezelés?

## Adminisztráció



Meggyőződés a jogi és szabályozási követelményeknek való megfelelésről



Szerződési, felhasználási feltételek megértése

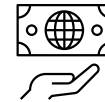


Szerződő felek kötelezettségeinek nyomon követése, teljesítések monitorozása



Módosítások, változások kezelése

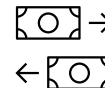
## Kockázat, pénzügyi vonatkozás



Árazás, feltételek, ösztönzők feltérképezése



Szerződés beteljesülését kockáztató faktorok, kikötések beazonosítása, átvilágítás



Felmerülő költségek, felelősségi korlátok, kártérítési kikötések beazonosítása



Sablon-szabványosítás, e-aláírás, szerződésaudit

# Miről szól a szerződéskezelés?

## Adminisztráció



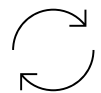
Meggyőződés a jogi és szabályozási követelményeknek való megfelelésről



Szerződési, felhasználási feltételek megértése



Szerződő felek kötelezettségeinek nyomon követése, teljesítések monitorozása



Módosítások, változások kezelése

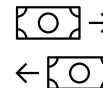
## Kockázat, pénzügyi vonatkozás



Árazás, feltételek, ösztönzők feltérképezése



Szerződés beteljesülését kockáztató faktorok, kikötések beazonosítása, átvilágítás



Felmerülő költségek, felelősségi korlátok, kártérítési kikötések beazonosítása



Sablon-szabványosítás, e-aláírás, szerződésaudit

Több (100) oldalnyi (jogi) szöveg precíz, hatékony feldolgozása elengedhetetlen

# Milyen kihívásokkal jár?



Manuális,  
időigényes  
feladatok

- Szerződések teljes szövegének, melléketek, kiegészítések, csatolmányok **áttekintése**
- Szabályozásnak való **megfelelés**, felhasználási feltételek **feltérképezése**,
- Az **információ-kinyerés** folyamata



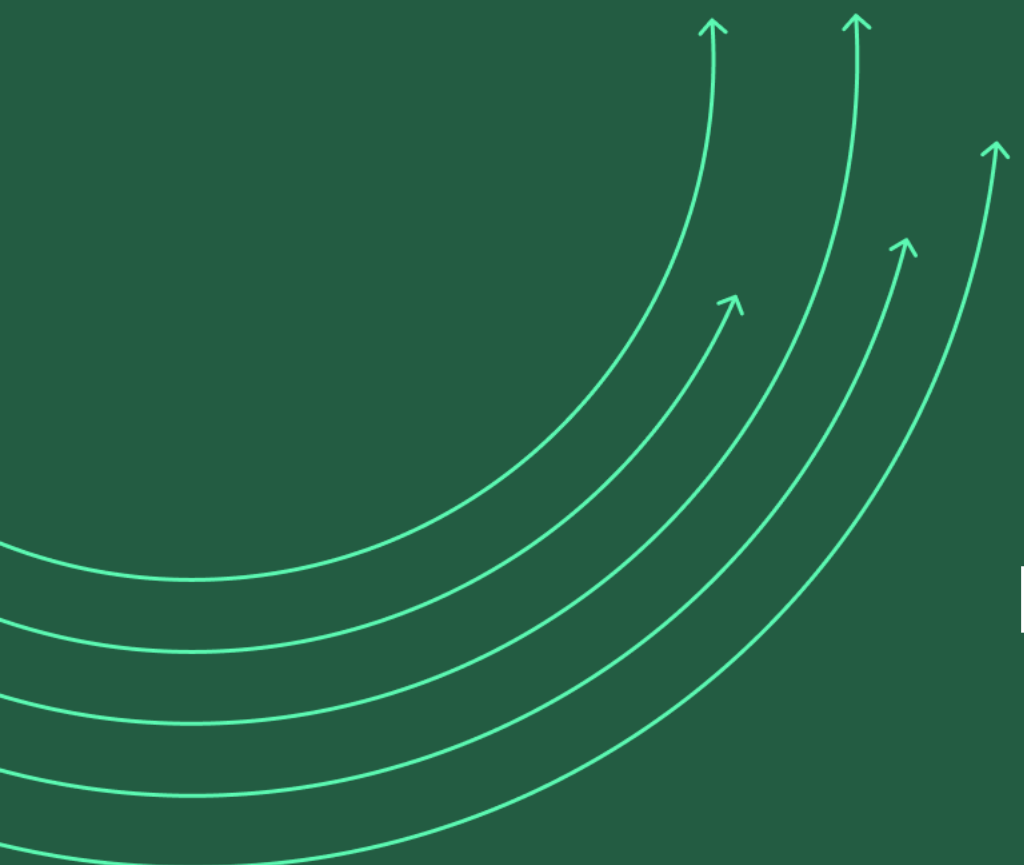
Átláthatóság,  
ellenőrizhetőség

- Több (100, 1000) szerződést átlátni, csoportosítani, **struktúráltnan kezelni és tárolni**
- Kockázatok, jogi környezetek, feltételek **monitorozása**, felügyelete



Hibák, mulasztás




- Szerződésbeni **hibás elemek** (elgépelések, rossz dátum, stb.) beazonosítása
- Manuálisan kinyert információk lehetnek **pontatlanok**, hibásak
- **Kihagyott elemek** (pl.: automatikus megújulás, fontos kondíció) okozta meglepetések



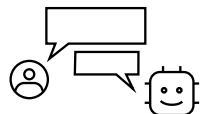
# Mesterséges Intelligencia



# Milyen üzleti haszon realizálható vele?

	Kihívás	MI hatás	Megoldás
	Kereskedelmi- és üzleti hitelszerződések átvizsgálása	Évi 360 000 órányi ügyvédi munka <b>másodpercek</b> alatt	Szerződés-struktúra és tartalom monitorozása felügyelet nélküli gépi tanuló algoritmussal
	Beszerezési folyamatokkal járó papírmunka mennyisége	<b>80% gyorsabb</b> draftolás, 50% gyorsabb formakitöltés	A DocuSign segítségével sablonok automatikus generálása és kitöltése
	Pereskedéssel, peres eljárásokkal járó költségek csökkentése	Átlagosan 6-10 órányi munka kevesebb, mint <b>2 perc alatt</b>	Kereseti reakció, ellenkérelem, viszontkereset automatikus generálása

# Milyen megoldásokat nyújt?



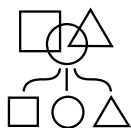
## Generative AI

- **Sablonírás**, szövegrészek átírása, újrafogalmazás, hozzáadás, kitöltés
- Szerződésekkel való **párbeszéd, tartalomösszegzés, kérdések** feltevése



## Információ- kinyerés

- Automatikusan **kinyerhető** (jogi) kifejezések, pl.: **szerződő felek, dátumok, pénzüsszegek, jogi referenciák, felelősségek, kikötések**
- Potenciális **kockázatos** faktorok, diszkrepanciák **megjelölése**



## Szöveg- szegmentáció

- Összefüggő szövegrészek, kitételek beazonosítása,
- Fájlok **tagolása** (mellékletek, kiegészítő csatolmányok, záradékok)
- Szöveg **tördelése** további feldolgozás céljából



## Szöveg- osztályozás

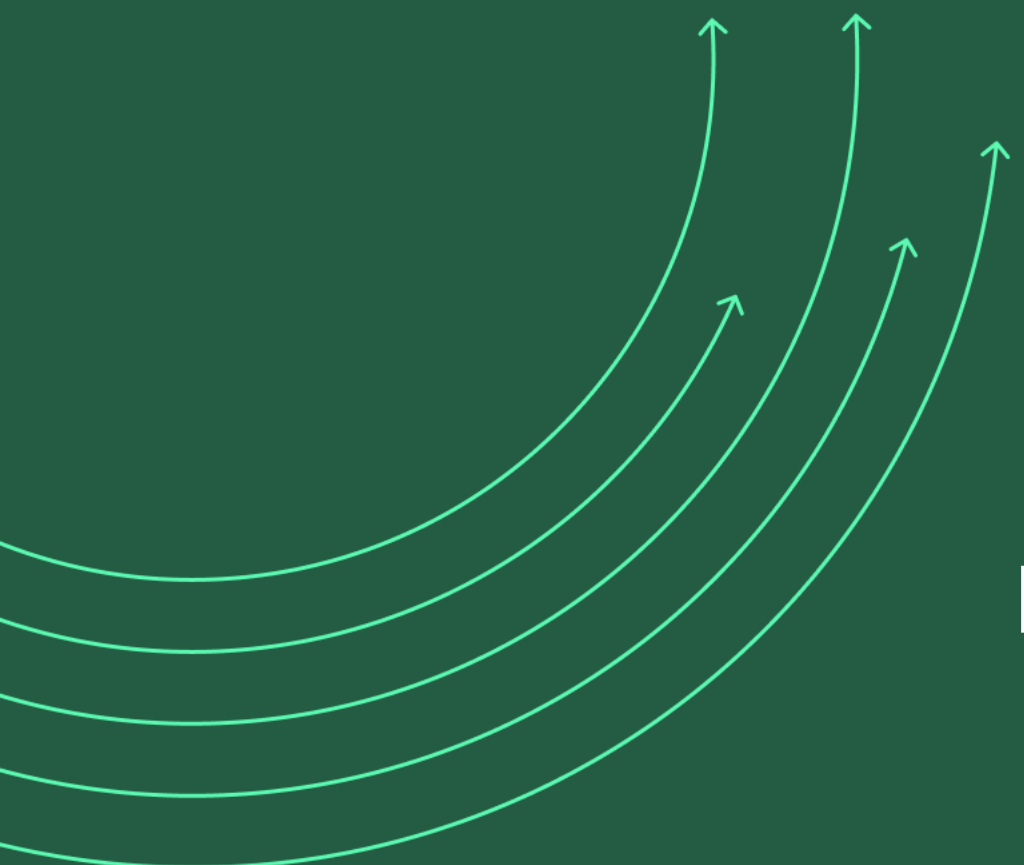
- **Szerződéstípusok** meghatározása (adásvételi, szállítási, bérleti, hitel, ...)
- **Nyelv**-szerinti (többnyelvűség), prioritási osztályozás



## Szemantikus keresés

- Fejlett keresőmotorokkal, kulcsszavak mellett körbeírással, tartalmi szűréssel **lokalizálható információ** szerződéseken belül
- Szerződések közötti **szűrés** strukturális és szemantikai alapon





# Példák és demó

# Use case #1: Információ-kinyerés névelem felismeréssel

## Korpusz jellemzői

## Névelemek / kifejezések



Szerződések, megállapodások, felhasználási feltételek, bírósági iratok, szabályzatok

- Szerződő felek, jogi hivatkozások,
- Alkalmazandó jog, felelősség korlátozása, kötbér



Pénzügyi riportok, jelentések, elemzések, biztosítási kötvények, követelések

- Pénzügyi, gazdasági mutatók, tickerek,
- Követelési összegek, kötvényszámok, baleseti bejelentők elemei



Orvosi nyilvántartások, feljegyzések, értékelések, diagnózisok, receptek, beteg űrlapok

- Betegségek, tünetek
- Gyógyszernevek, adagolás,
- Orvosi beavatkozások, készülékek



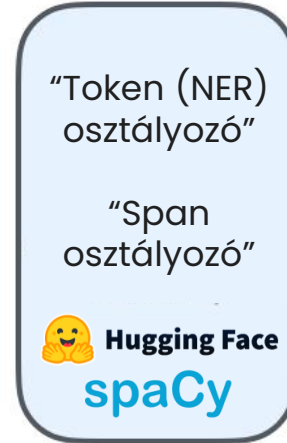
Termékleírások, számlák, beszerzési-logisztika, visszaküldési-visszatérítési szabályzatok

- Terméknevek és attribútumok, árinformáció,
- Vevő-eladó, szállítás, fizetés módja

# Use case #1: Információ-kinyerés névelem felismeréssel

```
"value": {  
  "start": 434,  
  "end": 444,  
  "text": "2023.08.01",  
  "labels": [  
    "Dátum"  
  ]  
}
```

```
"value": {  
  "start": 540,  
  "end": 568,  
  "text": "per-, teher-, és igénymentes",  
  "labels": [  
    "Jogi helyzet"  
  ]  
}
```



Személy<sup>[1]</sup> Pénz<sup>[2]</sup> Dátum<sup>[3]</sup> Jog<sup>[4]</sup> Jogi helyzet<sup>[5]</sup> Tárgy<sup>[6]</sup> Egyéb<sup>[7]</sup>

Az adásvételi szerződés létrejött a mai napon (2023.06.08 Dátum) alulírott felek között, jármű Tárgy adás-vétele tárgyában. Felek kijelentik, hogy az adásvétel visszerthes / ingyenes Jogi helyzet. (A megfelelő rész aláhúzendő.) A jármű Tárgy vételára: 4.000.000 Ft Pénz, azaz 4 millió forint Pénz. A járműnek Tárgy a vevő birtokába kerülési időpontja: 2023.08.01 Dátum. Felek kijelentik, hogy a forgalmi engedély Tárgy, és törzskönyv Tárgy átadása-átvétele megtörtént, az átadás-átvétel időpontja: 2023.08.01 Dátum. A tulajdonjog Jog változás hatálybalépésének időpontja: 2023.08.01 Dátum. Eladó Személy kijelenti, hogy a jármű Tárgy per-, teher-, és igénymentes Jogi helyzet.



## Modell építés

1 nagy (multiNER) modell, vagy névelemenként 1 modell

Welcome to the **Bank of China** .

ORG ———  
GPE ———

## SpanCat

Stuktúrálatlan kifejezések, egymást átfedő elemek, beágyazások






## Precision / Recall

'Kihagyás' rosszabb, mint helytelen felismerés (?)

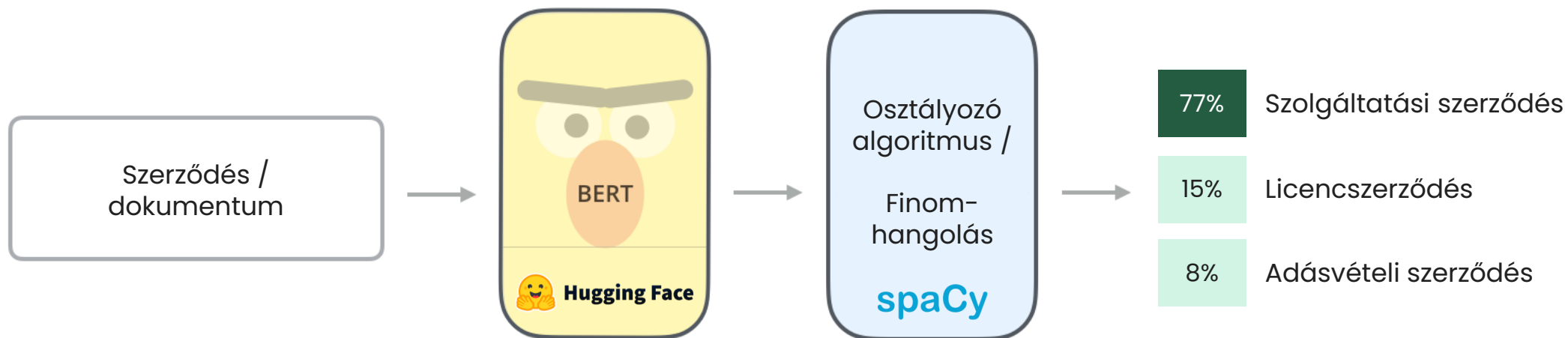
# Use case #2: Szövegosztályozás

## Korpusz jellemzői

## Osztályok, kategóriák

	Szerződések, megállapodások, felhasználási feltételek, bírósági iratok, szabályzatok	<ul style="list-style-type: none"><li>• Szerződéstípusok,</li><li>• Személyes-érzékeny szövegrészek,</li><li>• Nyelv / többnyelvűség detekciója</li></ul>
	Pénzügyi riportok, jelentések, elemzések, biztosítási kötvények, követelések	<ul style="list-style-type: none"><li>• Pénzügyi hírek, szentiment / emóció tartalmuk</li><li>• Biztosítási követelések, csalásfelderítés</li></ul>
	Orvosi nyilvántartások, feljegyzések, értékelések, diagnózisok, receptek, beteg űrlapok	<ul style="list-style-type: none"><li>• Orvosi szakterületek,</li><li>• Betegségkockázati szintek</li></ul>
	Termékleírások, számlák, beszerzési-logisztika, visszaküldési-visszatérítési szabályzatok	<ul style="list-style-type: none"><li>• Vásárlói vélemények, visszajelzések</li><li>• Termék / szolgáltatás leírások</li><li>• Beszerzési, operációs szerződések prioritizálása</li></ul>

# Use case #2: Szövegosztályozás



## Szemantika / struktúra

Szöveg struktúra,  
szógyakoriság, vagy jelentés  
fontosabb?



## Input adat

Teljes szöveg kell?  
Első bekezdés, esetleg a cím  
önmagában elegendő?



## Címkék és minőségük

Annotált adatok aránya  
Használható ZSL?  
Kevés adat végett szükséges FSL?

# Use case #3: Generative AI

## Korpusz jellemzői

## Alkalmazott Generative AI



Szerződések, megállapodások, felhasználási feltételek, bírósági iratok, szabályzatok

- Szerződés összegzés, kérdezés
- Szerződés kiegészítése, módosítási javaslat



Pénzügyi riportok, jelentések, elemzések, biztosítási kötvények, követelések

- Pénzügyi riportok automatikus generálása,
- Elemzések összegzése, információkinyerés
- Személyre szabott banki-biztosítási ajánlatok



Orvosi nyilvántartások, feljegyzések, értékelések, diagnózisok, receptek, beteg űrlapok

- Orvosi jelentés betegadatok alapján
- Orvosi jelentések összegzése, kérdések feltevése
- Orvosi szakvélemény kérése ChatBot felületen

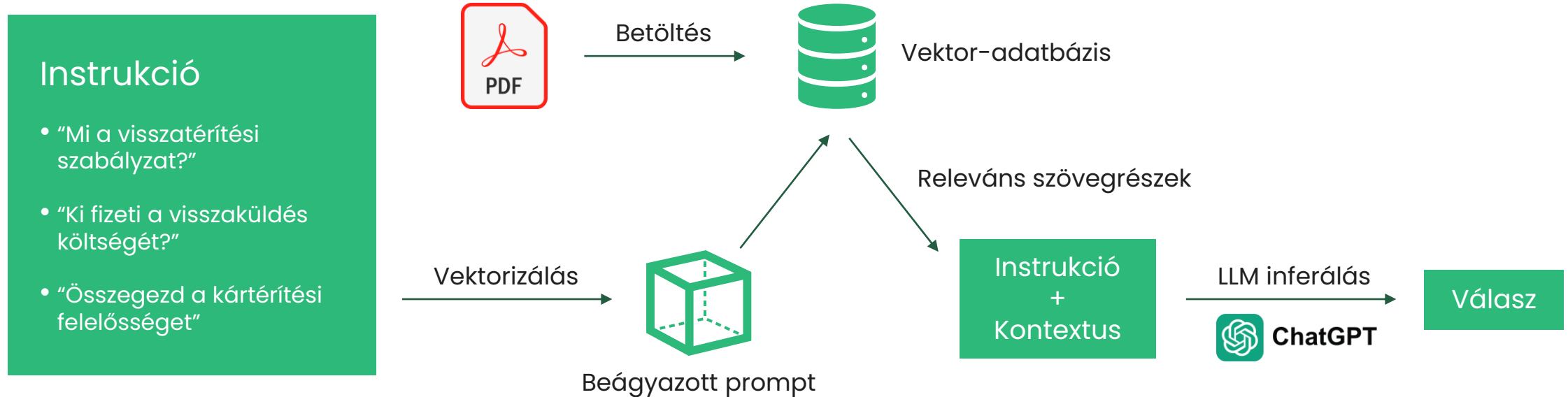


Termékleírások, számlák, beszerzési-logisztika, visszaküldési-visszatérítési szabályzatok

- Termékleírások generálása
- Közösségi media tartalomgyártás,
- Marketing kampányok megfogalmazása



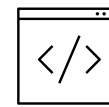
# Use case #3: Generative AI



**Tartalom érzékenysége**  
Harmadik félnél feldolgozhatjuk adatainkat?



**Költségek**  
API hívások vagy saját GPU-n futtatás egységára alacsonyabb?



**Finomhangolás vs Prompting**  
Hangoljuk-e az LLM-et saját korpuszunkon, vagy promptba pakoljuk az inputot?



**Validálás**  
Miképp győződjünk meg a generált szöveg minőségéről?

# Use case #3: Generative AI Demo

 **OpenAI**

+



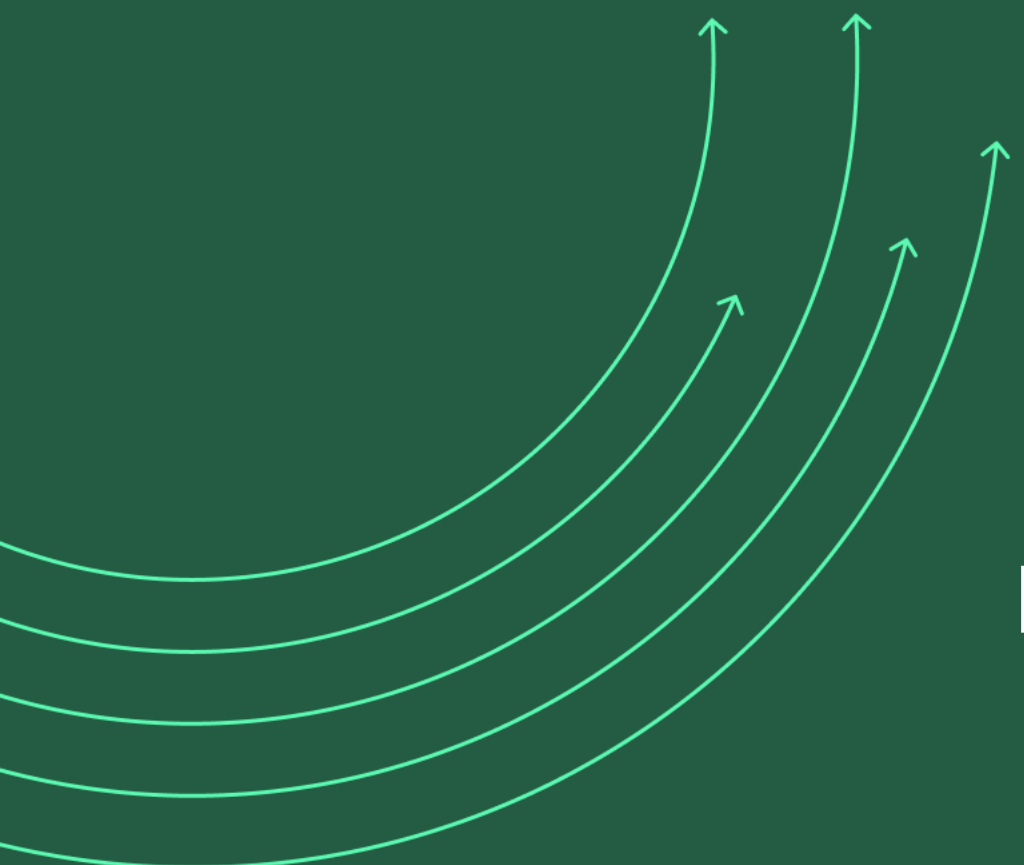
**LangChain**

+



**Kommunikáljunk a szerződéseinkkel,  
dokumentumainkkal!**

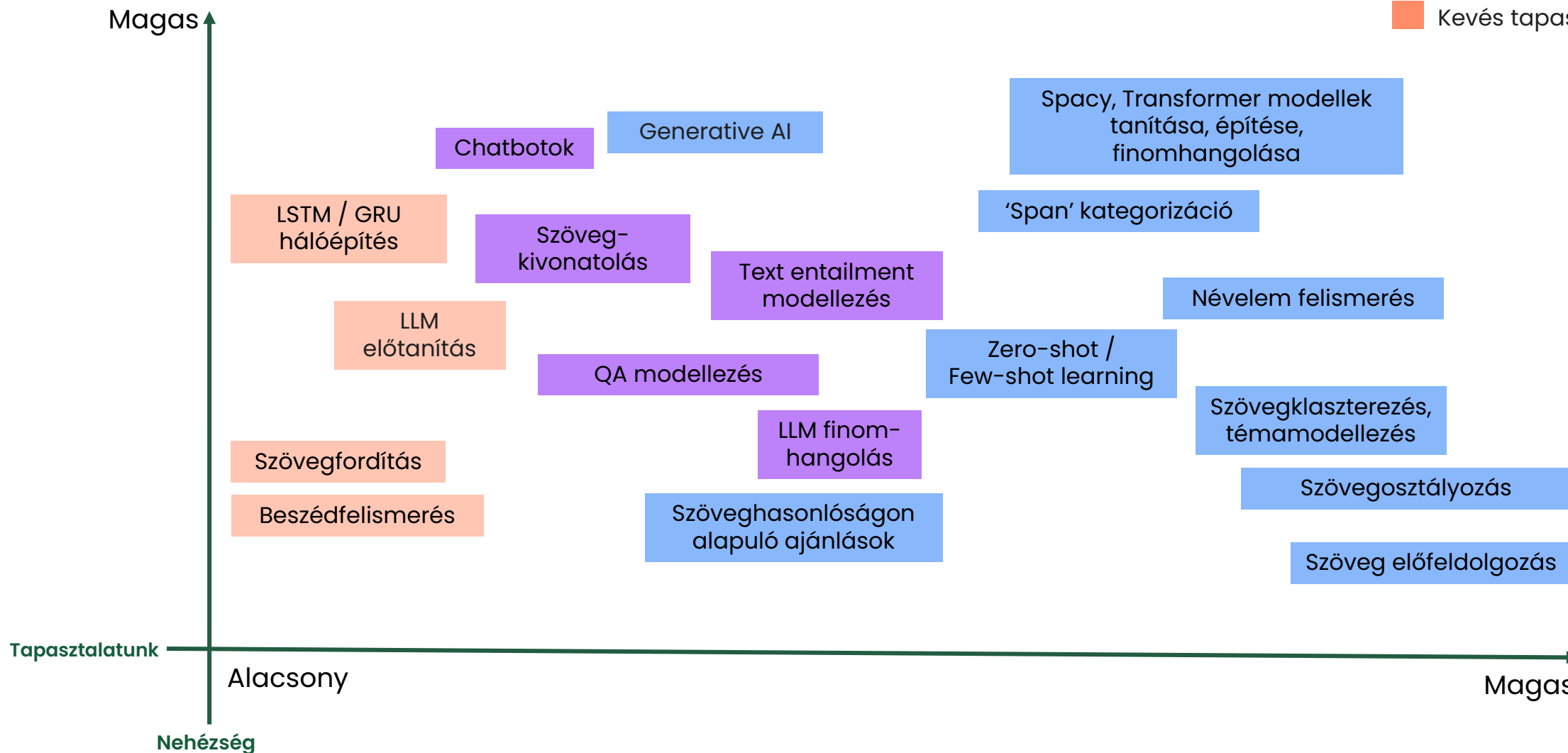
- Összegzés
- Kérdések feltevése
- Információ kinyerése
- Draftolás / sablonírás / módosítás



Hiflylabs x NLP

# Mit tudunk mi, >1 év külföldi NLP tapasztalat után?

- Fejlett, gyakorlott tudás
- Kompetens
- Kevés tapasztalat



# Milyen folyamat vezet az MI-vezérelt szoftverhez?

## Nyers szöveg integrációja

## Korpusz építés

## Modellfejlesztés

## Termékesítés



### Leírás

Többféle formátumban tárolt nyers szöveg gyűjtése és feldolgozása

Tisztítás és transzformáció, feldolgozható, modellezhető alakra hozatal, változók, címkék mellérendelése

Többféle LLM modellel való kísérletezés, ha szükséges előtanítás, finomhangolás, változó- és korpusz-igazítás

PoC → termékfunkció  
NLP megoldás integrációja  
Teljesítménykövetés, modell és korpusz frissítés



### Módszertan

Szövegfájlok  
OCR  
Kiterjesztések  
Speech-to-text  
Voice-to-text

Helyesírás, szegmentáció, stop-szavak, lemmatizáció, normalizáció; **Beágyazások** (szemantikus és struktúrális; tf-idf, BERT, word2vec, stb...)

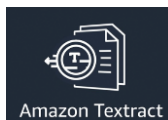
Szófaj- és névelem-felismerés, 'span' detekció, szöveg-osztályozás, szentiment-analitika, szövegösszegzés, témamodellezés, Generative AI

NoSQL adatbázisok, AWS Lambda, Docker konténerizáció, API, stb...



### Technológia

ABBYY



Hugging Face



spaCy

OpenAI



mongoDB



AWS Lambda



FastAPI

# Hiflylabs

Creating business value **in a world full of data**. We engineer data solutions and build apps that make an impact.

## Areas of competence:



Data Platform



Advanced Analytics



Analytics Engineering



Data Visualization



Web Application



Management & Digital Consulting



Mobile



IT & Support

Our primary focus is **Data**

**20+**  
years' experience  
in data industry

**150+**  
staff of data  
experts

**500+**  
data projects  
delivered



# Hiflylabs.

[hiflylabs.com/contact\\_us](https://hiflylabs.com/contact_us)

[linkedin.com/company/hiflylabs](https://linkedin.com/company/hiflylabs)



[linkedin.com/in/kristofrabay](https://linkedin.com/in/kristofrabay)

[github.com/kristofrabay](https://github.com/kristofrabay)