

Le Forum se déroule avec le soutien de :



La captation du Forum des archivistes est réalisée par Ikonokast





Session

Les attentions au classement:
temporalité et outils

Intervenant : Xavier Gillard -- UCLouvain & Archives de l'Etat (Be)

Lieu : Carré 1

Date : 28 mars 2025



avec
attention
archives, archivistes & sociétés

xelians

SAINT-GOBAIN

naoned



IRON
MOUNTAIN®



Prémium
vision
Archives Evolution

ARKEY c'est un peu une **chasse au trésor**
comme celles de notre enfance ...
Mais ici, **tout fonctionne à l'envers**



Les objectifs d'ARKEY

Interprétation

Comprendre le contenu des documents



Extraction

Obtenir une version des documents qui soit exploitable par une machine



Navigation

Aider l'utilisateur à atteindre son propre objectif de recherche archivistique



Les objectifs d'ARKEY

Interprétation

Comprendre le contenu des documents



Motivation

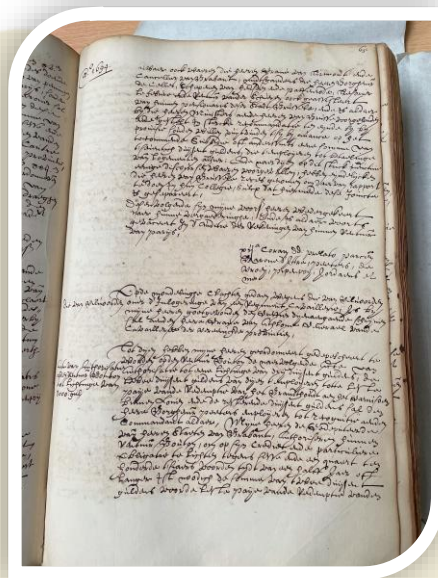


L'approche proposée: Machine-in-the loop

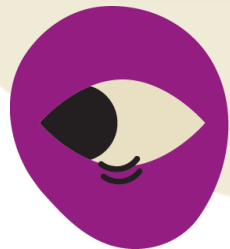
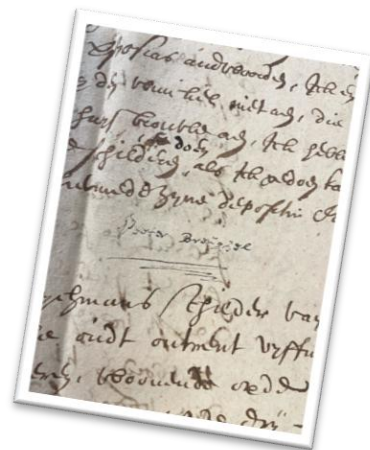
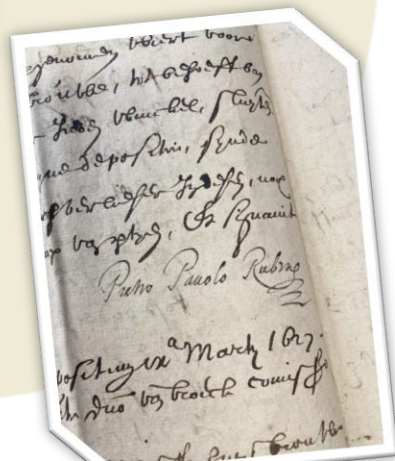


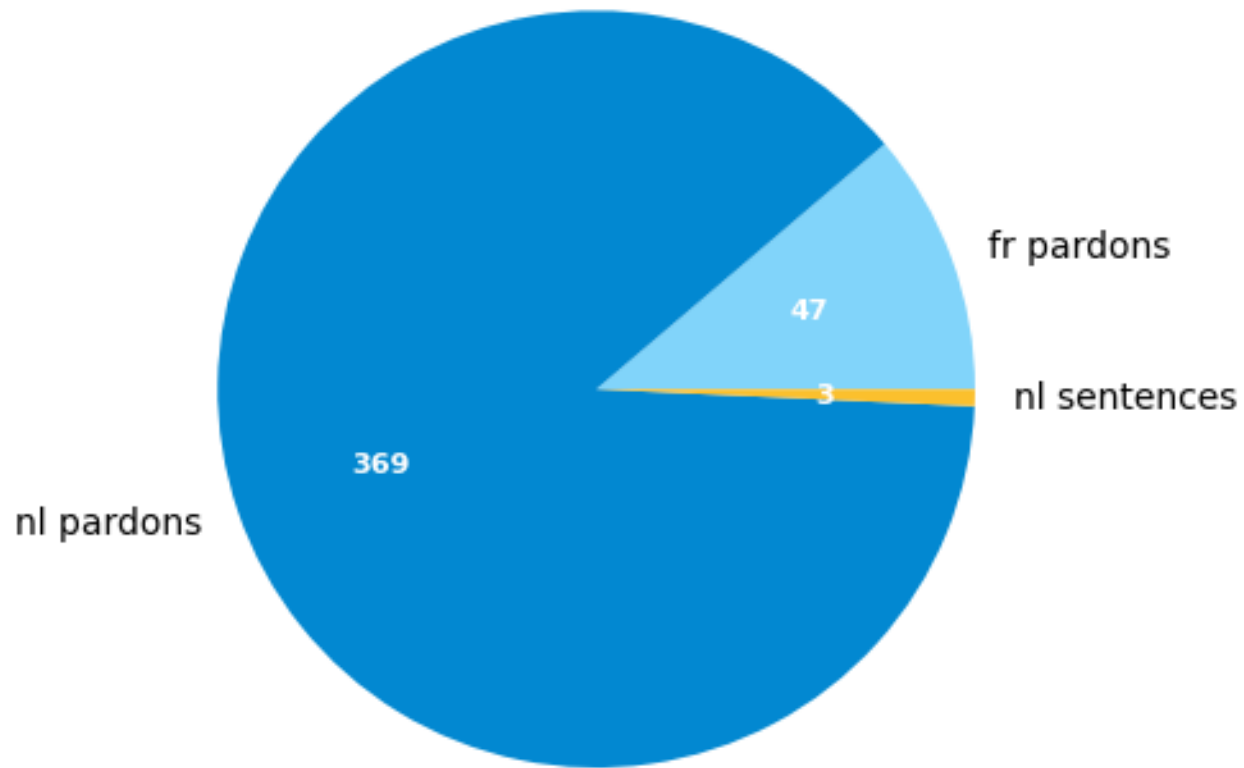
C'est toujours **l'humain qui réalise le travail**, mais il peut compter sur l'aide d'un « **assistant** » automatisé avec qui il peut faire des aller-retours

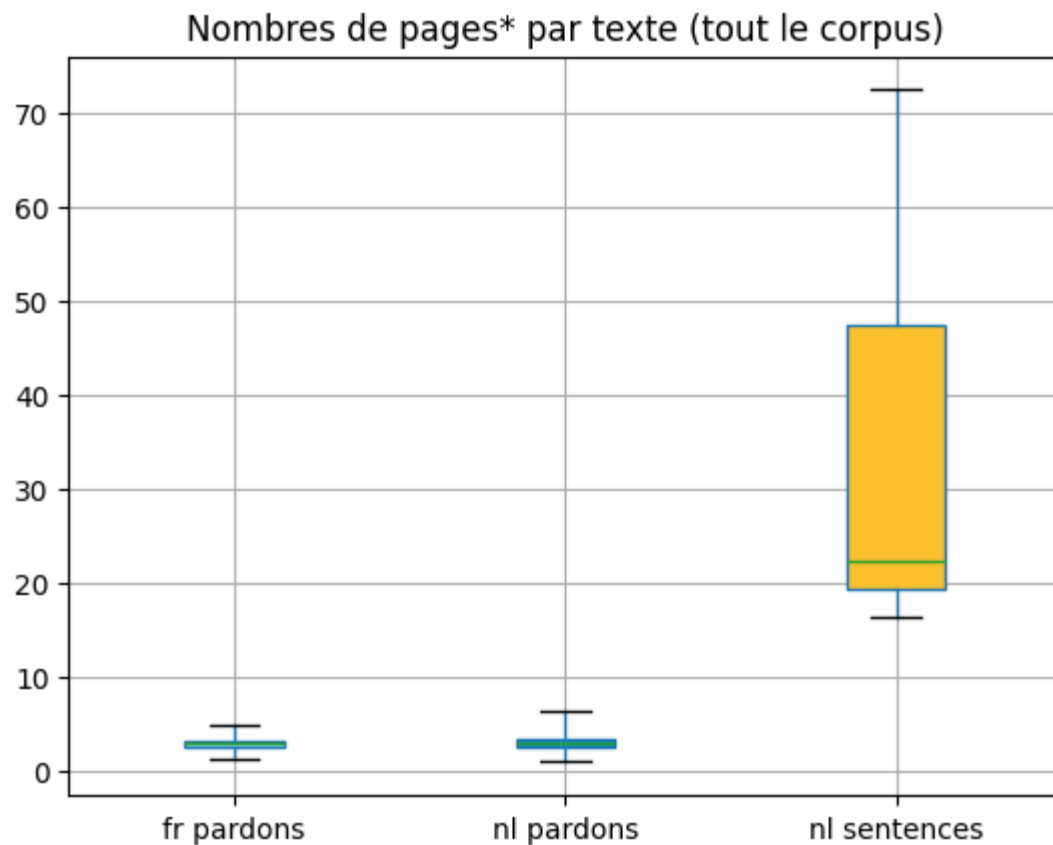
Le corpus



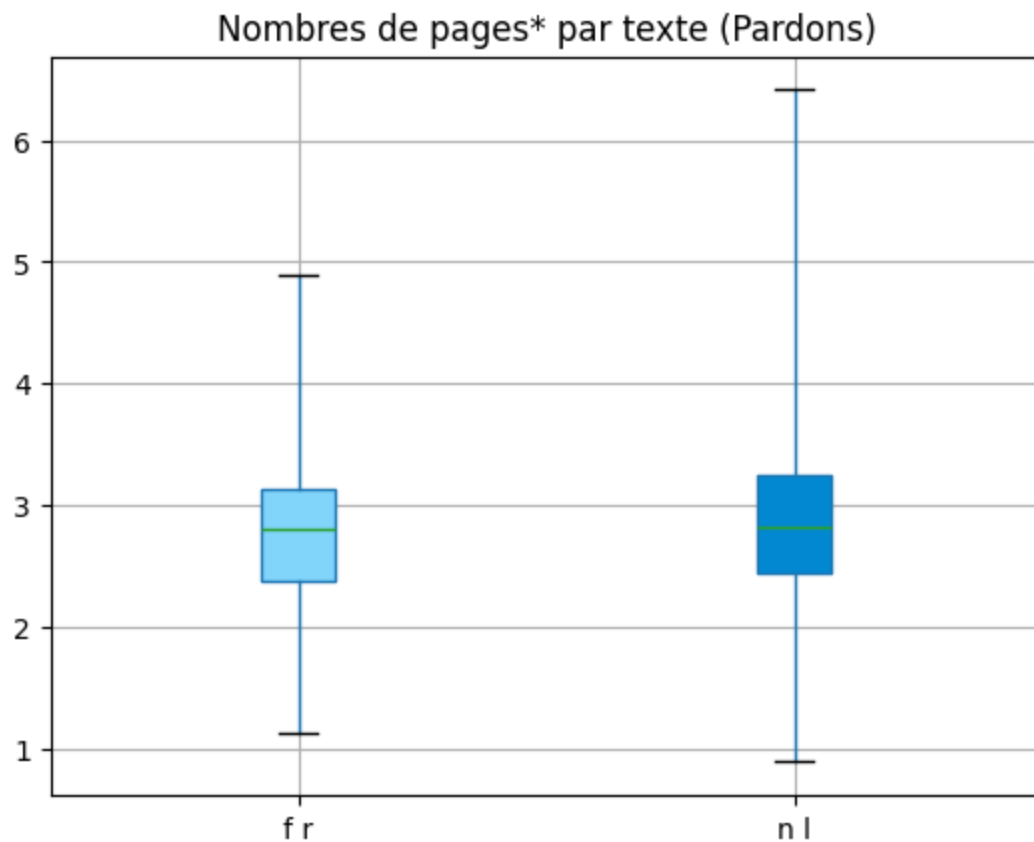
Lettres de pardons et de sentences du conseil de Brabant (XVe XVIIe)





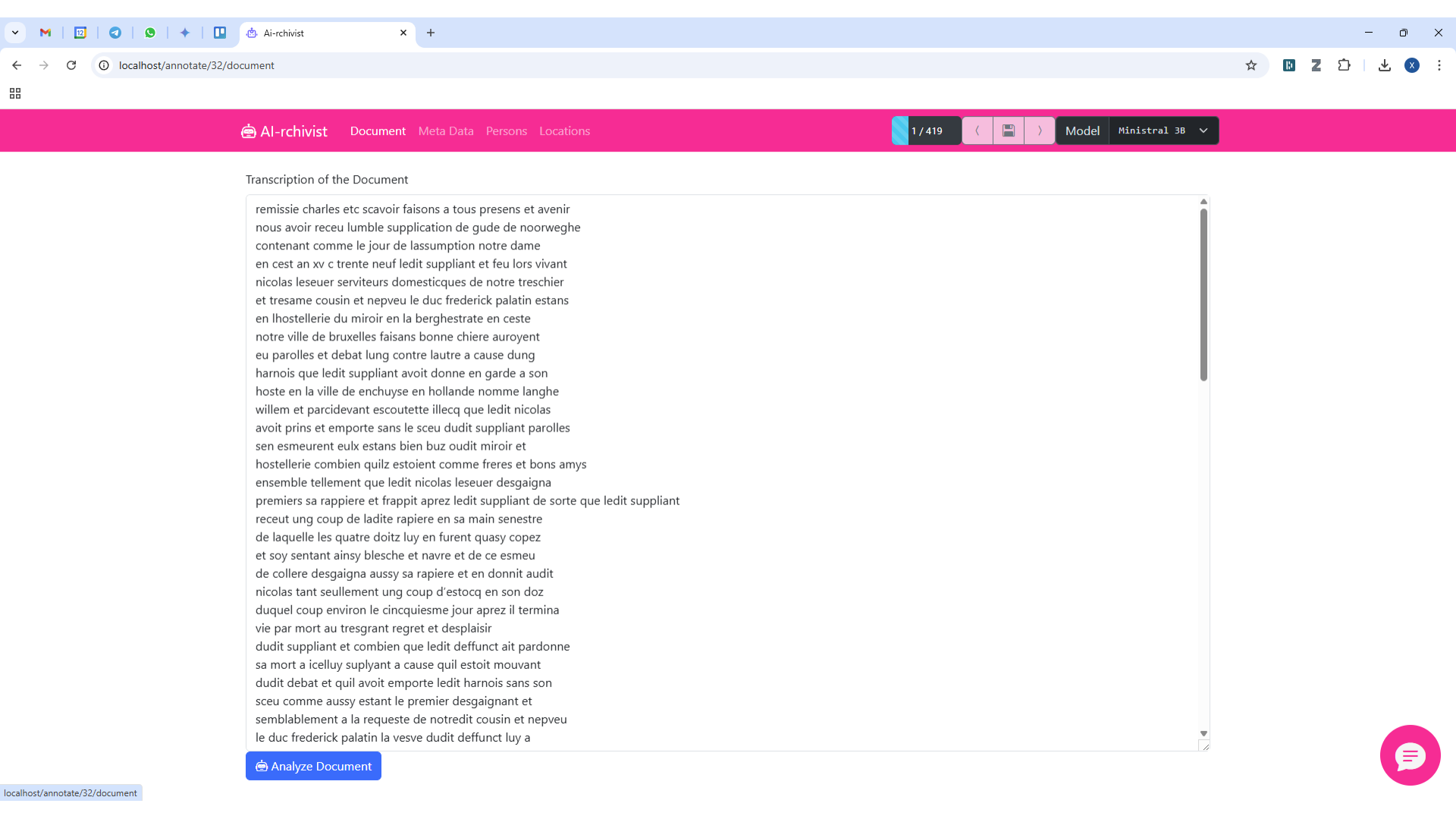


* Hypothèse de 400 mots par page



* Hypothèse de 400 mots par page





Transcription of the Document

remissie charles etc scavoir faisons a tous presens et avenir
nous avoir receu lumble supplication de gude de noorweghe
contenant comme le jour de lassumption notre dame
en cest an xv c trente neuf ledit suppliant et feu lors vivant
nicolas leseuer serviteurs domesticques de notre treschier
et tresame cousin et nepveu le duc frederick palatin estans
en lhostellerie du miroir en la berghestrate en ceste
notre ville de bruxelles faisans bonne chiere auroyent
eu parolles et debat lung contre lautre a cause dung
harnois que ledit suppliant avoit donne en garde a son
hoste en la ville de enchuyse en hollande nomme langhe
willem et parcidevant escoutette illecq que ledit nicolas
avoit prins et emporte sans le sceu dudit suppliant parolles
sen esmeurent eulx estans bien buz oudit miroir et
hostellerie combien quilz estoient comme freres et bons amys
ensemble tellement que ledit nicolas leseuer desgaigna
premiers sa rapiere et frappit aprez ledit suppliant de sorte que ledit suppliant
receut ung coup de ladite rapiere en sa main senestre
de laquelle les quatre doitz luy en furent quasy copez
et soy sentant ainsy blesche et navre et de ce esmeu
de collere desgaigna aussy sa rapiere et en donnit audit
nicolas tant seulement ung coup d'estocq en son doz
duquel coup environ le cinquiesme jour aprez il termina
vie par mort au tresgrant regret et desplaisir
dudit suppliant et combien que ledit deffunct ait pardonne
sa mort a icelluy suplyant a cause quil estoit mouvant
dudit debat et quil avoit emporte ledit harnois sans son
sceu comme aussy estant le premier desgaignant et
semblablement a la requeste de notredit cousin et nepveu
le duc frederick palatin la vesve dudit deffunct luy a

Analyze Document

localhost/annotate/32/metadata

1 / 419ModelMinistral 3B

Type of the DocumentRemission

Act Date

mil cinq cens trente neuf

Facts Date

en cest an xv c trente neuf

Summary in English

The king grants a remission to Gude de Noorweghe for the murder of his servant Nicolas Le Seueur.

Summary in French

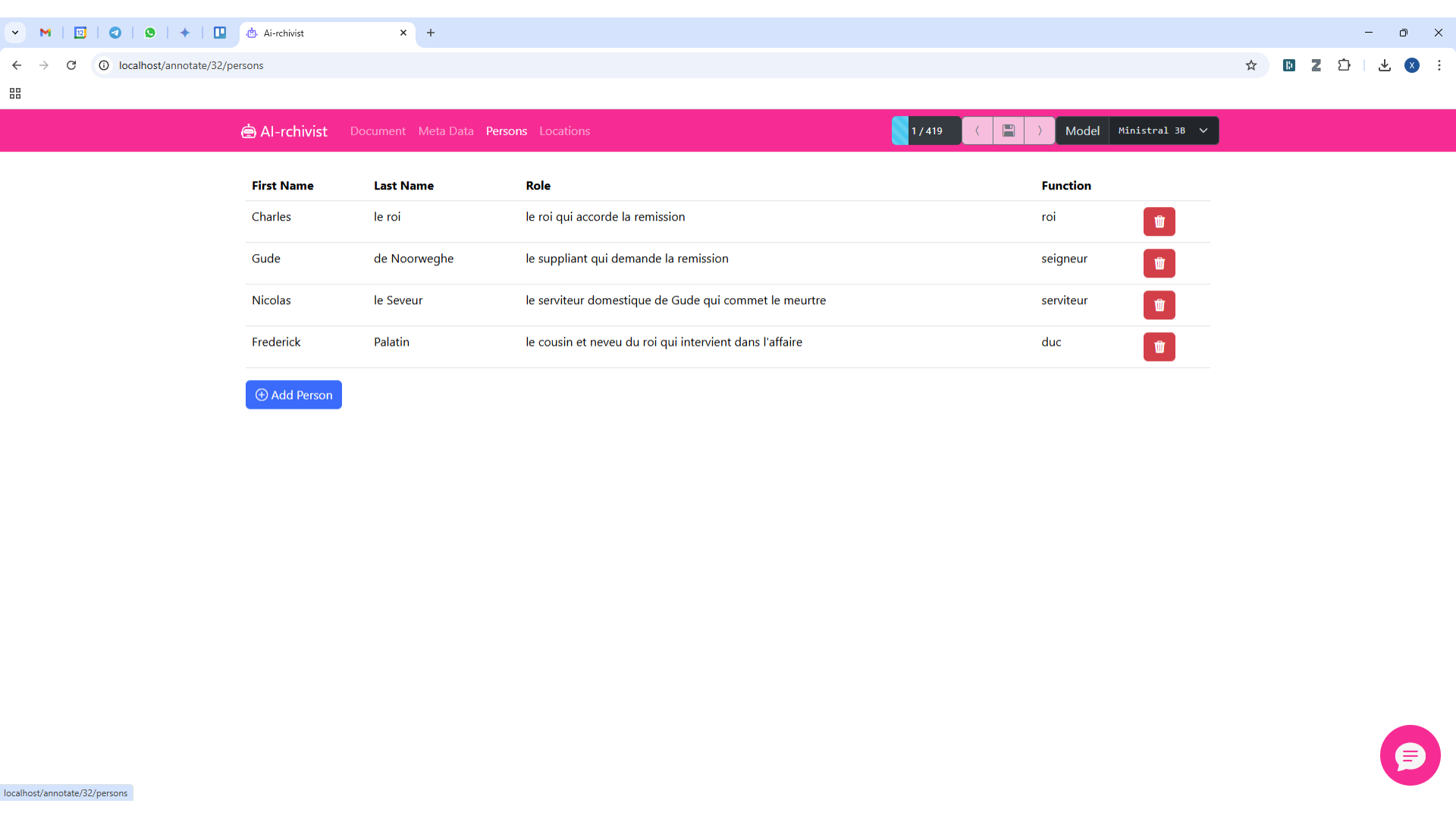
Le roi accorde une remission à Gude de Noorweghe pour le meurtre de son serviteur Nicolas Le Seueur.

Summary in Dutch

De koning verleent een remissie aan Gude de Noorweghe voor de moord op zijn dienaar Nicolas Le Seueur.

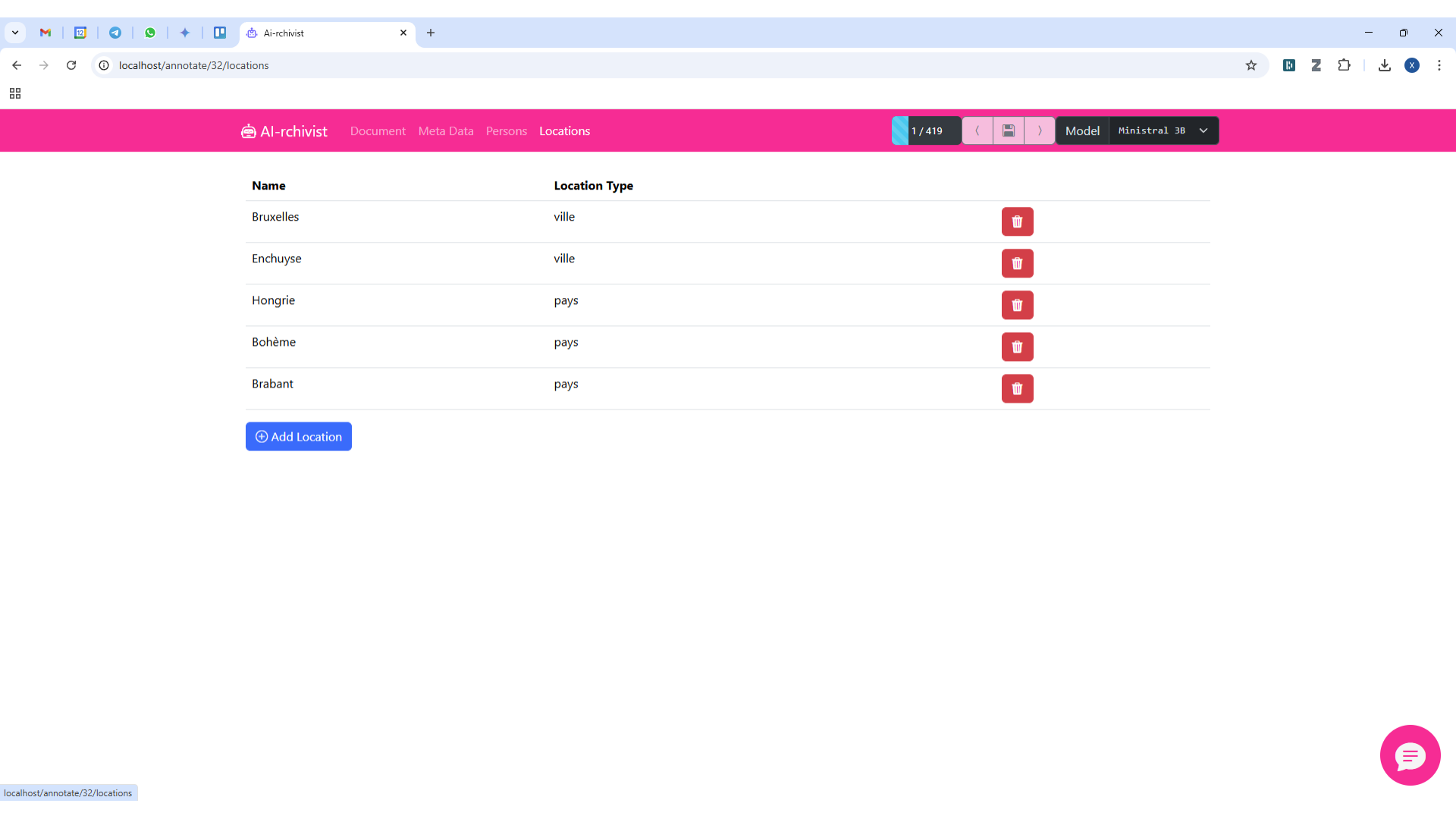
Summary in German

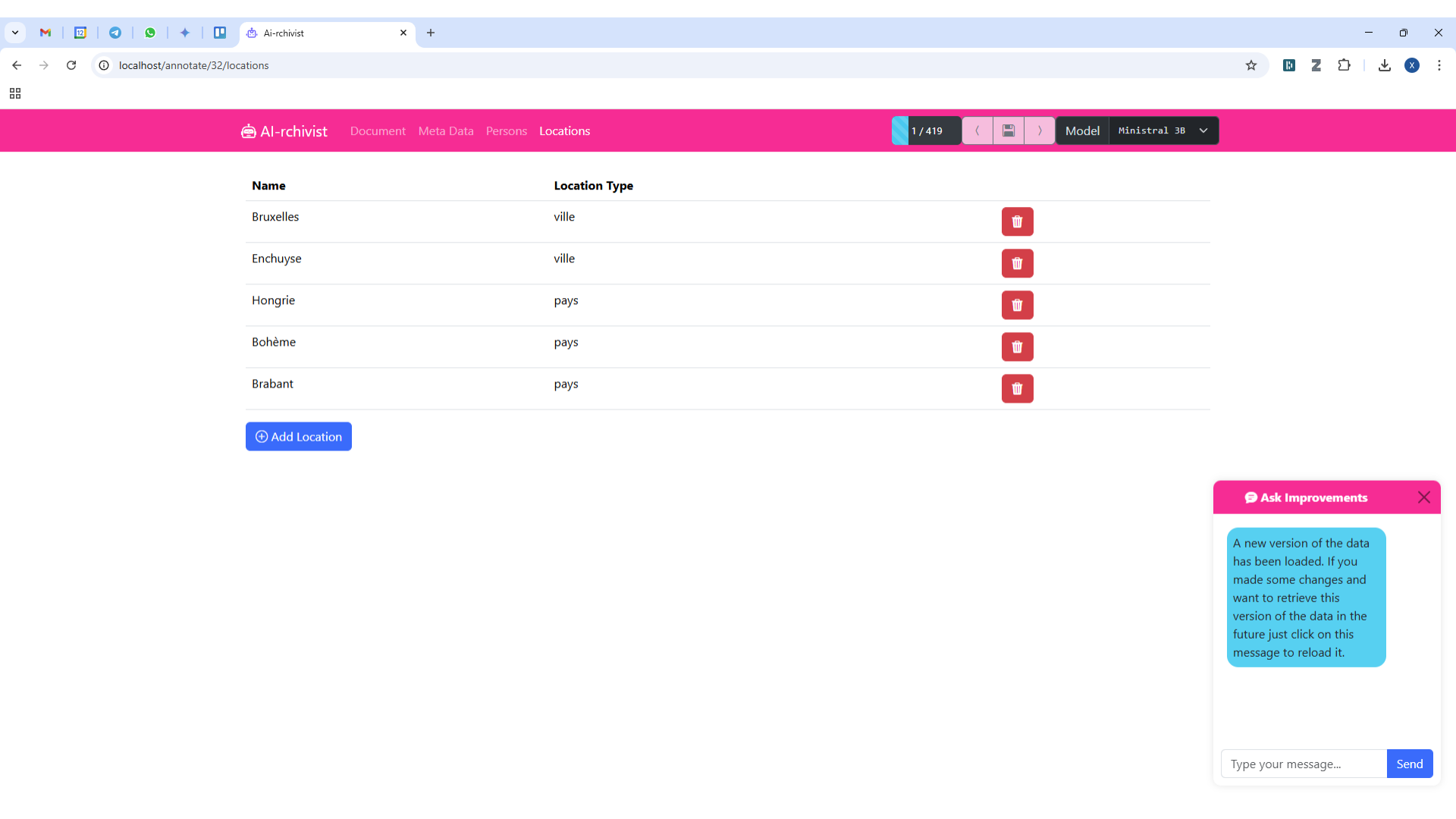
Der König erteilt eine Remission an Gude de Noorweghe für die Ermordung seines Dieners Nicolas Le Seueur.



First Name	Last Name	Role	Function	
Charles	le roi	le roi qui accorde la remission	roi	
Gude	de Noorweghe	le suppliant qui demande la remission	seigneur	
Nicolas	le Seueur	le serviteur domestique de Gude qui commet le meurtre	serviteur	
Frederick	Palatin	le cousin et neveu du roi qui intervient dans l'affaire	duc	

[+ Add Person](#)





Name

Location Type

Bruxelles

ville



Enchuyse

ville



Hongrie

pays



Bohème

pays



Brabant

pays



+ Add Location

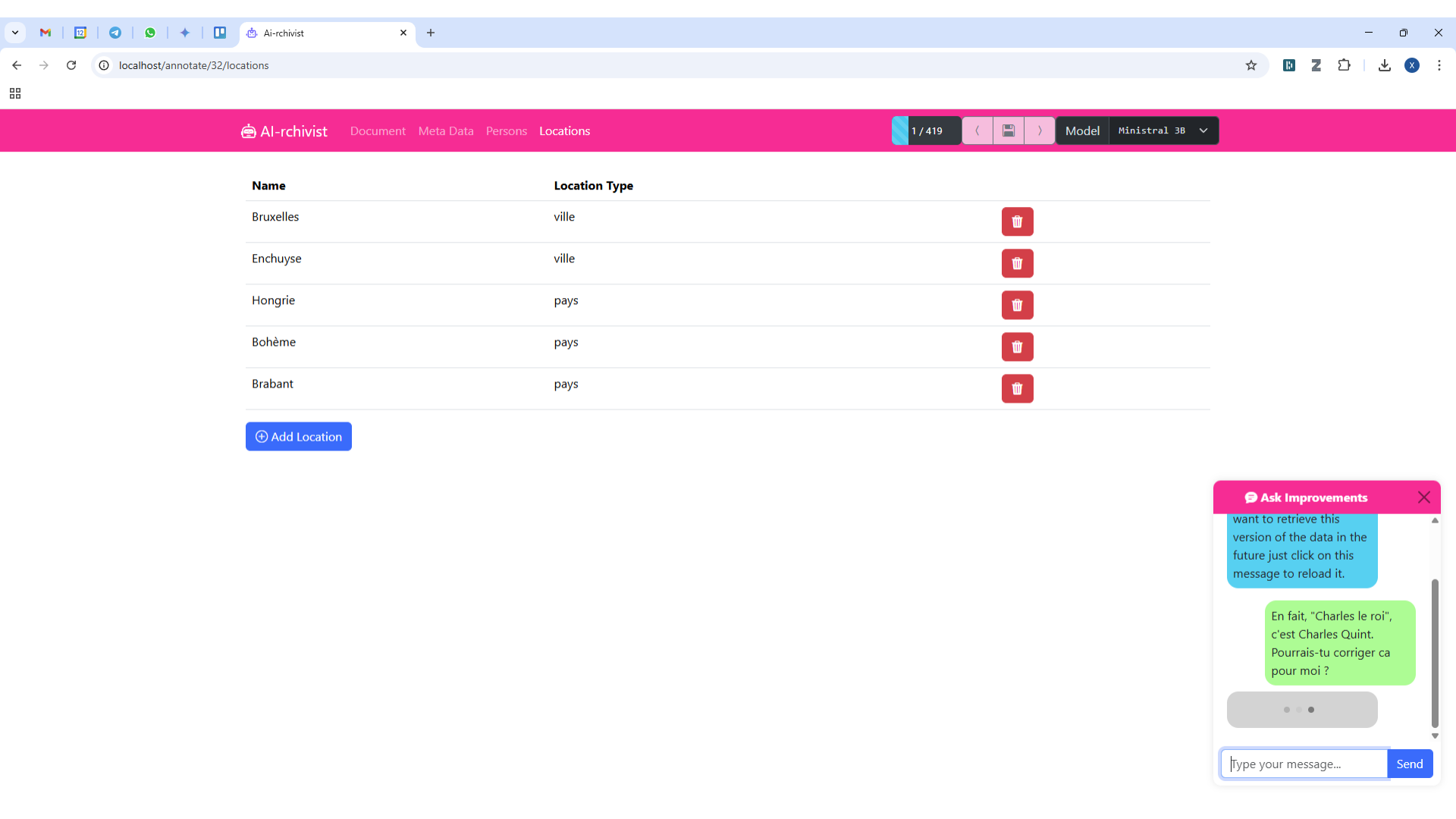
Ask Improvements



A new version of the data has been loaded. If you made some changes and want to retrieve this version of the data in the future just click on this message to reload it.

Type your message...

Send



Name

Location Type

Bruxelles

ville



Enchuyse

ville



Hongrie

pays



Bohème

pays



Brabant

pays



⊕ Add Location

Ask Improvements

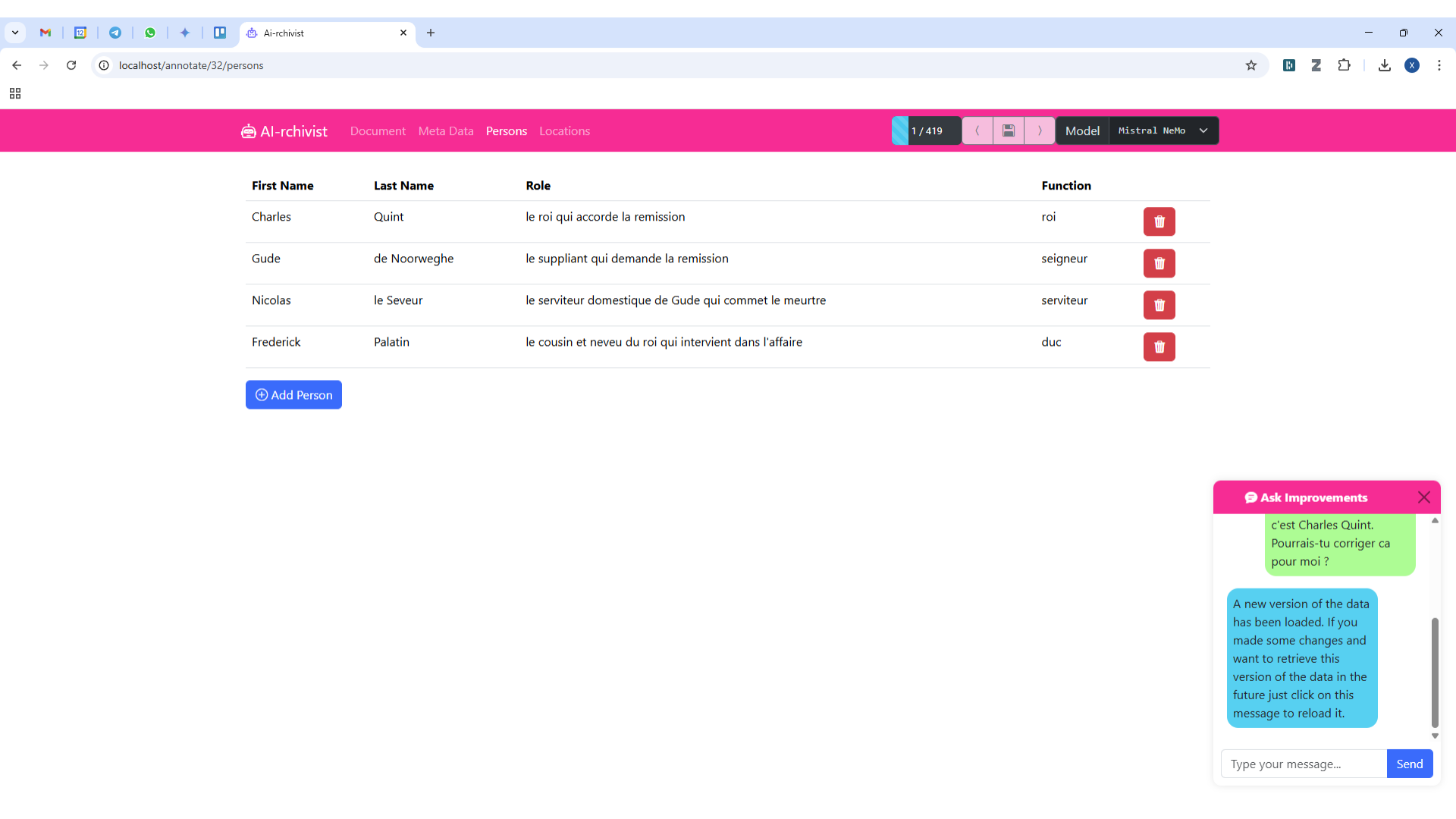


want to retrieve this version of the data in the future just click on this message to reload it.

En fait, "Charles le roi", c'est Charles Quint. Pourrais-tu corriger ça pour moi ?



Send



First Name

Last Name

Role

Function

Charles

Quint

le roi qui accorde la remission

roi



Gude

de Noorweghe

le suppliant qui demande la remission

seigneur



Nicolas

le Seueur

le serviteur domestique de Gude qui commet le meurtre

serviteur



Frederick

Palatin

le cousin et neveu du roi qui intervient dans l'affaire

duc



Add Person

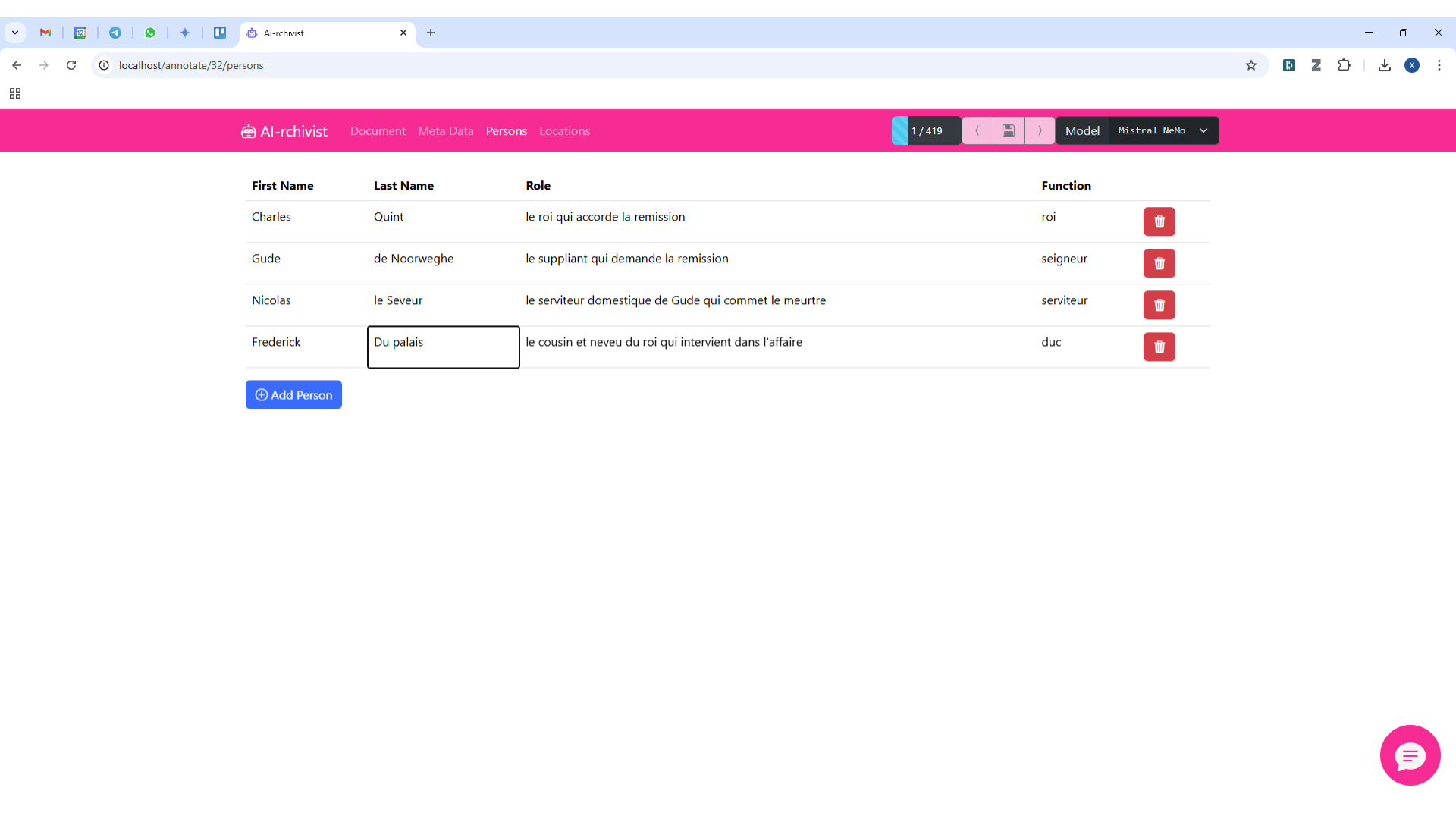
Ask Improvements

c'est Charles Quint.
Pourrais-tu corriger ca
pour moi ?

A new version of the data
has been loaded. If you
made some changes and
want to retrieve this
version of the data in the
future just click on this
message to reload it.

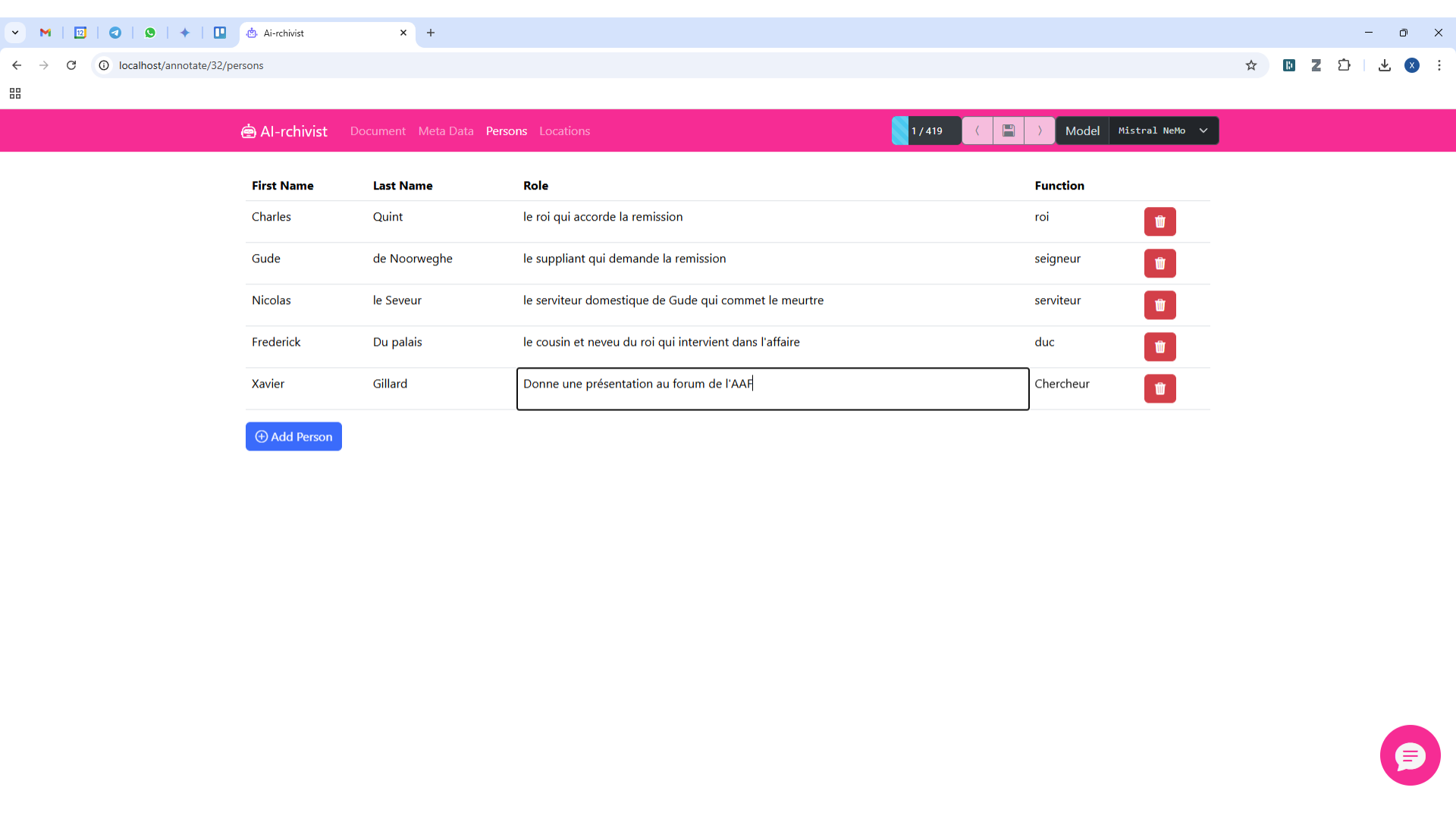
Type your message...

Send



First Name	Last Name	Role	Function	
Charles	Quint	le roi qui accorde la remission	roi	
Gude	de Noorweghe	le suppliant qui demande la remission	seigneur	
Nicolas	le Seueur	le serviteur domestique de Gude qui commet le meurtre	serviteur	
Frederick	Du palais	le cousin et neveu du roi qui intervient dans l'affaire	duc	

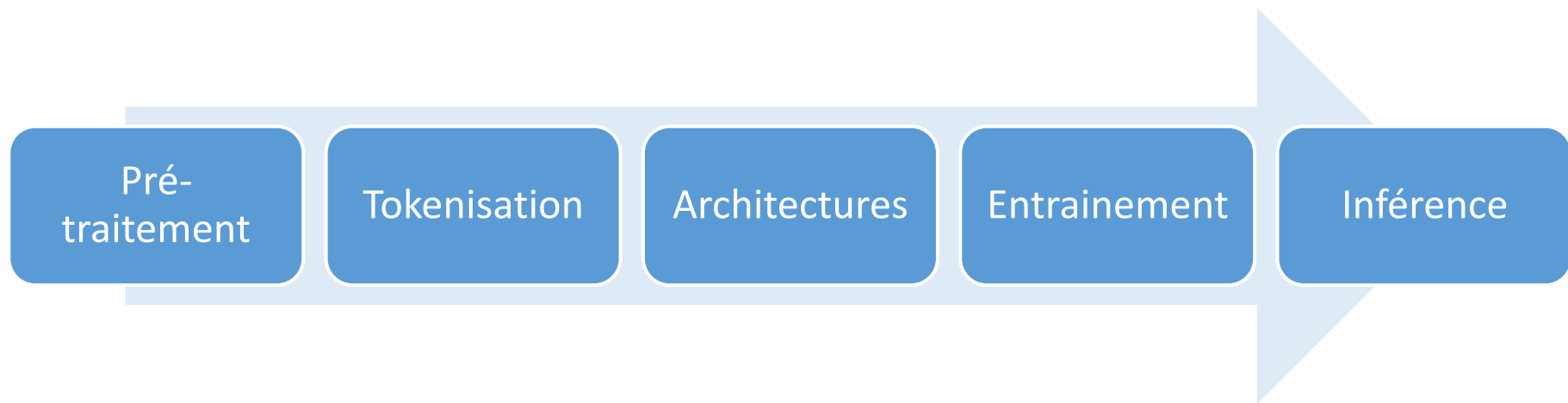
[+ Add Person](#)



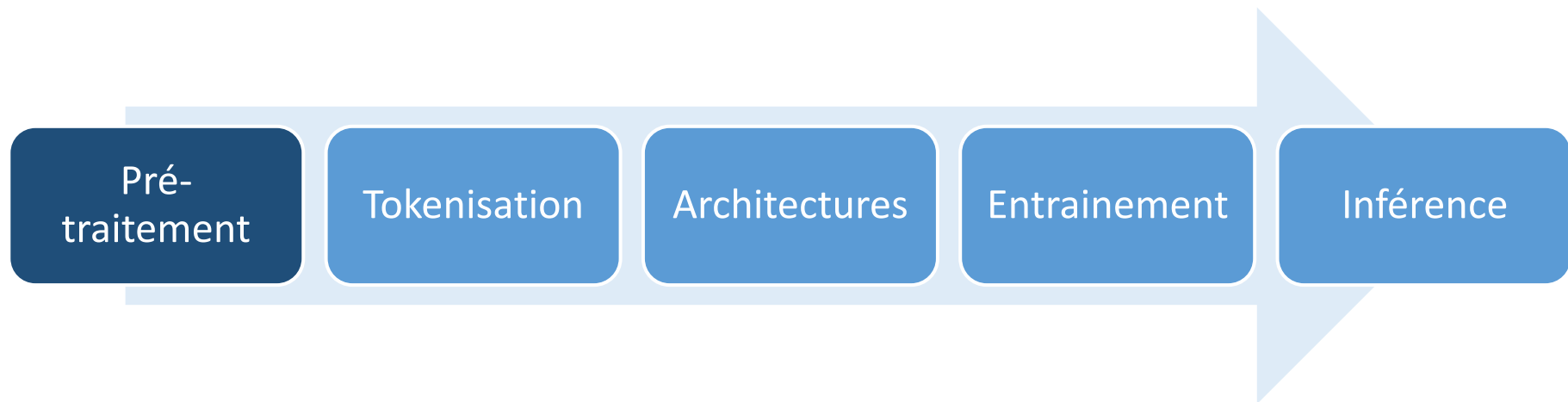
ON DIRAIT DE LA
MAGIE, MAIS
COMMENT ÇA
MARCHE ?



Méthodologie



Méthodologie



farent souffrent et laissent ledit impoye de nre pnté grace
gntan pardon et remission / Selon et par manne / q dit est
plamentement paisiblement / sanz et esle. / Car en ly se
misist en domier ne souffrir et fce fait ind en domie
en royaume ne en biens aucun desloziers en empeschement
au royaume / Mais son royaume ainsi detenu par domier
pour la cause dte / misist en farent misist a plain
et entiere delivran Ensemble / Et bien non rouspignes
santms en sont pour et aces / fcs en empeschites / Car
ainsy nous pleust se estre fait / En testimony de ce
nous avons fait misist nre seel a reb entes / Donné
en nre ville de Brucelles le xxvj^e Jours de may / l'an
de gde mil rmgrens et douze / Et des Roignes de nous
empereurs / d'hanoy de germanie le xxvj^e et de hongrie
le xxvj^e / Et sur le Royleyn avoit escript ainsi / Par l'empereur
et monseigneur l'archiduc / Et leux rousers / Et signé
Dintre /

11 June 1861

[illegible]

W

AutoSave

Off









224_636_489-491_Jan van Bol... • Saved ▾

🔍

Gillard Xavier

GX

—

□

×

FileHomeInsertDrawDesignLayoutReferencesMailingsReview

🛡️

PROTECTED VIEW

Be careful—files from the Internet can contain viruses. Unless you need to view, click the [Protected View](#) link below this message.

Naam: Annick Decoster

Datum: 27/2/2023

ID 224

Archiefreferentie: 510-001-636

Scans: 489-491

Naam suppliant[en]: Jan van Bollenberge

Gecorrigeerd: ja ☐ nee ☐

Transcriptie [f. 488 v°] Kaerle etc. Allen den genen die desen onsen
sien Saluyt Wy hebben ontfangen die oetmoedige supplicatie
van janne van bollebergen poirter en[de] vischcoopere binnen
onse stede van loevenen Inhouden[de] hoe al eyst zoe dat de
voirs. suppl[iant] hem altyts wel en[de] eerlick geregeert ende
gedragen heeft soe wel in zyn voirs. ambacht van vischcoop\$
als onder die coopluden en[de] ande[re] daermede hy te doen[e]
gehadt heeft Click dat hy altyt van goede fame ende

Page 1 of 91631 wordsText Predictions: On

W

AutoSave

Off









242_637_22-25_Jan Slave.doc... • Saved ▾

🔍

Gillard Xavier

GX

—

□

×

FileHomeInsertDrawDesignLayoutReferencesMailingsReviewViewZoteroHelp

🛡️

PROTECTED VIEW

Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View.

Enable Editing

✕

Gratiebrief voor Jan Slave
ARA, *Rekenkamers*, 637, f.20-22v
Getranscribeerd door Gert Gielis

[f.20] Kaerle etc. doen te weten[e] allen lieden iegenwoirdighe
En[de] toecomende dat wy ontfanen hebben die oitmoidige supplica[tie]
van Janne Slave ticheldecke[re] oudt o[m]trent xxii jae[re]n
ingeseten poirte[re] ende tegewoirdelyc gevangen by onsen
amman in ons[er] stad van bruess[el] inhouden[de] hoe dat den
winter lestled[en] is groot geweest en[de] zeer lange geduert
heeft gehadt dat hy suppl[iant] nyet oft nyet vele en heeft con[n]en
gewinnen en[de] binnen den selven winte[re] tot noch toe is
geweest grooten dieren tyt midts welcker aermoede en[de]
gebreke om hem en[de] zyn[en] huysvr[ouwe] den cost te gecrygen[e]
en[de] anderssins by ingeven[e] en[de] temptatien vand[er] vraue
is gebuert dat hy suppliant des saterdaighs voer sint
iansmisse lestleden omtrent drie uren naeder noenen
is geweest in onse hof binnen ons[er] voers. stad van
bruess[el] en[de] com[m]en[de] aende ierste doere staen[de] opden windel-
praet aende beckerie welcke doere hy gesloten

Page 1 of 61634 wordsText Predictions: On

nouveau 1

```

1  kaerle etc. allen den genen die desen onsen brief zullen
2  sien saluyt wy hebben ontfangen die oetmoedige supplicatie
3  van janne van bollebergen poirter ende vischcoopere binnen
4  onse stede van loevenen inhoudende hoe al eyst zoe dat de
5  voirs. suppliant hem altyts wel ende eerlick geregeert ende
6  gedragen heeft soe wel in zyn voirs. ambacht van vischcoop$
7  als onder die coopluden ende andere daermede hy te doene
8  gehadt heeft oick dat hy altyt van goede fame ende
9  name geweest heeft sonder achterhaelt oft verwon$
10 geweest te hebben van enigen stukken weerdich zynd$
11 van pugnacie desen nyet tegenstaende onsen meyer
12 van loeven oft zyn stedehouderen heeft hem onlanx
13 leden gecallengert ende aengesproken voirs scepenen aldaer
14 hem imponerende ende te laste leggende dat hy
15 op sinte barthelemeus dage lestleden int dit jaer
16 xvc twintich ten welcken tyde wy inde voirs.
17 onse stadt van loevenen waren met een wetsteen
18 gewegen ende veroch vercocht zoude hebben eenen
19 hovelinc zekere zalm tot vyf pont toe sonder die
20 gewegen te hebben met geseighden gewichten soe
21 hy sculdich was van doen ende dat hy tot diversen
22 stonden binnen eenen jare herweerts metten voirs.
23 wetsteen zyn zalm uuytgewegen ende vercocht
24 zoude hebben zoe dat opten selven sinte barthelemeus
25 dage henryck de meester diener van onsen
26 voirs. meyer die recht tegen over die vischbanck
27 vanden voirs. suppliant stont zoude totten selven suppliant
28 geseyt hebben ick calengere u dairaf daerop hy
29 suppliant vraechde waeraf daerop de voirs. henryck
30 de meester weder zeyde van dat ghy weght
31 met uwen wetsteen ende wel weet dat ghy
32 nyet wegen en moecht dan met geseighde gewichten
33 waerop de voirs. suppliant antworde dat is myn scul$
34 want den wetsteen meer dan vyf pont weght ende
35 dat hy daer nyet meer gewichts en hadde segge$
36 voirt en weght hy nyet meer dan vyf pont soe
37 geve ick u den besten zalm die ick op myn banck
38 hebbe hoewel de voirs. hendrick seet ende mainteneert
39 dat de voirs. wetsteen metten geseighde gewichten
40 maer bevonden was wegende vier ponden zeere
41 luttel uuytslaende seggende daeromme totten voirs.
42 suppliant ghy en hebt nyet gedaen als een eerlick man
43 dat ghy met desen wetsteen gewegen hebt voer
44 vyf pont ende hy en wecht maer scaers viere pont

```

Normal text file length: 10181 lines: 200 Ln: 1 Col: 57 Pos: 57

*nouveau 1 - Notepad++

Fichier Édition Recherche Affichage Encodage Langage Paramètres Outils Macro Exécution Modules d'extension Documents ?

nouveau 1

```

1  kaerle etc. doen te wetene allen lieden iegenwoirdighe
2  ende toecomende dat wy ontfanen hebben die oitmoedige supplicatie
3  van janne slave ticheldeckere oudt omtrent xxii jaeren
4  ingeseten poirtere ende tegewoirdelyc gevangen by onsen
5  amman in onser stad van brussel inhoudende hoe dat den
6  winter lestleden is groot geweest ende zeer lange geduert
7  heeft gehadt dat hy suppliant nyet oft nyet vele en heeft connen
8  gewinnen ende binnen den selven wintere tot noch toe is
9  geweest grooten dieren tyt midts weloker aermoede ende
10 gebreke om hem ende zynen huysvrouwe den cost te georygene
11 ende anderssins by ingevene ende temptatien vander vraue
12 is gebuert dat hy suppliant des saterdaighs voer sint
13 iansmisse lestleden omtrent drie uren naeder noenen
14 is geweest in onse hof binnen onser voers. stad van
15 brussel ende commende aende ierste doere staende opden windel-
16 graet aende beckerie welcke doere hy gesloten
17 vande met twee nagelen van latysers heeft hy de
18 selve nagelen met zynen werckhamer uuyt gebroeken
19 ende ginc den windelgraet oppe tot boven opten soldere
20 vander cameran vander tapisserien ende vander cameran
21 van onsen juwelen dwelc boven al eenen soldere is ende
22 beu-vindende dat den selven soldere boven de voers. camere
23 vander tapisserien nyet al en was geberdert heeft met
24 eenen zeele hem tusschen die ribben afgelaten inde voers.
25 camere vander tapisserien aldaer hy nyet met allen
26 en heeft genomen maer bevindende die duere van
27 bynnen gesloten met eenen stecke heeft dien vuytgetrocken
28 ende is uuyter cameran gegaen ende soe voerts den voers.
29 windelgraet toetreckende die voers. doere beneden
30 aenden windelgraet die hy te voeren af treckende
31 die voers. nagelen opgedaen hadde maer des anderen
32 daighs nae sint ianssmisse voers. midts der aermoeden
33 ende grooter temptacien voers. is hy suppliant wederomme
34 gecommen omtrent ix oft x uren voere der noenen int
35 voers. hof onse hof ende is duer de voers. nederste
36 doere die hy gevonden heeft toestaende zoe hy die gelaten
37 hadde maer nyet toegesloten wederomme den voers. windelgraete
38 opgegaen tot opten voers. soldere ende van desen soldere is
39 voerts gegaen tot op eenen andere windelgraet ende zoe voerts
40 tot voere die duere vander cameran vanden voers. juwelen
41 aen welcke doere hy tslot achterweerts dede dat
42 bleef sonder sluyten staende maer daer teynden was $$$
43 eenen grendel van binnen vernagelt ende gesloten dat hy
44 daer nyet doer en mochte maer heeft d$er

```

Normal text file length: 10511 lines: 187 Ln: 1 Col: 10 Pos: 10 Windows (CR LF) UTF-8 INS

Pré-Traitement

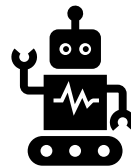
Perd volontairement une partie de
l'information afin de la rendre
exploitable par une machine



Standardisation

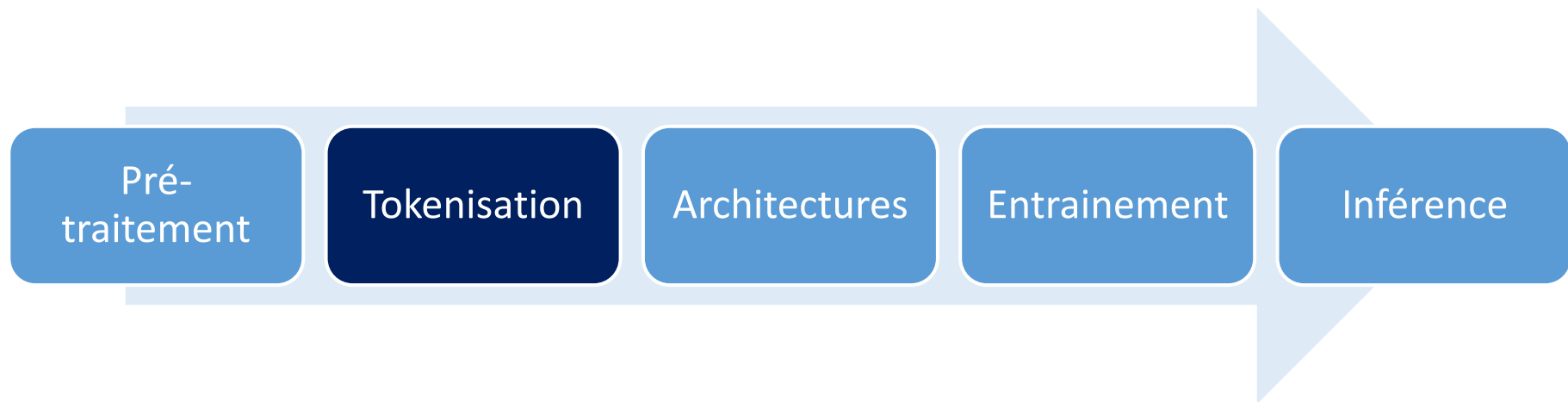


Travail Humain



Exploitable

Méthodologie



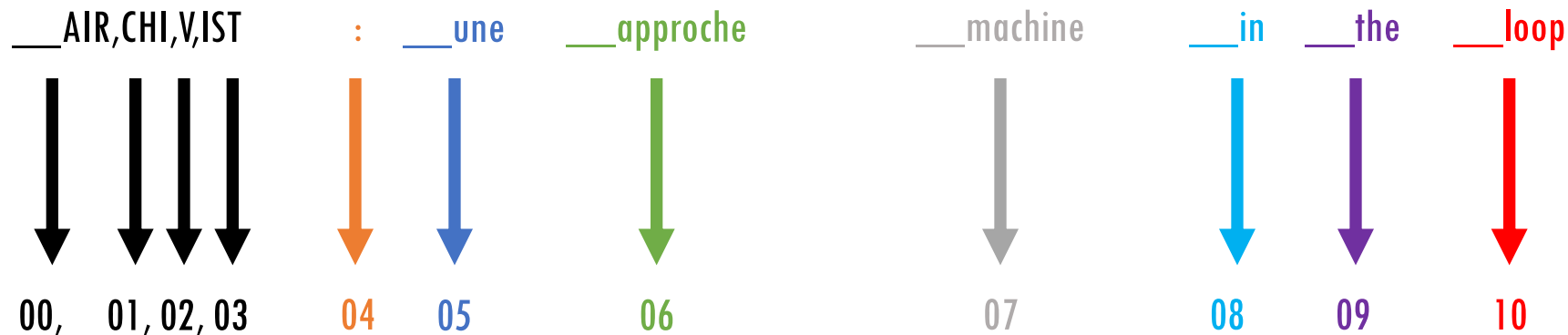
Tokenisation ... ça fait mal ?

Les **machines** ne connaissent pas de mots,
juste **un ensemble fini de nombres**. Il faut
donc utiliser un **vocabulaire fini** pour
transformer le texte en une suite de nombres

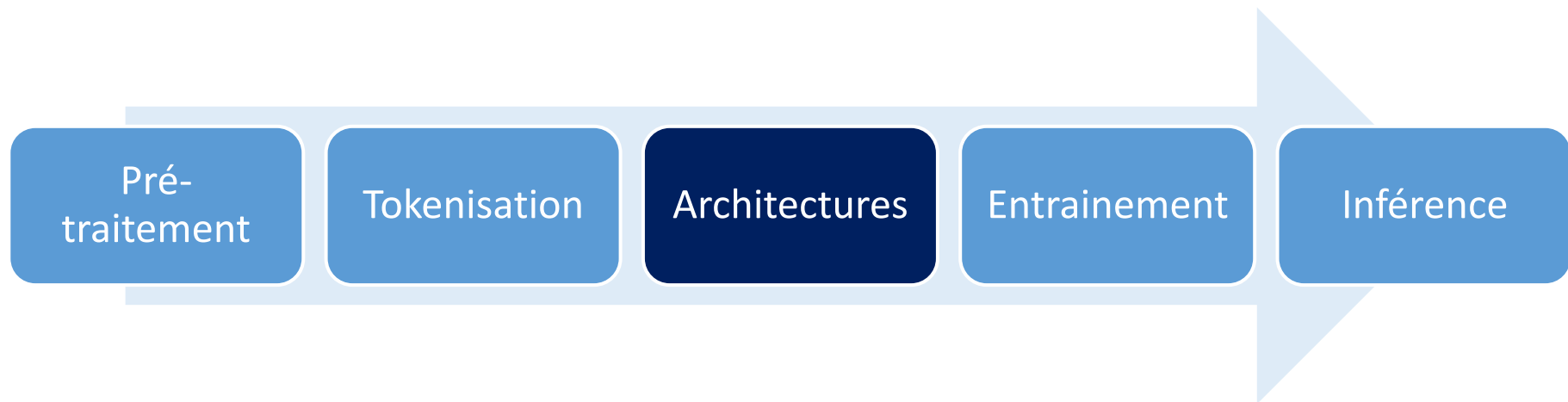
Exemple AIRCHIVIST: une approche machine in the loop

Token	Identifiant
__AIR	00
CHI	01
V	02
IST	03
:	04
__une	05
__approche	06
__machine	07
__in	08
__the	09
__loop	10

AIRCHIVIST : une approche machine in the loop



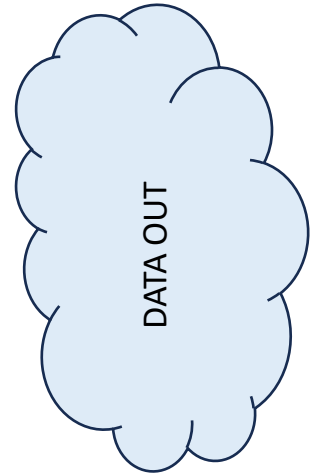
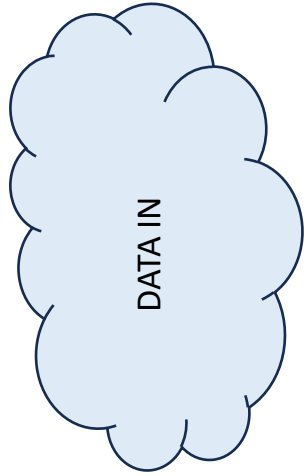
Méthodologie





ARCHITECTURES

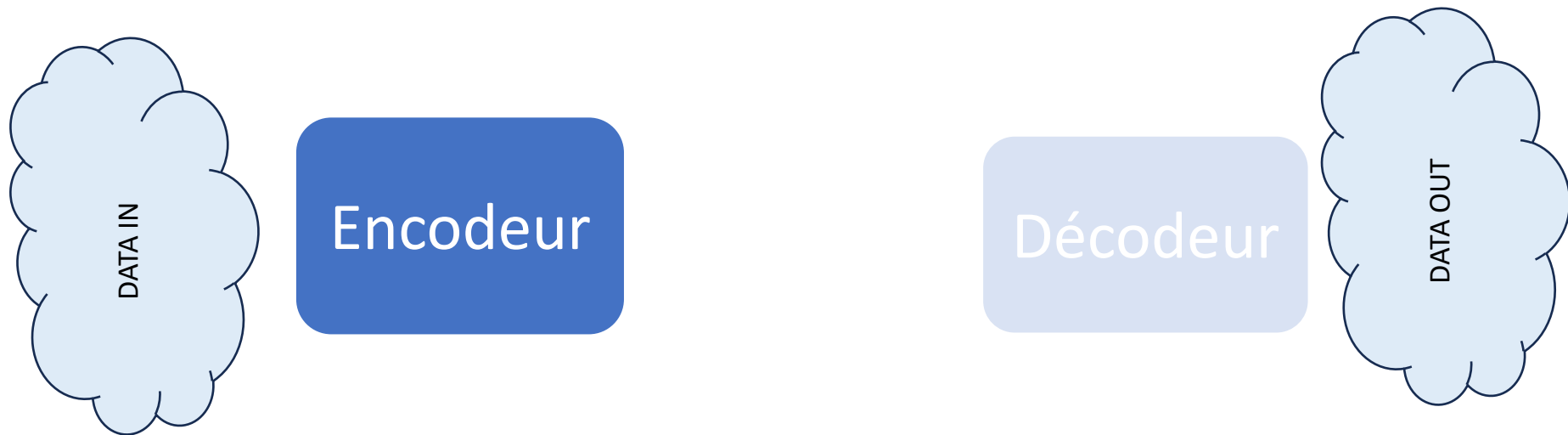
Architecture générale



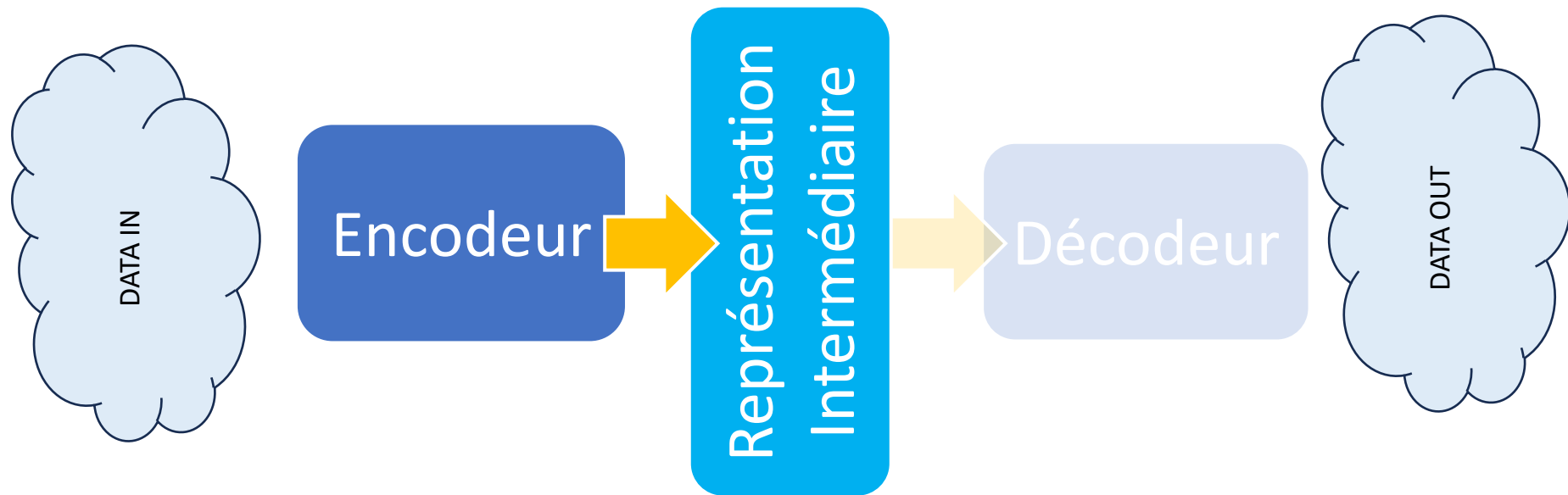
Architecture générale



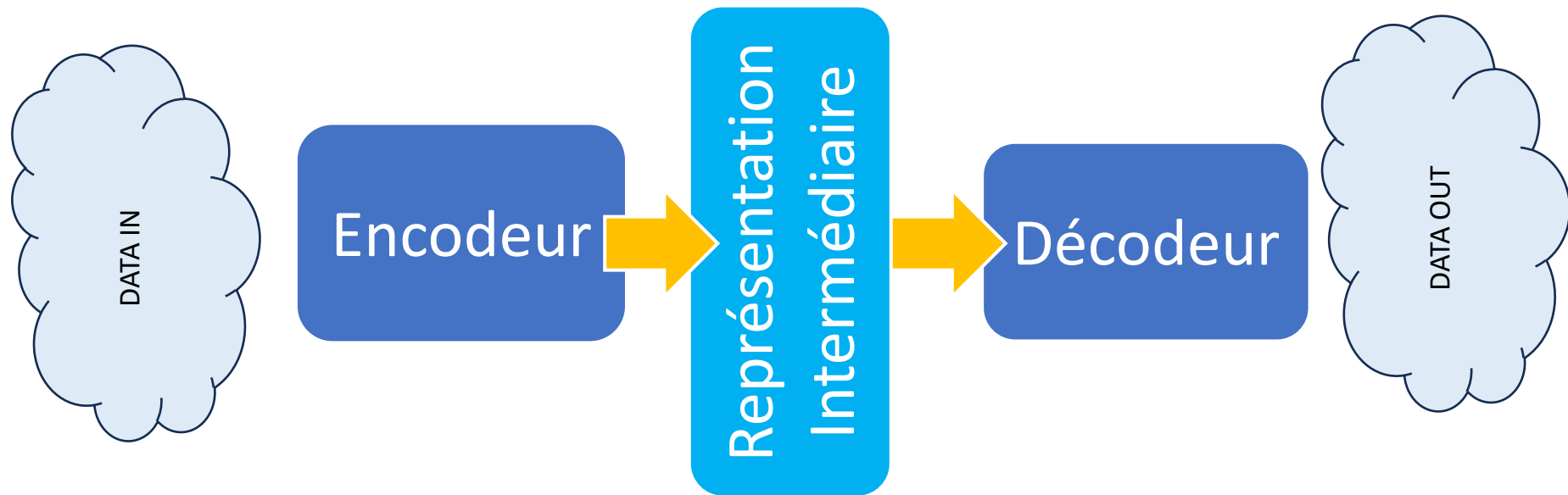
Architecture générale



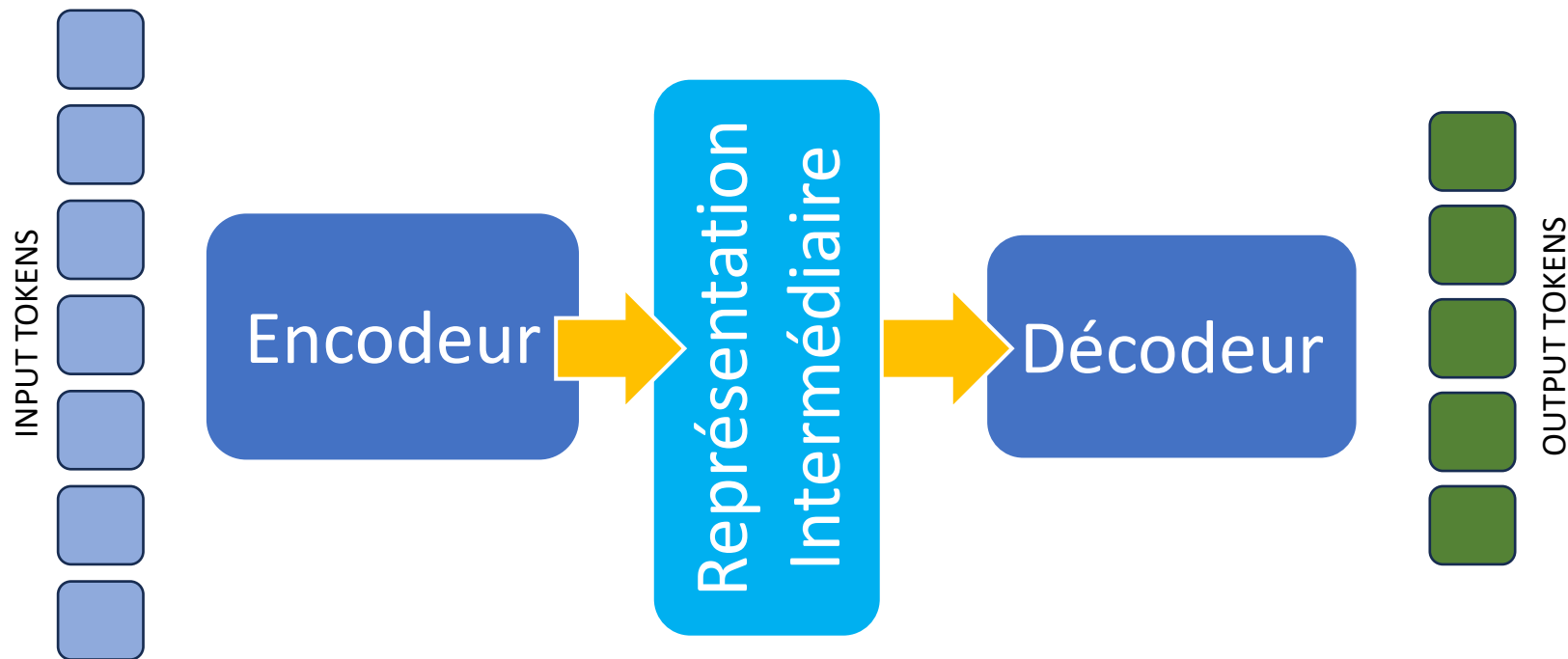
Architecture générale



Architecture générale



Architecture générale





C'EN EST TROP!
... même pour de super serveurs!



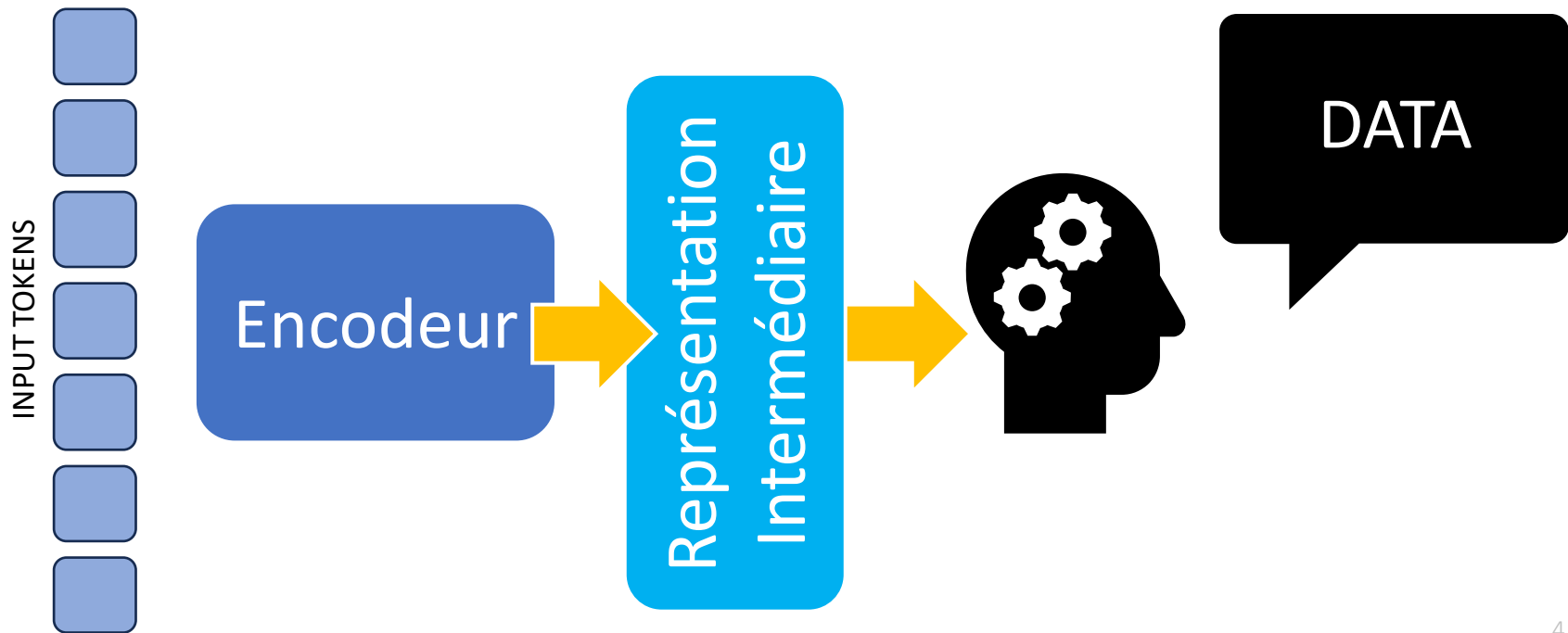
3 facons de
couper





?? BERT ??

1. Encodeur uniquement (BERT, RoBERTa)

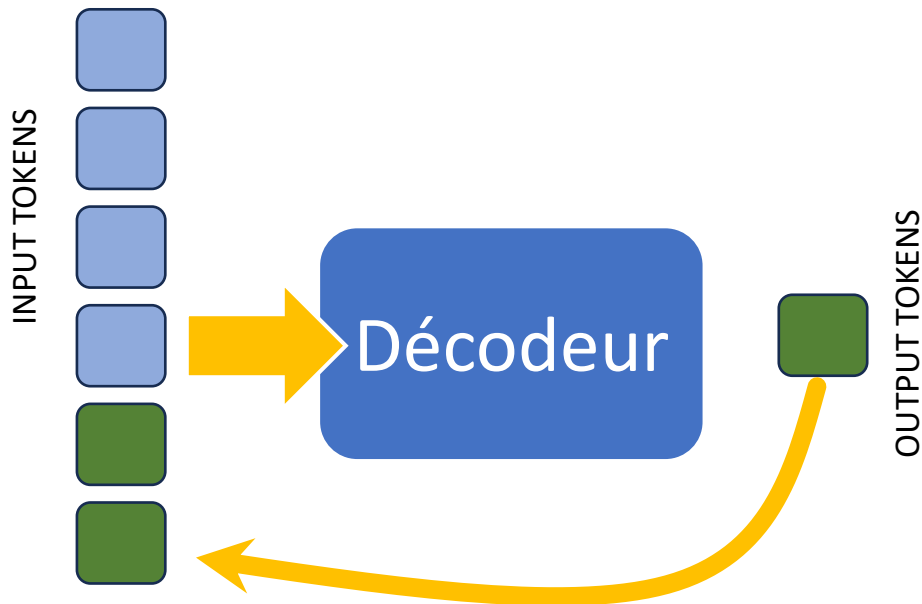




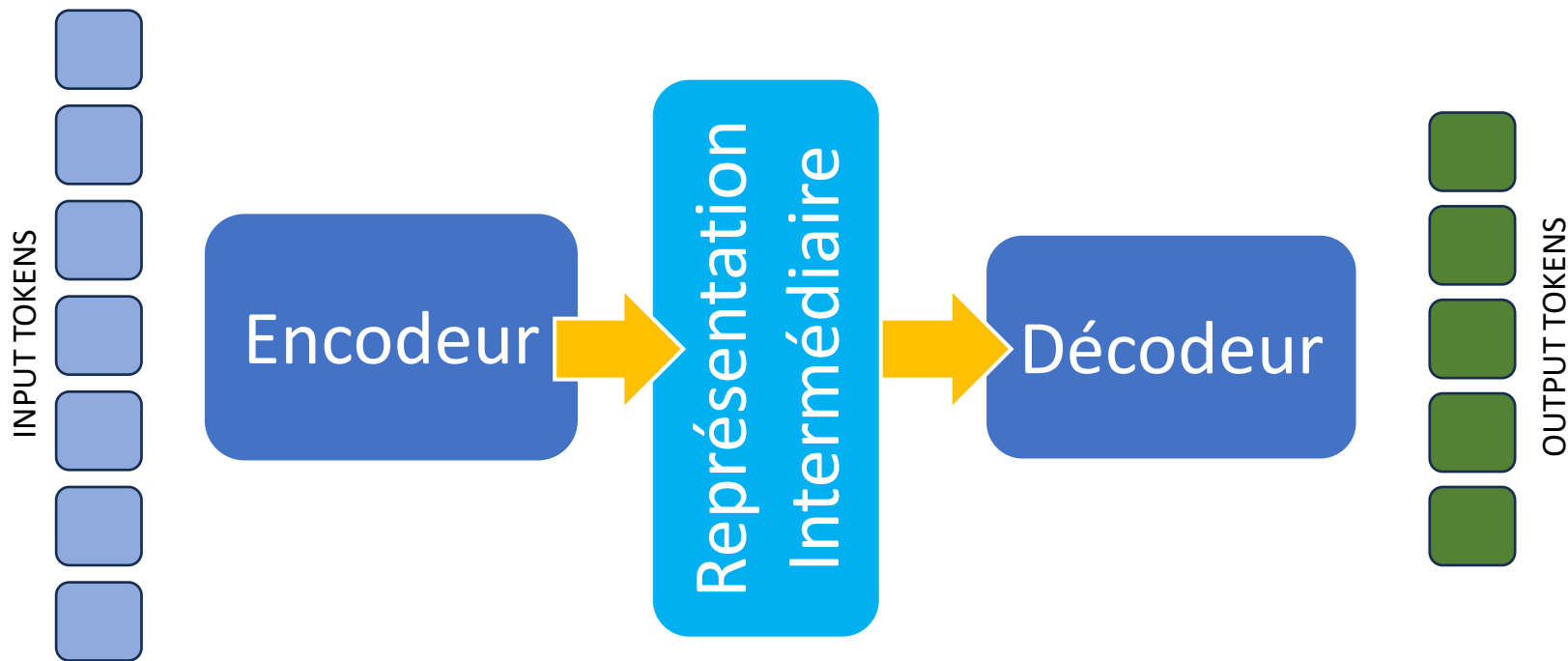
Décodeur uniquement



2. Décodeur uniquement (GPT, MISTRAL, ...)



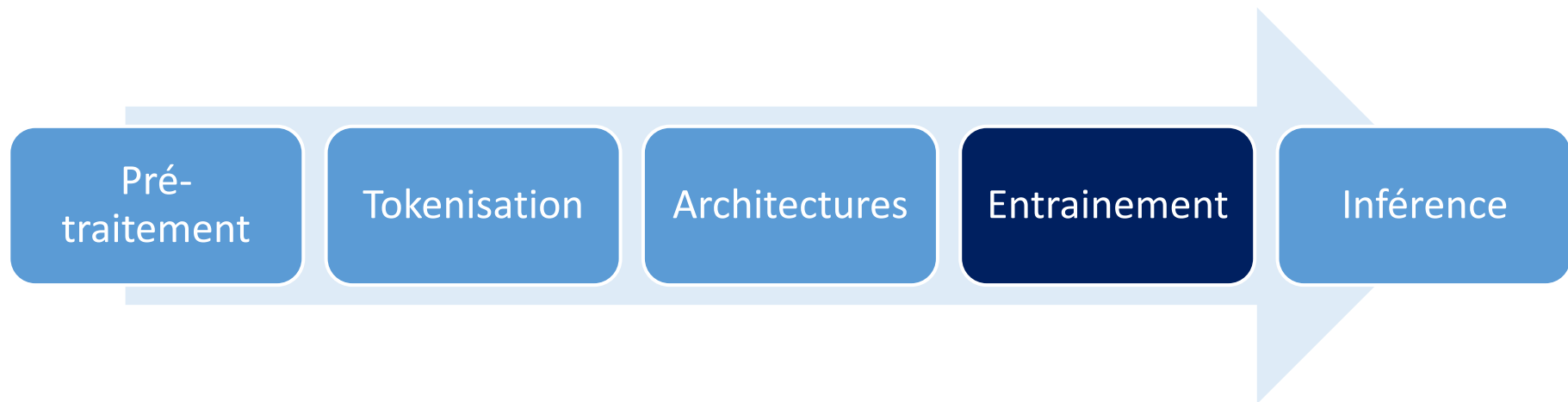
3. On ne coupe pas (transformer complet)



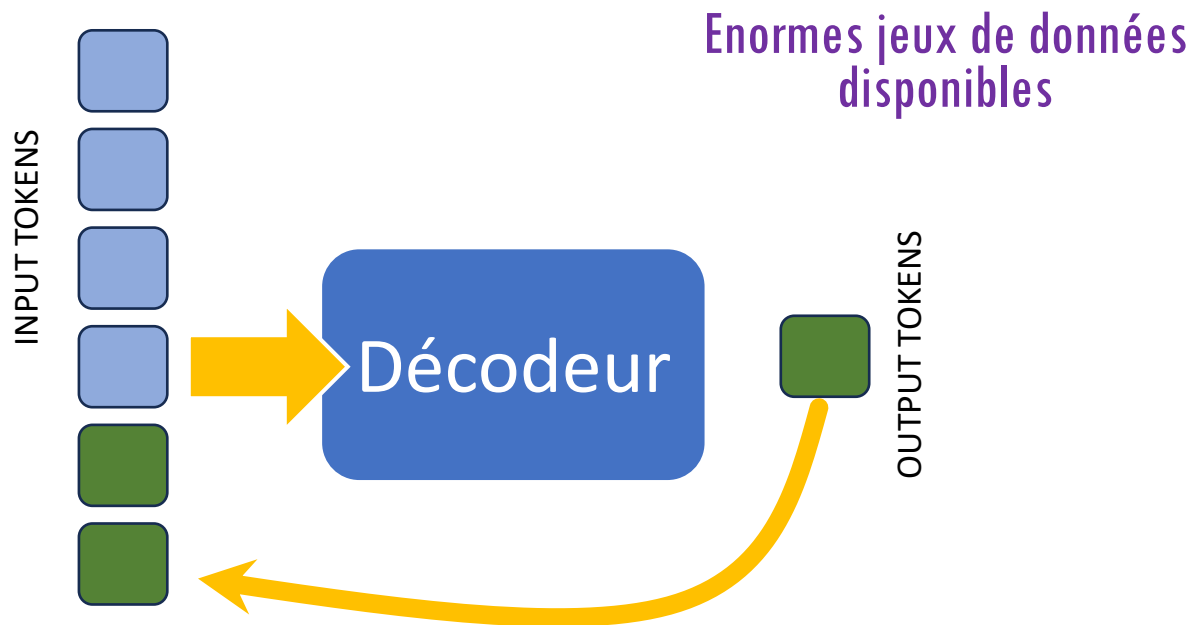
A low-angle, backlit photograph of a crowd of people at a party. The scene is filled with numerous balloons in various colors (pink, blue, yellow, green). In the foreground, several people are silhouetted against a bright, hazy light source. One person in the center-left is making a heart shape with their hands. Other people are reaching up, some holding balloons. The overall atmosphere is festive and celebratory.

D'où vient la popularité des modèles décodeur ?

Méthodologie



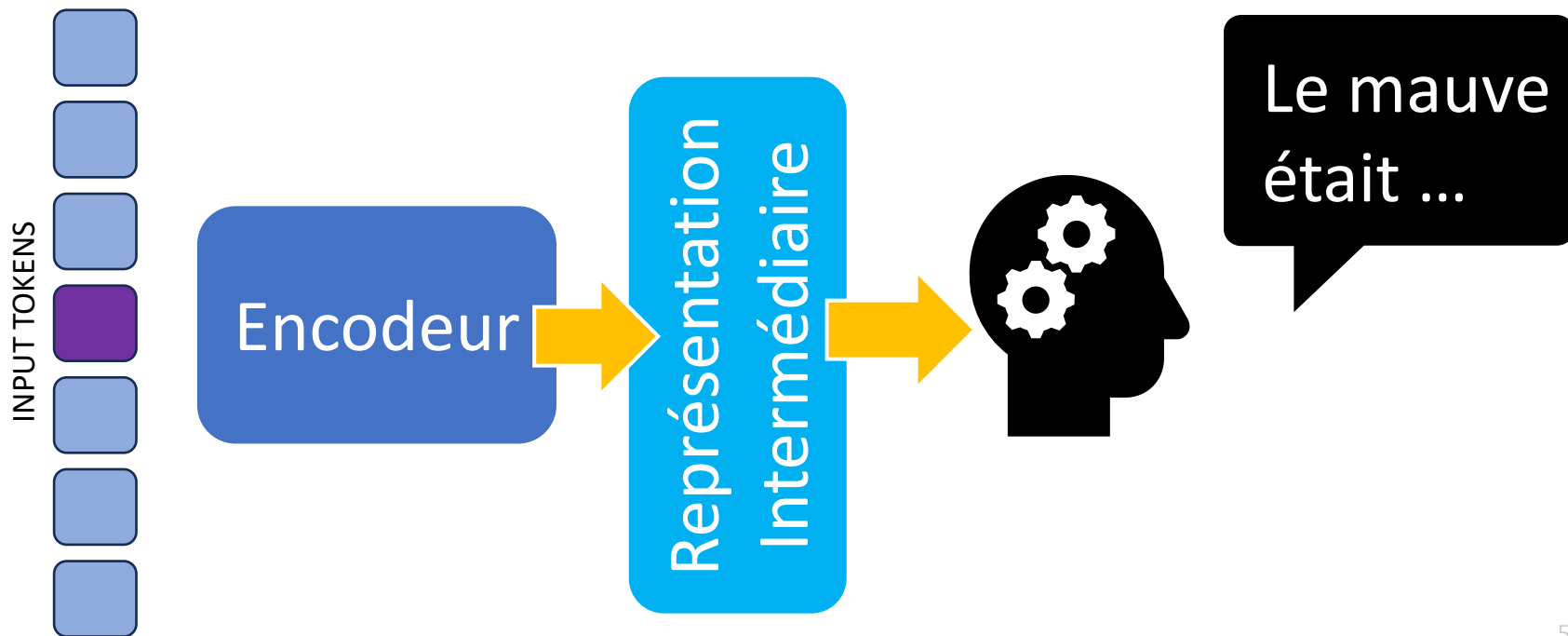
Entraînement auto supervisé



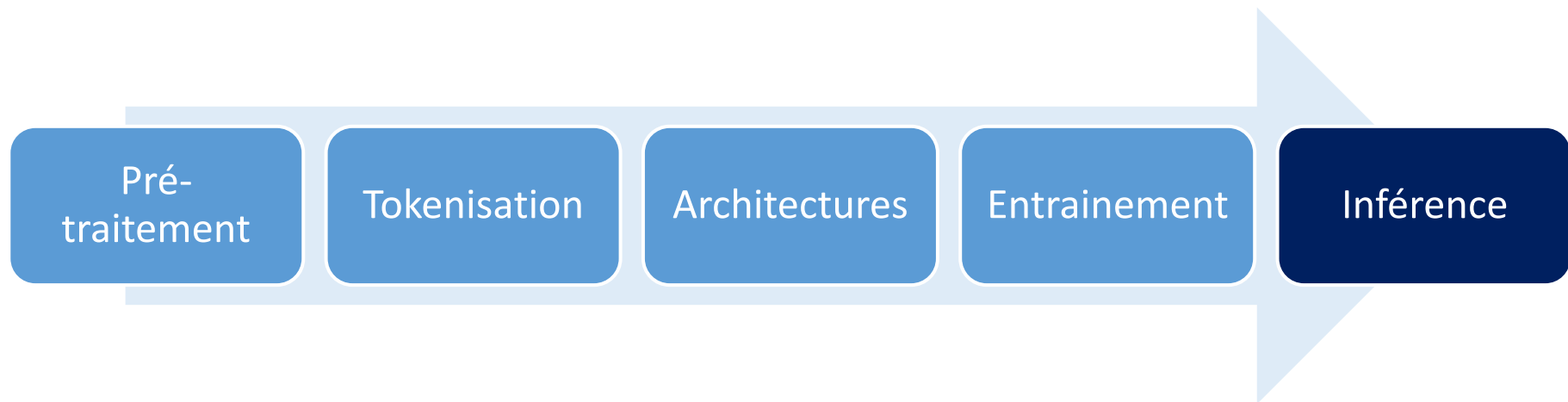
ET POUR LES MODELES
UNIQUEMENT ENCODEUR
IL N'Y A RIEN A FAIRE?



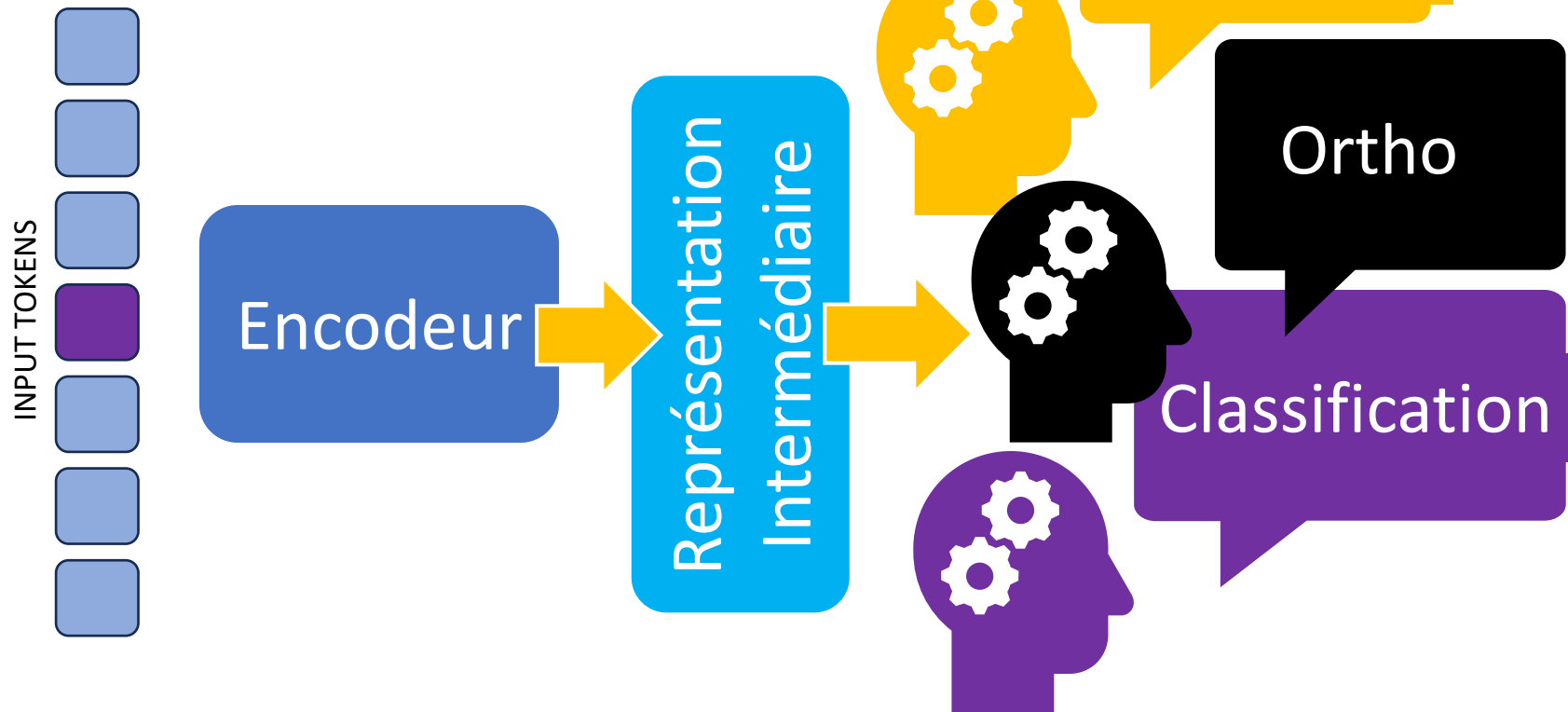
MLM PRETRAINING



Méthodologie

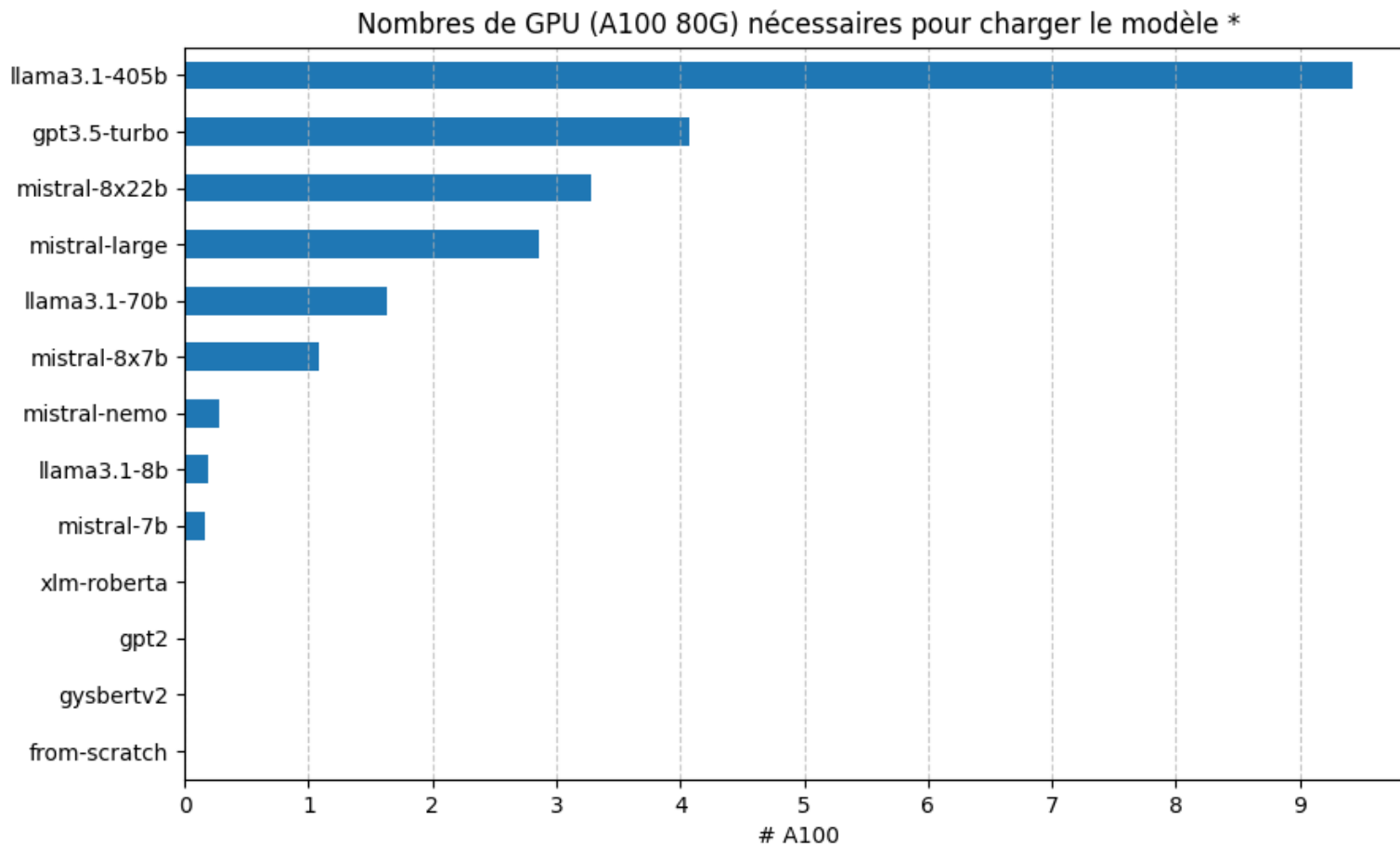


PRETRAINING – Pourquoi ?



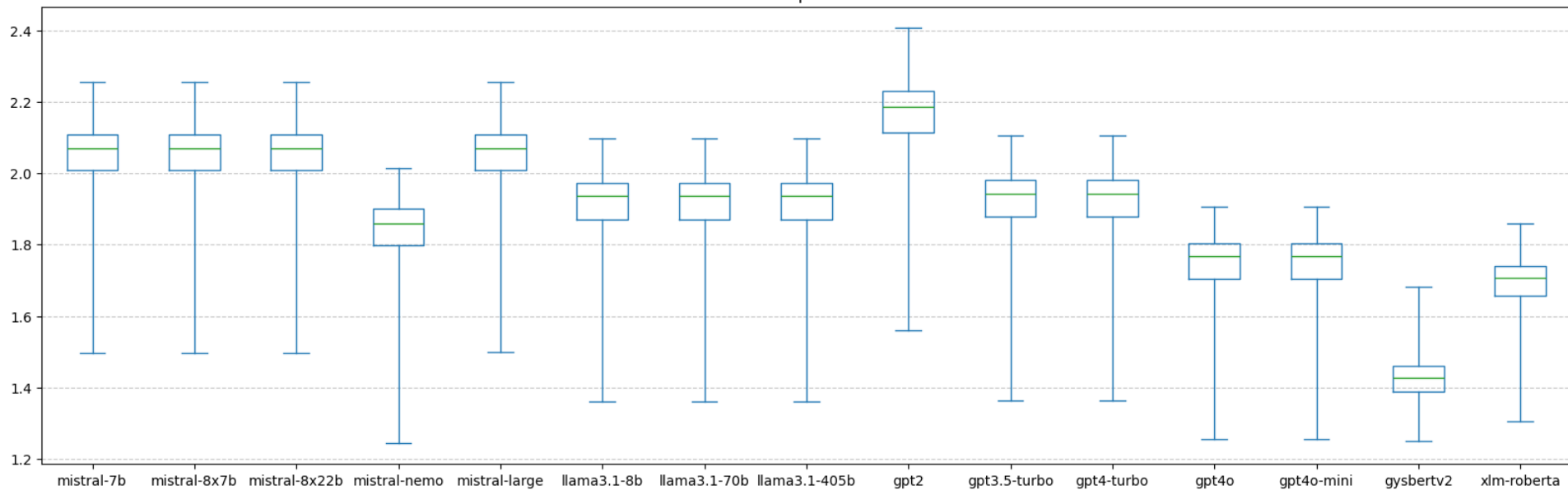
The image features the word "FAIL" spelled out in large, three-dimensional wooden block letters. The letters are light-colored wood with visible grain. The letter 'A' has a circular hole in its center. The letters are set against a dark, textured background. A semi-transparent dark horizontal band runs across the middle of the image, containing the word "LIMITATIONS" in white, uppercase, sans-serif font.

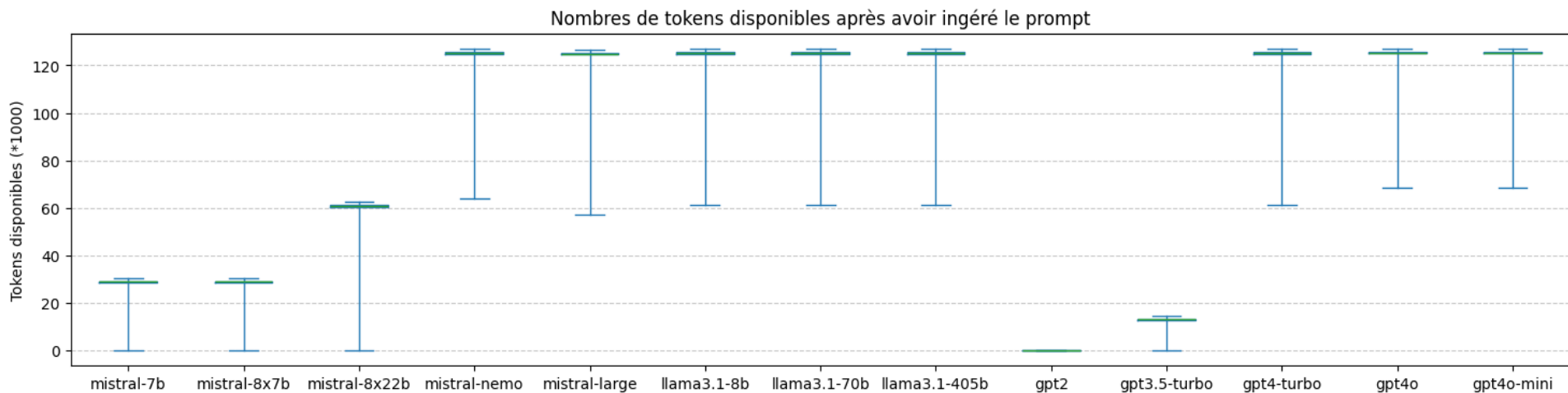
LIMITATIONS

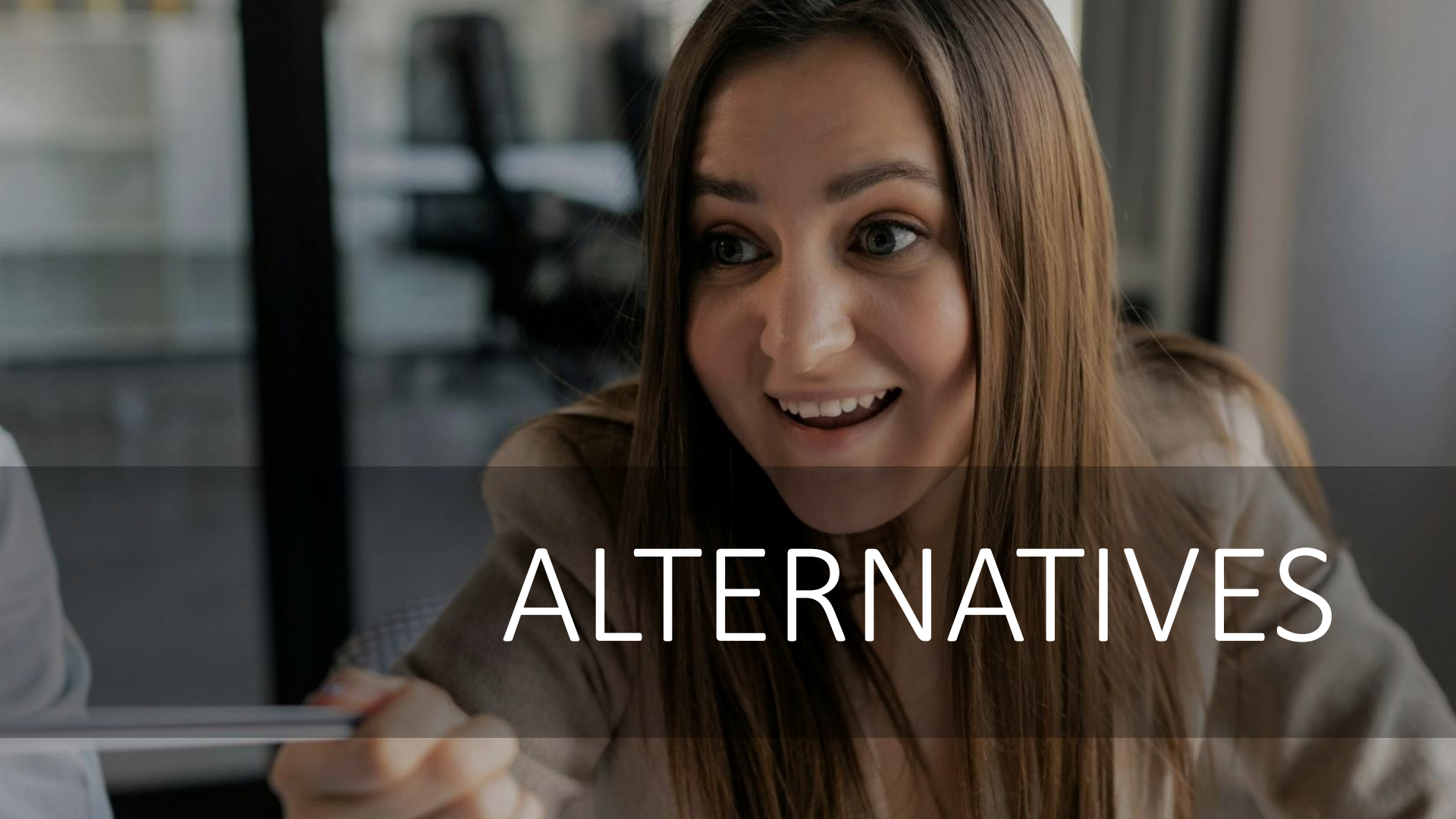


* En supposant que les poids soient représentés sur 16 bits

Ratio du nombre de tokens comparé à celui d'un tokenizer "idéal"

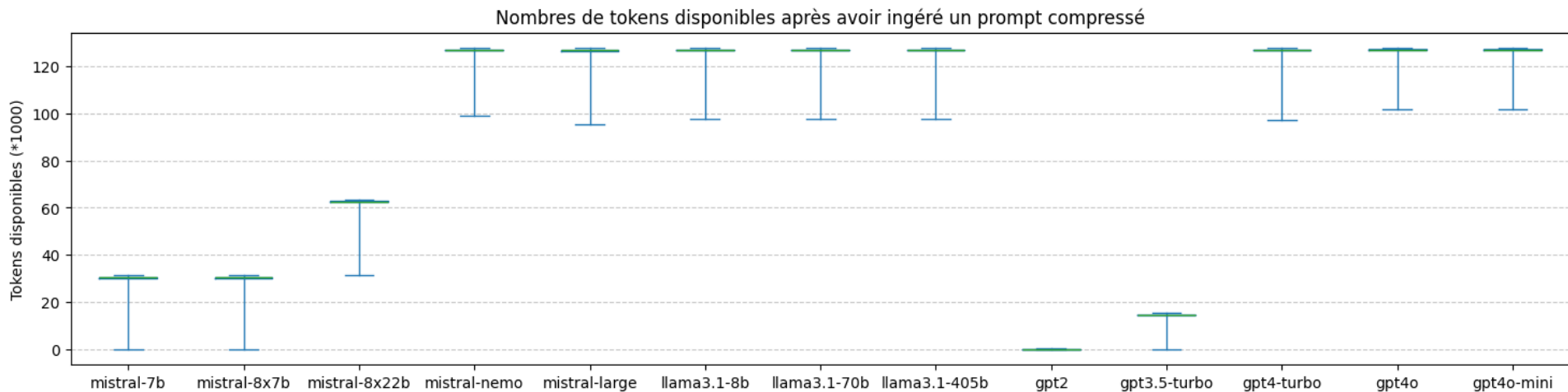




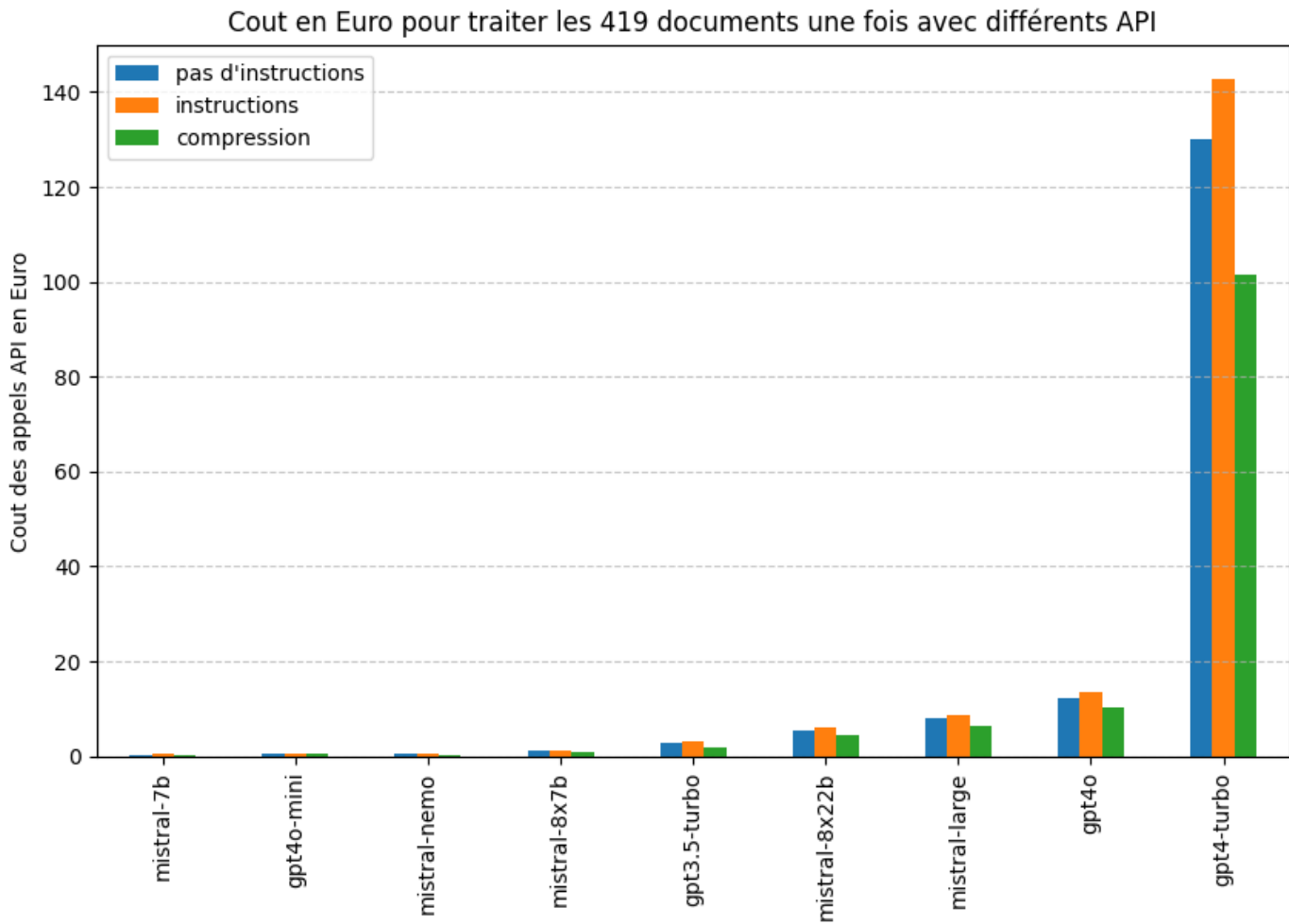
A close-up photograph of a young woman with long, straight brown hair and a bright smile. She is looking slightly off-camera to her left. She is wearing a light-colored, possibly beige or cream, button-down shirt. The background is a blurred office environment with shelves and papers. A semi-transparent dark horizontal band runs across the middle of the image, and the word "ALTERNATIVES" is written in large, white, sans-serif capital letters across this band.

ALTERNATIVES

Compression de prompt



Appels API





?? BERT ??



Expérience en cours

