

# How to Build a Smart Search System Using SAS<sup>®</sup> Viya<sup>®</sup> With Microsoft Azure and SingleStore DB

Paul Koot, Notilyze

Michel Philippens, SAS

#ExploreSAS

# SAS EXPLORE

Level Up Your Skills in AI and Analytics

Sept. 11-14 // Las Vegas





**notilyze**

FORWARD THROUGH DATA

# About Us

- Since 2016
- 50+ Clients
- 20 Employees
- Europe
- Local datacenters
- SAS partnership
- Security focus



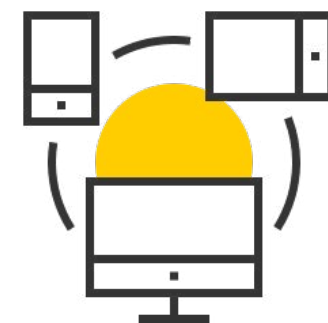


We operationalize data *as a service*, using:



**ANALYTICS**

We offer end-to-end services on our Analytics Platform based on SAS technology.



**SOFTWARE DEVELOPMENT**

We believe in smart tailor-made solutions.

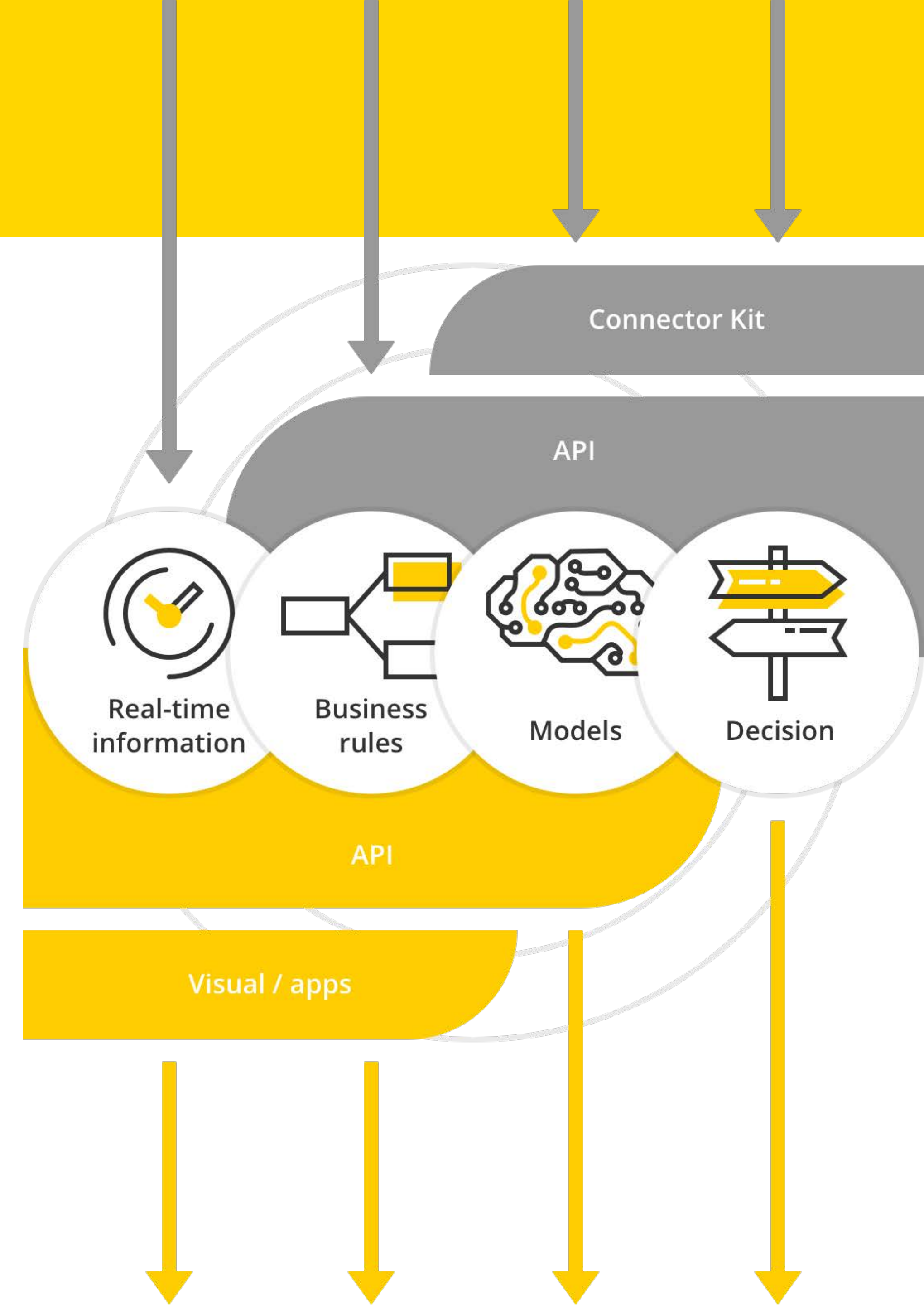


**CLOUD**

A robust infrastructure determines the user experience of your application.

# Analytics Services

- Energy pattern recognition
- Text analytics services
- Payment insights
- Customer analysis
- Factory optimization
- IoT device monitoring



# Business Context Customer

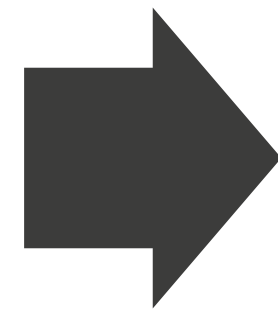


FEDERALE OVERHEIDSDIENST  
Werkgelegenheid, Arbeid  
en Sociaal Overleg



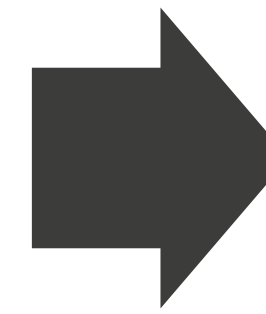
# Business Context

1. Legal experts receive various questions on labor laws



2. Legal experts search Q&A database

Vraag #	Omschrijving	Jaar	Type document	CLETA code	Vraagsteller	Status	Gerelateerde documenten
42507	Art. 30 §1 van de wet van 3 juli 1978 betreffende de arbeidsovereenkomsten (verder: arbeidsovereenkomstenwet) - Koninklijk Besluit van 28 augustus 1963 betreffende het behoud van het normaal van de werklieden, de dienstboden, de bedienden en de werknemers aangeworven voor de dienst op binnenscheep, voor afwezigheidsdagen ter gelegenheid van familiegebeurtenissen of voor de vervulling van staatsburgerlijke verplichtingen of van burgerlijke opdrachten - Kort verzuim/klein verlof - Huwelijk in het buitenland	2019	Briefwisseling	Teaag kort verzuim	Mathias Roman	Officieel beantwoord	
43724	Art. 30, § 1 WAO van 3 juli 1978 - K.B. Klein Verlet van 28 augustus 1963 - Overtijden vader - Crematie en begrafenis op verschillende daten	2020	Briefwisseling	Teaag kort verzuim Ti. Covid-19 maatregelen	Anan Roelens	Officieel beantwoord	
46457	Wet van 27 juni 2021 tot uitbreiding van het rouwverlof bij het overlijden van een partner of een kind en tot het flexibiliseren van de opname van het rouwverlof - Wijziging van het KB van 28 augustus 1963 betreffende het behoud van het normaal van de werklieden, de dienstboden, de bedienden en de werknemers aangeworven voor de dienst op binnenscheep, voor afwezigheidsdagen ter gelegenheid van familiegebeurtenissen of voor de vervulling van staatsburgerlijke verplichtingen of van burgerlijke opdrachten - Wijziging van de artikelen 52, 70 en 71 van de wet van 3 juli 1978 betreffende de arbeidsovereenkomsten - Inwerking van de periode van het gewaarborgd loon ingeval het rouwverlof wordt gevolgd door ziekteverlof	2021	Briefwisseling	Teaag kort verzuim	Anan Roelens	Officieel beantwoord	
39452	Artikel 30 §1 van de wet van 3 juli 1978 betreffende de arbeidsovereenkomsten - KB van 28 augustus 1963 betreffende het behoud van het normaal loon van de werklieden, de dienstboden, de bedienden en de werknemers aangeworven voor de dienst op binnenscheep, voor afwezigheidsdagen ter gelegenheid van familiegebeurtenissen of voor de vervulling van staatsburgerlijke verplichtingen of van burgerlijke opdrachten - Klein verlof - Beging 'kwalificatiedienst' - Racht op klein verlatekort verzuim - Huwelijk broer - Nachtoploegen - Priestertes die verspreid liggen over twee astronomische dagen - Artikel 30, § 1 van de wet van 3 juli 1978 betreffende de arbeidsovereenkomsten - Koninklijk besluit van 28 augustus 1963 betreffende het behoud van het normaal loon van de werklieden, de dienstboden, de bedienden en de werknemers aangeworven voor de dienst op binnenscheep, voor afwezigheidsdagen ter gelegenheid van familiegebeurtenissen of voor de vervulling van staatsburgerlijke verplichtingen of van burgerlijke opdrachten	2017	Briefwisseling	Teaag kort verzuim	An Sofie Craninx Familiehulp vzw	Officieel beantwoord	
39934	Art. 30, § 1 van de wet van 3 juli 1978 betreffende de arbeidsovereenkomsten - KB van 28 augustus 1963 betreffende het behoud van het normaal loon van de werklieden, de dienstboden, de bedienden en de werknemers aangeworven voor de dienst op binnenscheep, voor afwezigheidsdagen ter gelegenheid van familiegebeurtenissen of voor de vervulling van staatsburgerlijke verplichtingen of van burgerlijke opdrachten	2017	Briefwisseling	Teaag kort verzuim	Kirsty Beutens, Group S	Officieel beantwoord	
44064	Art. 30, § 1 van de wet van 3 juli 1978 betreffende de arbeidsovereenkomsten - KB van 28 augustus 1963 betreffende het behoud van het normaal loon van de werklieden, de dienstboden, de bedienden en de werknemers aangeworven voor de dienst op binnenscheep, voor afwezigheidsdagen ter gelegenheid van familiegebeurtenissen of voor de vervulling van staatsburgerlijke verplichtingen of van burgerlijke opdrachten - Klein verlof - Overtijden adoptievader	2020	Briefwisseling	Teaag kort verzuim	minnow De Boeck - Group S	Officieel beantwoord	



3. Legal experts formulate an answer





# Limitations Current Search Process

The screenshot shows a search interface with a table of results. The table has columns for 'Vraag #', 'Jaar', and 'Type docu'. The rows contain various legal questions and their corresponding years and document types. Four yellow callout boxes with question marks are overlaid on the interface, each asking a question about search limitations.

Vraag #	Jaar	Type docu
42507	2019	Briefwissel
43724	2020	Briefwissel
46457	2021	Briefwisseling
39452	2017	Briefwisseling
39934	2017	Briefwissel
44964	2020	Briefwissel

Can we search in one language and get results in two languages?

Can we search based on articles and laws mentioned in the text?

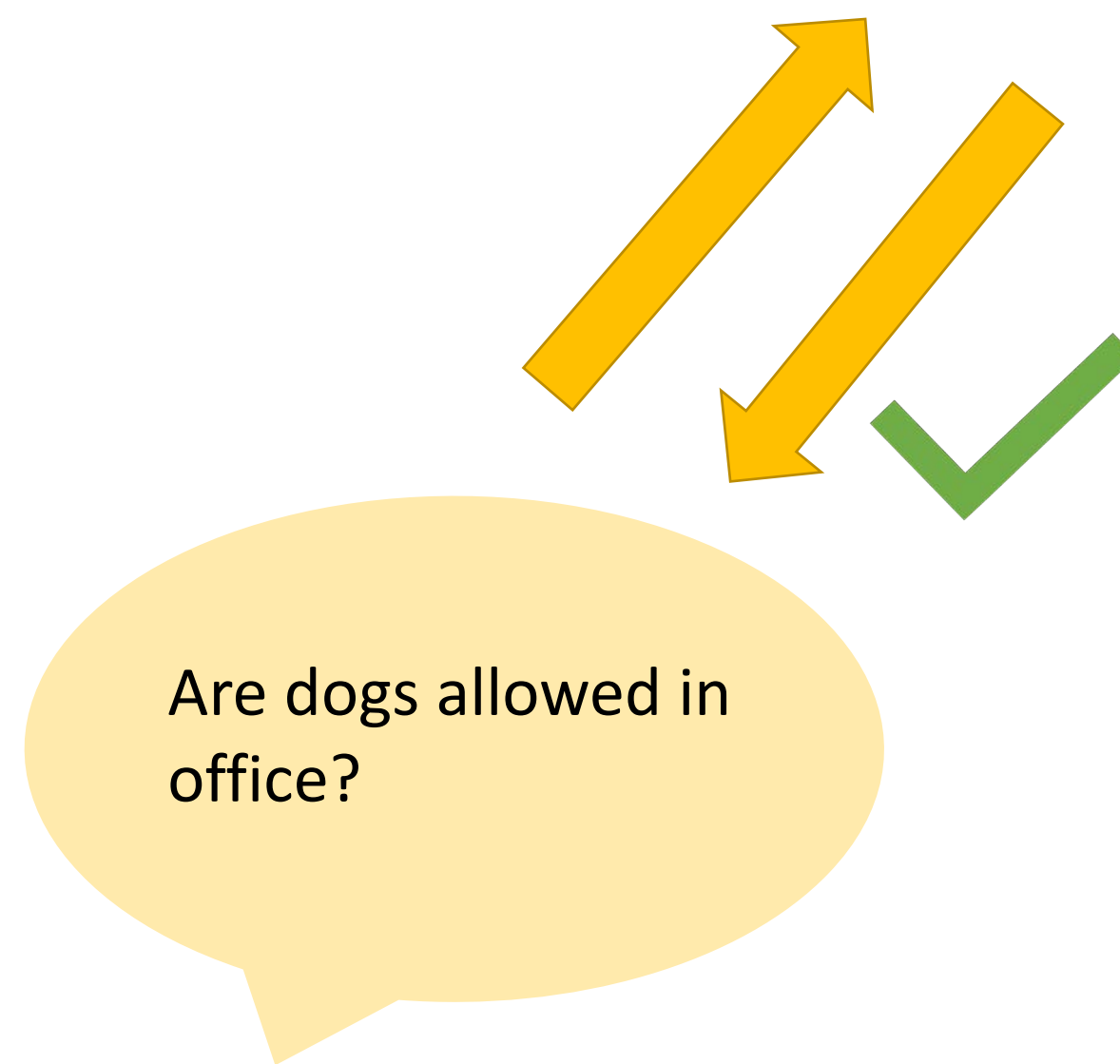
Can we search on multiple keywords?

Can we automate the "labelling" of documents?

# Limitations Current Search Process

## Q&A Database

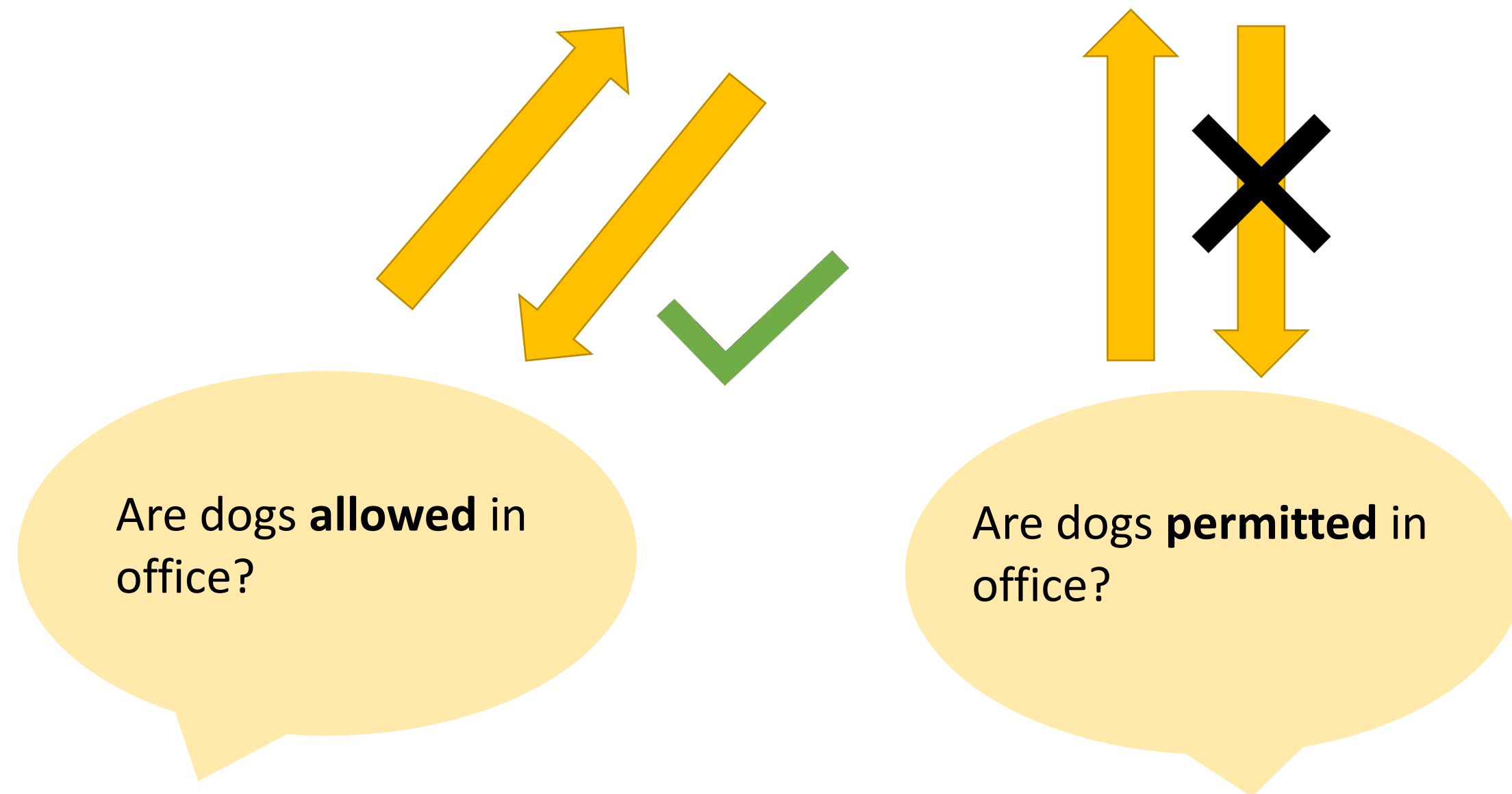
	Question	Answer	Label
1	Are dogs allowed in office?	Yes, according to Law 3.4 from...	Dogs
2	Holidays for 2nd wedding?	No, according to Law 6.5.4 ...	Wedding
3	....	....	....



# Limitations Current Search Process

## Q&A Database

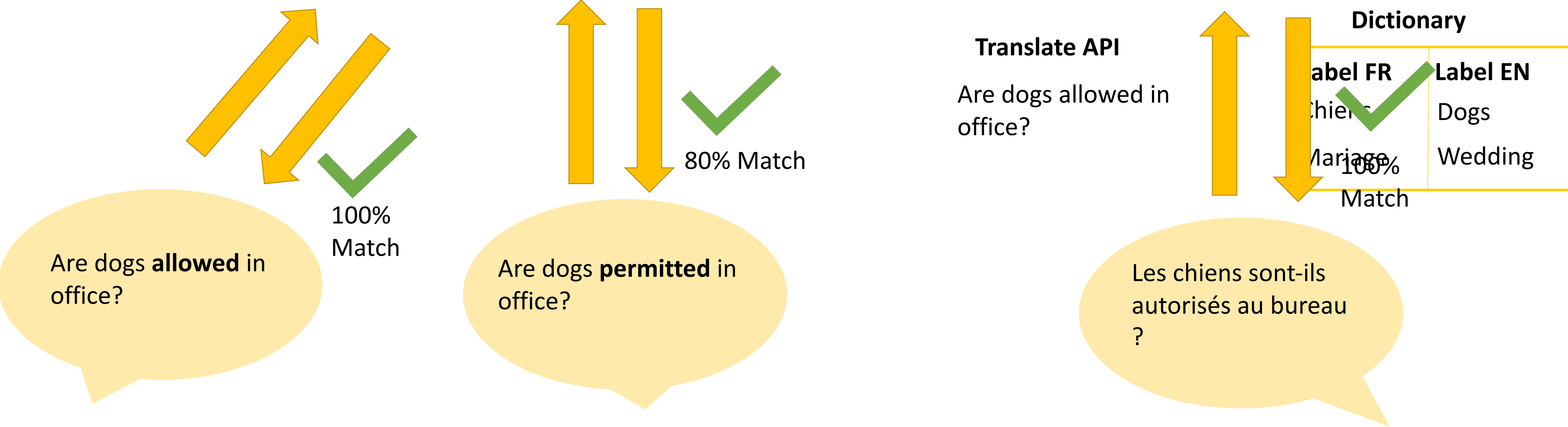
	Question	Answer	Label
1	Are dogs allowed in office?	Yes, according to Law 3.4 from...	Dogs
2	Holidays for 2nd wedding?	No, according to Law 6.5.4 ...	Wedding
3	....	....	....



# New Solution

## Q&A Database System

	Question	Answer	Label
1	Are dogs allowed in office?	Yes, according to Law 3.4 from...	Dogs
2	Holidays for 2nd wedding?	No, according to Law 6.5.4 ...	Wedding
3	....	....	....



# ChatGPT to the Rescue?

MI

Dear ChatGPT, I have a question about the interpretation of Belgian labor law, specifically regarding the law of October 3, 2022. It concerns the following question: Does the right to 3 instances of taking the first sick day without a certificate reset when changing employers throughout the year, or not? Can you provide an answer in no more than 10 lines?



I'm not aware of any specific law passed on October 3, 2022, as my knowledge goes up until September 2021. However, based on general principles of Belgian labor law, rights such as taking the first sick day without a certificate are typically associated with the individual employee and are not usually reset upon changing employers during the year. These rights are often considered personal and continuous, regardless of employment changes, but the specific interpretation would depend on the exact provisions of the law in question. For accurate and up-to-date information, consult legal experts or official sources.

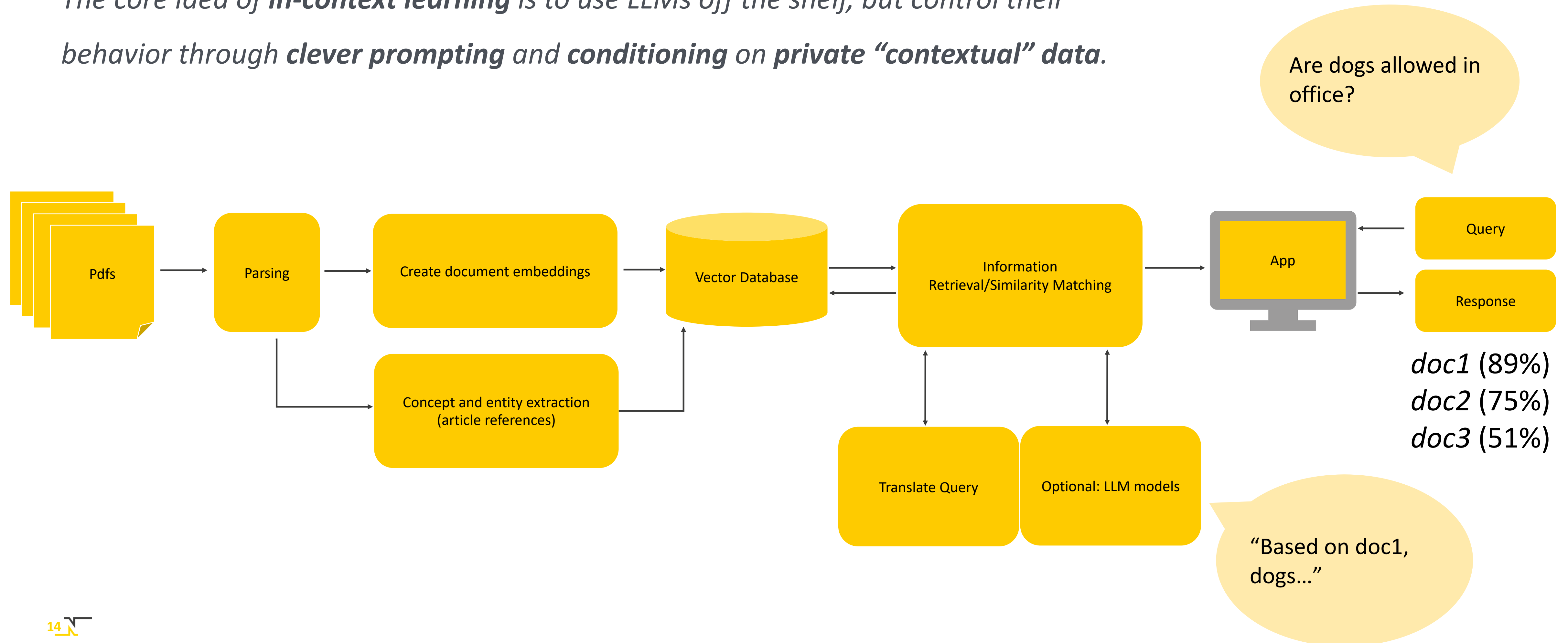


Off-the-shelf LLM:

- Risk of hallucination
- Training data is limited to public data and limited in time
- Lack of transparency: on what data was this model trained?
- Training on business context is expensive (cost + compute cost)

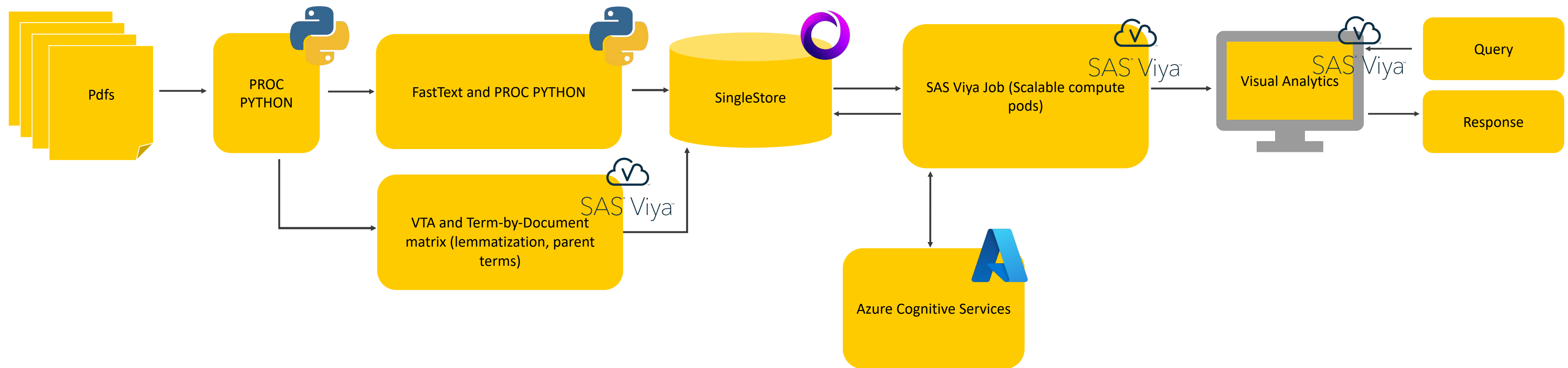
# Enter Smart Search (In-Context Learning)

The core idea of *in-context learning* is to use LLMs off the shelf, but control their behavior through *clever prompting* and *conditioning on private “contextual” data*.



# Enter Smart Search (In-Context Learning)

*The core idea of **in-context learning** is to use LLMs off the shelf, but control their behavior through **clever prompting** and **conditioning** on private “contextual” data.*



# SAS Viya Job (Scalable Compute Pods)

- Upon starting job; start pod = overhead
- Reusable pod: always standby
- Saves 10 seconds startup time
- Multiple reusable pods could be reserved for the job
- Server resources dedicated to system can be up/downscaled



# Word/Document Embeddings – 2 ways

TF-IDF:

- Term-by-document matrix
- Based on ‘Parent terms’
- Lemmatization
- “Old school” word embeddings

$$\log \left( \frac{\#term\_match_{doc}}{\#term\_total_{doc}} / \frac{\#doc\_match_{corpus}}{\#doc\_total_{corpus}} \right)$$

TF=Term

Frequency

IDF=Inverse

Document Frequency

Embeddings:

- FastText<sup>[1]</sup>
- SIF<sup>[2]</sup>
- Cosine similarity

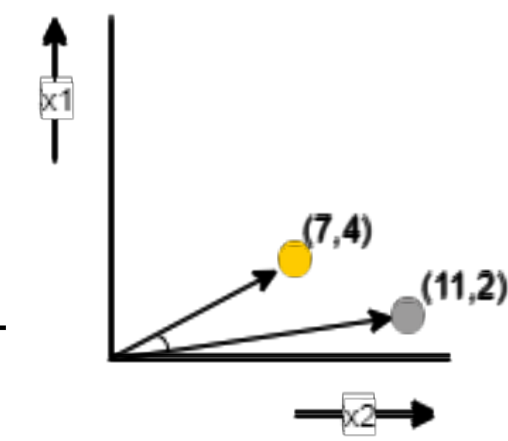
John is **cycling** to school every day.

Anne was **driving** to school, when...

$$\vec{A} = [7,4]$$

$$\vec{B} = [11,2]$$

$$\cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|}$$



[1] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, 5, 135-146.  
[2] Arora, S., Liang, Y., & Ma, T. (2017). A simple but tough-to-beat baseline for sentence embeddings. In *International conference on learning representations*.

# Open-Source Integration in the Cloud

The screenshot displays the SAS Environment Manager interface for monitoring jobs. The main content area is titled "Jobs and Flows" and is divided into "Monitoring" and "Scheduling" tabs. The "Monitoring" tab is active, showing a list of jobs under the heading "Monitor (6)".

**Jobs Filter:**

- Statu... (Refresh)
- Canceled
- Succes...
- Failed
- Pending
- Running
- Timed ...
- Unknown

**Created By Filter:**

- Filter
- admin
- admi...
- admi...
- amali...
- arjen....
- asch...

**Table Data:**

Name	Start Date	Run Time	Status	Created By	Log
02trainModelFastTextEmbeddingsWordsEP300.sas	August 23, 2023 09:33:03 PM	07:01:27.160	Successful	paul	<a href="#">Download</a>
02trainModelFastTextEmbeddingsWordsEP200.sas	August 23, 2023 05:39:01 PM	04:22:40.370	Successful	paul	<a href="#">Download</a>
02trainModelFastTextEmbeddingsWordsEP150.sas	August 23, 2023 05:38:58 PM	03:15:31.735	Successful	paul	<a href="#">Download</a>
02trainModelFastTextEmbeddingsWordsEP100.sas	August 22, 2023 09:56:31 PM	02:16:00.894	Successful	paul	<a href="#">Download</a>
02trainModelFastTextEmbeddingsWords.sas	August 22, 2023 09:56:22 PM	01:04:52.798	Successful	paul	<a href="#">Download</a>
02trainModelFastTextEmbeddingsWords.sas	August 22, 2023 11:21:37 AM	00:28:45.704	Successful	paul	<a href="#">Download</a>

**Additional UI Elements:**

- Search bar: 02trainModelFastTextEmb...
- Filter by: August 17, 2023 to August 24, 2023
- View: 20 (Dropdown)
- Displaying 1-6 (Navigation)
- Apply button

# Demo – VA Dashboard

**Smart Search System FOD WASO v2.0**

1994 tot 2023

Filter complete database

**Filter complete database - FOD WASO**  
Gebruik bovenstaande filters om te zoeken op CLETA codes, genoemde wetsartikelen in de documenten of andere eigenschappen. Dubbelklik op een document in de lijst voor een preview.

Nummer van de vraag	Omschrijving	Jaar	Subtype	CLETA Codes	Wetsartikelen
01022016	Sociale verkiezingen - Artikel 35 van de wet van 3 juli 1978 betreffende de arbeidsovereenkomsten - Occulte periode - Ontslag om dringende reden - Werkwijze - Belang van aanvraag tot re-integratie - Opstart specifieke procedure beschermde werknemers na re-integratie - Rechtspraak Hof van Cassatie	2016	Briefwisseling	3aif, 3aic	Art. 35 van de wet van 03
01032016	Elections sociales - Art. 14 & 30 Loi du 4.12.2007 - Détermination de l'horaire : désaccord non constaté - interpellation du CLS hors délai - absence de recours	2016	Briefwisseling	3acba, 3acha	Art. 30 de la loi du 04/12,
01042019	Sociale verkiezingen - Drempelberekening - Begrip werknemer - Opleidingen, leerlingen en kleine statuten	2019	Briefwisseling	3abb	De wet van 04/12/2007, l De wet van 01/01/2020.
01042019	Sociale verkiezingen - Drempelberekening - Begrip werknemer - Opleidingen, leerlingen en kleine statuten	2019	Briefwisseling	3abb	
01122015	Elections Sociales - Condition d'éligibilité (ancienneté) - ancienneté conventionnelle suite à reprise après faillite - Article 19 de la Loi du 20 septembre 1948 portant organisation de l'économie	2015	Briefwisseling	3abf	La loi du 20/09/1948.
02052016	Sociale Verkiezingen - Vervanging kandidaat - Beëindiging van de arbeidsovereenkomst in onderling akkoord - Artikel 38 van de wet van 4 december 2007	2016	Briefwisseling	3acbe, 1faa	
02052016 b	Elections sociales - Art 49, al. 2 Loi 4.12.2007 - Temps de travail Electeurs, membres du bureau et témoins	2016	Briefwisseling	3acbj, 3acbg	La loi du 16/03/1971, Art 04/12/2007.
02122011	Sociale verkiezingen - Drempelberekening - Begrip werknemer - Werknemer met arbeidsovereenkomst met Belgische werkgever die permanent in het buitenland tewerkgesteld is en daar ook onderhevig is aan sociale zekerheid	2011	Briefwisseling	3abf, 3abb, 3abe	

Actieve CLETA filters: 0  
Actieve Wetsartikel filters: 0  
Actieve Document Type filters: 0

**Documenten in geselecteerd dossier**

- 01022016.pdf
- 01032016.pdf
- 01042019.pdf
- 01042019vraag.pdf
- 01122015.pdf
- 02052016\_b.pdf
- 02052016.pdf
- 02072019\_bijlage\_vraag.pdf
- 02072019.pdf
- 03022016.pdf
- 03022016a.pdf
- 03022016b.pdf
- 03032016.pdf
- 03062016\_2.pdf
- 03122015.pdf

# Productivity Increase

The screenshot shows a web interface for document management. At the top left, there is a dropdown menu labeled 'Selecteer' and a search bar with a 'Zoeken' button. Below this is a table with columns for 'Vraag #', 'Jaar', and 'Type docu'. The table contains several rows of data, including question numbers and document types. Four yellow callout boxes with question marks and green checkmarks are overlaid on the interface, each containing a question about search capabilities.

Vraag #	Jaar	Type docu
42507	2019	Briefwissel
43724	2020	Briefwissel
46457	2021	Briefwisseling
39452	2017	Briefwisseling
39934	2017	Briefwissel
44964	2020	Briefwissel

Can we search in one language and get results in two languages?

Can we search based on articles and laws mentioned in the text?

Can we search on multiple keywords?

Can we automate the "labelling" of documents?

# Summary Business Benefits

- **Productivity:**
  - Improved search experience
  - Decreased time to find relevant documents
  - One platform that incorporates end-to-end solution
- **Trust:**
  - Simple transparent algorithms can do the trick
  - No hallucinations
  - Embeddings created from business-specific corpus

# Summary Business Benefits

- **Performance:**
  - Everything in cloud with Viya
    - Fast
    - Scalable
    - Updates remote
    - Secure
  - Vector database with SingleStore
    - Fast embeddings
    - Integrated in Viya
    - All data stored in one place (less data movement)

# Questions?



**Paul Koot**

Data Scientist

[paul@notilyze.com](mailto:paul@notilyze.com)

[www.notilyze.com](http://www.notilyze.com)





Turn your data into **value** and **operationalize** it fast!

#ExploreSAS

# SAS EXPLORE

Level Up Your Skills in AI and Analytics

Paul Koot

Michel Philippens

Joep Bom

Anthony Severyns

Frederik Vandenberghe

Wouter van de Weghe

Wesley Wuyts

Kaat Tastenhoye

Luca Barbanotti

