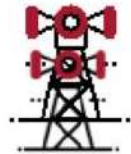




COUR DE CASSATION

Open data et Pseudonymisation des décisions de Justice

OBJECTIFS DE L'OPEN DATA



Garantir la publicité des décisions de justice



Assurer la transparence de la justice



Renforcer la confiance dans la justice



Assurer une meilleure **sécurité juridique**

ENJEUX DE L'OPEN DATA



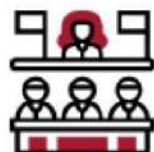
Concilier la **mise à disposition de décisions** lisibles avec la **protection des données personnelles**.



Rendre les **informations accessibles** et créer des **outils méthodologiques**.



Open data et analyses des pratiques professionnelles.



Modification de **l'office du juge**?

ETAPES DE L'OPEN DATA



Cadre légal

L'article L.111-13 du code de l'organisation judiciaire :

Sous réserve des dispositions particulières qui régissent l'accès aux décisions de justice et leur publicité, les décisions rendues par les juridictions judiciaires sont mises à la disposition du public à titre gratuit sous forme électronique.

Les nom et prénoms des personnes physiques mentionnées dans la décision, lorsqu'elles sont parties ou tiers, sont occultés préalablement à la mise à la disposition du public. Lorsque sa divulgation est de nature à porter atteinte à la sécurité ou au respect de la vie privée de ces personnes ou de leur entourage, est également occulté tout élément permettant d'identifier les parties, les tiers, les magistrats et les membres du greffe.

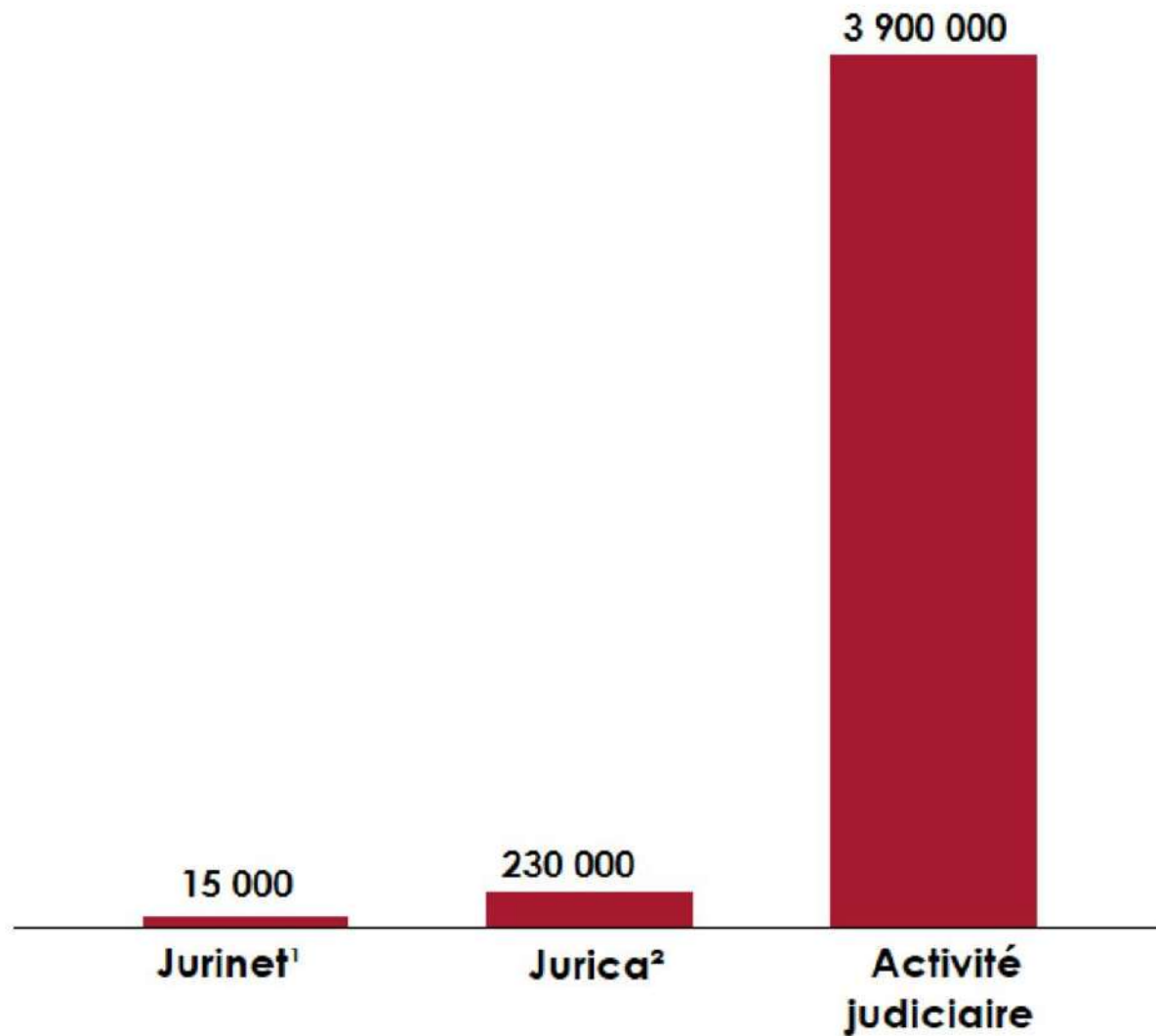
Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées. La violation de cette interdiction est punie des peines prévues aux articles 226-18, 226-24 et 226-31 du code pénal, sans préjudice des mesures et sanctions prévues par la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

Mise en œuvre: cadre réglementaire

Décret n°2020-797 du 29 juin 2020:

- La Cour de cassation est responsable de la mise à la disposition du public, sous forme électronique, des décisions de justice rendues par les juridictions judiciaires,
- Une occultation « socle » des noms et des prénoms des personnes physiques, et une occultation « renforcée », décidée, dans le cas où, malgré l'occultation des nom et prénoms, la mise à disposition de la décision est de nature à porter atteinte à la sécurité ou au respect de la vie privée des personnes physiques mentionnées au jugement ou de leur entourage,
- La décision d'occulter tout autre élément d'identification est prise par le président de la formation de jugement ou le magistrat ayant rendu la décision en cause lorsque l'occultation concerne une partie ou un tiers.
- Lorsque l'occultation concerne un magistrat ou un membre du greffe, la décision est prise par le président de la juridiction concernée.
- Toute personne intéressée peut introduire, à tout moment, devant un magistrat de la Cour de cassation désigné par le premier président, une demande d'occultation ou de levée d'occultation des éléments d'identification visés dans le cadre de l'occultation « renforcée ». Cette décision peut faire l'objet d'un recours devant le premier président de la Cour de cassation dans les deux mois suivant sa notification.

DEFI DE VOLUME: FLUX MOYEN ANNUEL DECISIONS



¹ Décisions de la Cour de cassation

² Décisions des cours d'appel

DÉFI : PROTECTION DONNÉES PERSONNELLES



Les décisions de justice doivent être **pseudonymisées avant diffusion.**



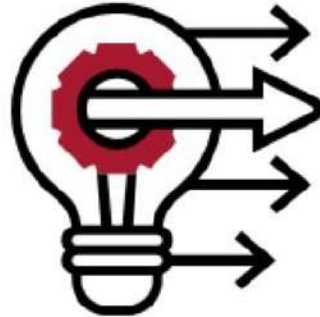
Deux niveaux de pseudonymisation : une occultation « socle » pour prénoms et nom personnes physiques et occultations complémentaires décidées par le magistrat (atteinte vie privée et sécurité)



Faciliter la mise en œuvre : **élaboration de doctrines d'harmonisation** par la constitution de groupes de travail internes (Cour de cassation et cours d'appel) – conformément à la **délibération de la CNIL du 6 février 2020**

UN CALENDRIER PROGRESSIF





Besoin **d'internaliser les**
compétences

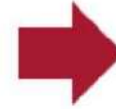
LES PROGRAMMES D'ENTREPRENEUR D'INTERET GENERAL

Entrepreneur d'intérêt général (eig3)

Entrepreneur d'intérêt général (eig4)



LES PROGRAMMES ENTREPRENEUR D'INTÉRÊT GÉNÉRAL



PRESTATAIRE

Moteur déterministe et interface d'annotation, évolutions difficiles.

prestataire externe

EIG 3

Moteur de pseudonymisation avec intelligence artificielle

data scientists

EIG 4

Interface d'annotation innovante – humain et la machine

développeurs et designer

CATÉGORIES REPÉRÉES PAR LE MOTEUR DE PSEUDONYMISATION

- Noms et prénoms de personne physique (parties ou tiers)
- Noms et prénoms de professionnel (juge, greffier, avocats)
- Différentes dates civiles (naissance, décès et mariage)
- Adresses
- Personnes morales
- Le cadastre
- Les plaques d'immatriculation
- Les comptes bancaires
- Le numéro identifiant INSEE
- Numéro de téléphone/fax/mail

CE QUE FAIT LE MOTEUR DE PSEUDONYMISATION PAR MACHINE LEARNING



Le système va apprendre l'ensemble **des règles empiriques** permettant d'obtenir le comportement désiré.

Le modèle va **intégrer de manière exhaustive** les différents contextes dans lequel une entité peut apparaître.

LES PRÉREQUIS À UN PROJET D'APPRENTISSAGE AUTOMATIQUE



Corpus de décisions de justice

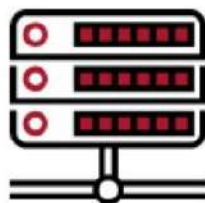


Documents annotés (Annotateurs)



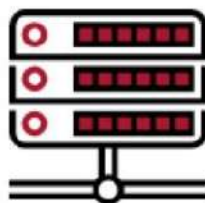
Puissance de calcul (gpu)

BASES DE DONNEES NATIONALES (STOCK)



JURINET

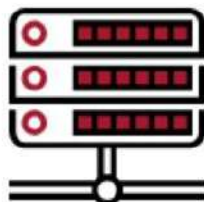
867 403 décisions*



JURICA

2 122 665 décisions*

Administrées
par la Cour de
cassation

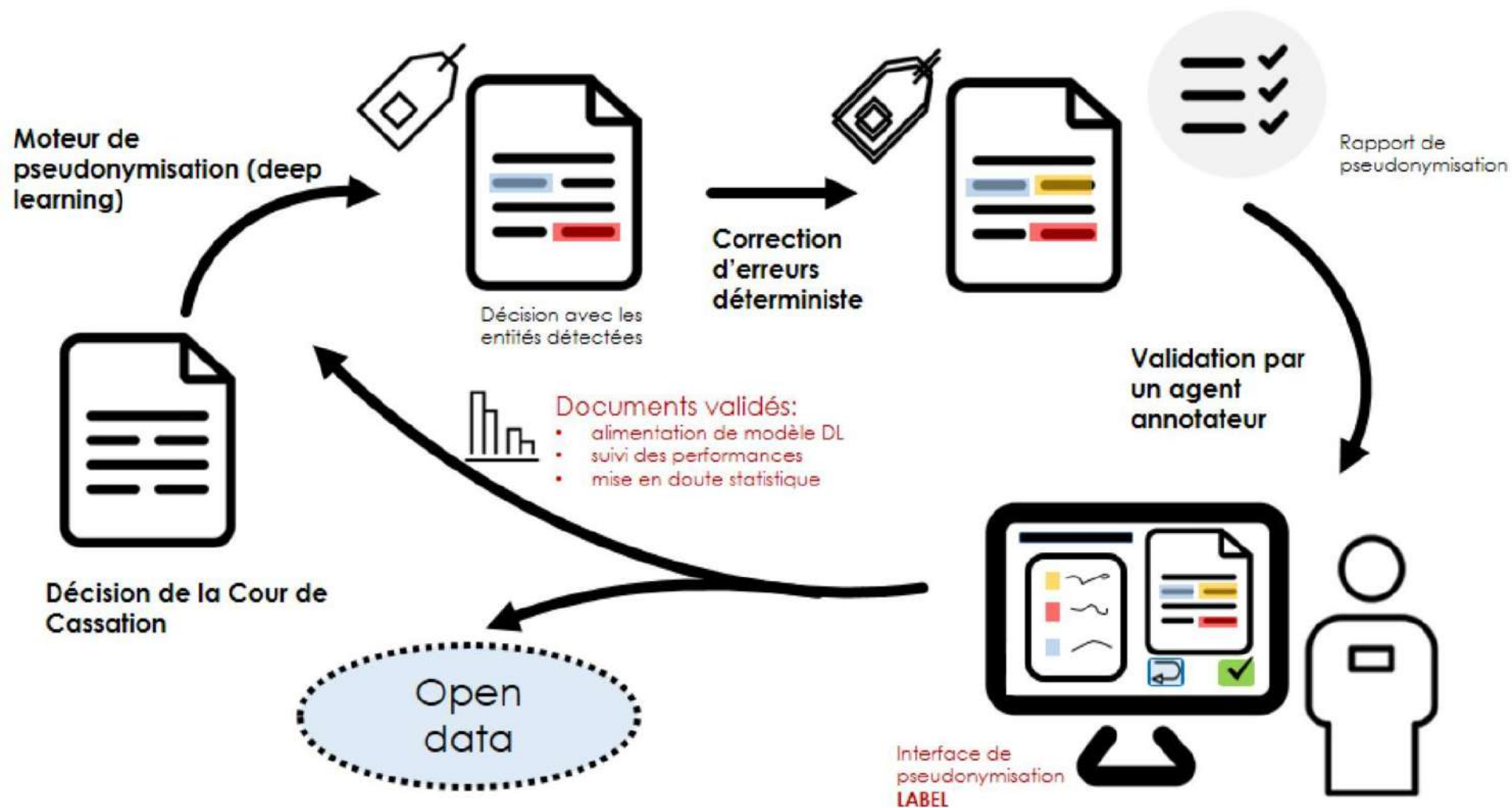


LEGIFRANCE

650 398 décisions*

Pseudonymisées
par la cour de
cassation

LA VUE GLOBALE DE PSEUDONYMISATION A VENIR






LA CHANTIER EN COURS DE DATA SCIENCE EN SYNERGIE AVEC LA NOUVELLE INTERFACE

Moteur de pseudonymisation

-  Liste des entités détectées par le modèle nlp
-  Les corrections déterministes
-  Rapport de pseudonymisation (mise en doute déterministe)







Label: Interface de pseudonymisation


-  Validation & correction par agent annotateur
-  Gestion de documents
-  Suivi de l'annotation manuelle





Méta-analyse: amélioration continue


-  Suivi de performances
-  Enrichissement de tests unitaires
-  Calcul du seuil de la mise en doute statistique
-  Sélection de documents pour l'entraînement supplémentaire du modèle nlp



 Alimentation de la base d'entraînement

 Alerte de mise en doute statistique

 Alertes de performance

 Documents validés par les agents : données « gold »

PROJET LABEL : DES LEVIERS D'ACCÉLÉRATION MULTIPLES



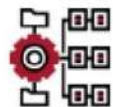
Une méta-analyse des annotations produites pour une amélioration continue.



Une relecture différenciée des décisions en fonction de leur sensibilité.



Une refonte ergonomique.



Une clarification et simplification des règles d'annotation et d'occultation.

DIFFUSION RENOVEE DE LA JURISPRUDENCE



Nouveau site internet de la Cour de cassation rénové



Moteur de recherche de jurisprudence performant



Merci !