

L'assistant juridique **propulsé** par intelligence artificielle

Marco LOPEZ Clément SAULNIER Olivier LE PESTIPON



LE PROBLÈME

Difficile pour un public non averti de connaître la législation en vigueur



Code du travail ~ 3000 pages.



DES RECHERCHES FASTIDIEUSES ET UN MANQUE DE TEMPS

Des dizaines de livres, titres, chapitres, sections, articles.



VOCABULAIRE SPÉCIFIQUE, DIFFICULTÉS À COMPRENDRE

Le jargon juridique n'est pas toujours clair.



LA SOLUTION

HarvAI: Un chatbot intelligent qui répond directement à toutes vos questions juridiques.

RECHERCHE LES ARTICLES
DE LOI PERTINENTS

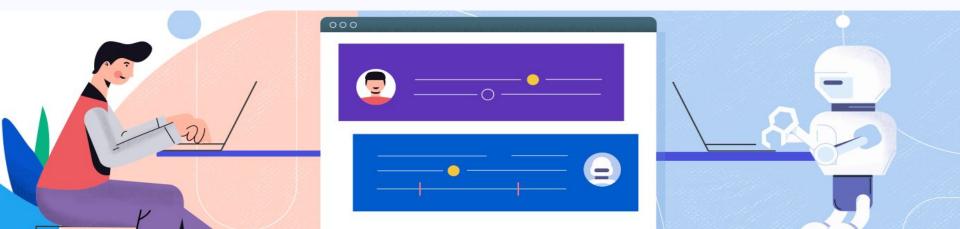


FORMULE UNE RÉPONSE EN SE BASANT SUR LES ARTICLES



PREMIÈRE VERSION SUR LE CODE DE LA ROUTE





HarvAl est organisé en quatre blocs.



DATA COLLECTION

Récupération des textes de lois sur Légifrance.



PREPROCESSING DES DONNÉES

Formatage des données pour les modèles.

QUESTION DE L'UTILISATEUR





MODÈLE « RETRIEVER »

Recherches des articles de loi pertinents.



MODÈLE « READER »

Extraction de la réponse à la question.



INTERFACE DE CHAT STREAMLIT

Affichage de la réponse.



LES DONNÉES

Le preprocessing des textes est une étape indispensable.

ARTICLES

DATAFRAME

Code de la route



~1 200 articles de lois

CODE

PDF - Code de la route



~ 350 PAGES

PREPROCESSING









USER_INPUT

Quelle est la peine pour conduite en état d'ivresse?

OUTPUT

quel peine conduite état ivresse

HarvAl est organisé en quatre blocs.



DATA COLLECTION

Récupération des textes de lois sur Légifrance.



PREPROCESSING DES DONNÉES

Formatage des données pour les modèles.

QUESTION DE L'UTILISATEUR





MODÈLE « RETRIEVER »

Recherches des articles de loi pertinents.



MODÈLE « READER »

Extraction de la réponse à la question.



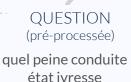
INTERFACE DE CHAT STREAMLIT

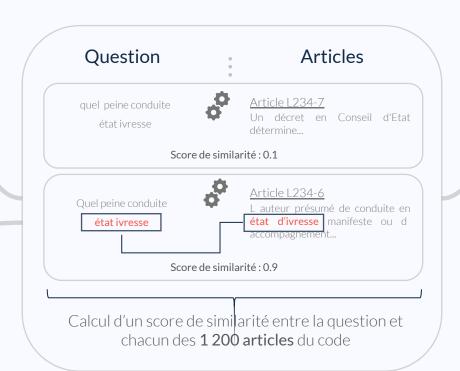
Affichage de la réponse.



L'algorithme qui retrouve les articles pertinents.







OUTPUT

Les 4 articles ayant le meilleur score

Article L234-6

L auteur présumé de conduite en état d'ivresse manifeste ou d accompagnement...

Article L234-1

Même en l'absence de tout signe d'ivresse manifeste, le fait de conduire...

HarvAl est organisé en quatre blocs.



DATA COLLECTION

Récupération des textes de lois sur Légifrance.



PREPROCESSING DES DONNÉES

Formatage des données pour les modèles.

QUESTION DE L'UTILISATEUR





MODÈLE « RETRIEVER »

Recherches des articles de loi pertinents.



MODÈLE « READER »

Extraction de la réponse à la question.



INTERFACE DE CHAT STREAMLIT

Affichage de la réponse.

LE READER 😕

Le modèle qui répond à une question en utilisant un texte court

CONTEXTE

4 articles du Code de la route retournés par le retriever

Article L234-6

Article L234-1

Article L213-3

Article L224-8

QUESTION

Choisi par l'utilisateur via l'interface:

« Quelle est la peine pour conduite en état d'ivresse ? »

MODÈLE CAMEMBERT

Modèle de question réponse basé sur CamemBERT pré entraîné et amélioré en utilisant 3 datasets en français.

Article L234-6:[...]

Article L234-1:

[...] l' air expiré égale ou supérieure à 0,40 milligramme par litre est puni de deux ans d'emprisonnement et de 4 500 euros d'amende II. Le fait de conduire un véhicule en état d'ivresse manifeste est puni des mêmes peines [...]

Article L213-3: [...] Article L224-8: [...]

etalab

RÉPONSE

« deux ans d'emprisonnement et de 4 500 euros d amende. »

HarvAl est organisé en quatre blocs.



DATA COLLECTION

Récupération des textes de lois sur Légifrance.



PREPROCESSING DES DONNÉES

Formatage des données pour les modèles.

QUESTION DE L'UTILISATEUR





MODÈLE « RETRIEVER »

Recherches des articles de loi pertinents.



MODÈLE « READER »

Extraction de la réponse à la question.



NOS RECHERCHES

Comment rendre HarvAI plus performant?



Modèles de retrievers : Dense vs. Sparse

Deux types de modèles très différents, avec et sans réseaux de neurones.



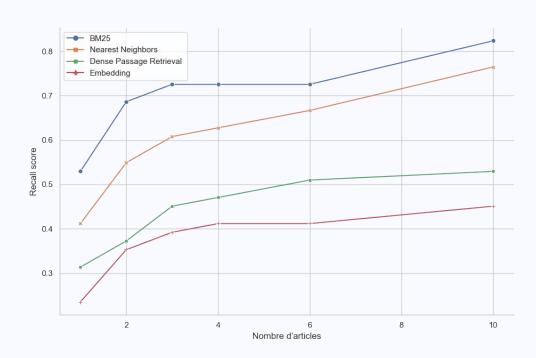
Méthode de scoring des retrievers

Création d'un dataset de questions/réponses sur mesure et d'une fonction d'évaluation.



Evaluation de quatre modèles différents sur notre dataset

Les modèles "*sparsc*" sont plus performants sur nos données et avec notre approche.



NEXT STEPS

Quels axes d'amélioration?

DATA

Utiliser d'autres corpus (code pénal, code civil, jurisprudence...).

Entraîner les modèles sur un dataset labellisé.



PREPROCESSING

Ajouter un preprocessing supplémentaire pour transformer le vocabulaire juridique en langage plus courant.



"Écouteurs" vs.

"Tout dispositif susceptible d'émettre du son"

USER INTERFACE

Création d'une application dédiée.





L'assistant juridique **propulsé** par intelligence artificielle

Marco LOPEZ Clément SAULNIER Olivier LE PESTIPON



