



GATES

Déployer de l'ingénierie pour gérer les données

Projet IDEX Justice Algorithmique des élections (JADE) 2022-2025

Romain Rambaud, Professeur de Droit Public, CRJ
Louis Maritaud, Ingénieur de Recherche GATES

Description du projet

- **Constat** : montée en puissance de la « justice prédictive », de la « jurimétrie » mais peu de travaux en droit public, volonté de développer un projet en IA
- **Hypothèse / contentieux des élections politiques** : contentieux ancien, standardisé, normé, chiffré, donc plus réceptif aux outils de la justice algorithmique.... **Problématique** : quelle rationalité = notion de sincérité du scrutin, écart de voix ? (v. R. Rambaud, AJDA)
- **Méthodologie** : hypothèse qu'une « méthodologie » de la justice algorithmique se développe MAIS qu'elle est actuellement surtout développée par des start-ups / Legal tech donc échappent à la recherche publique et au service public (thèse A. Hafsaoui).

Élaboration technique des bases de données JADE

- Les bases de données Décisions : la problématique de la (non)préparation des données / métadonnées par les juridictions (XML – Zip Datagouv)

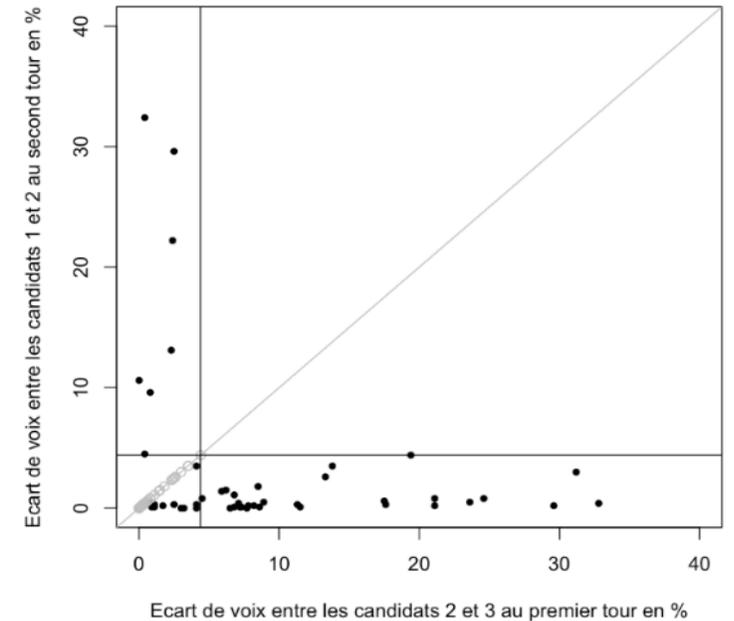
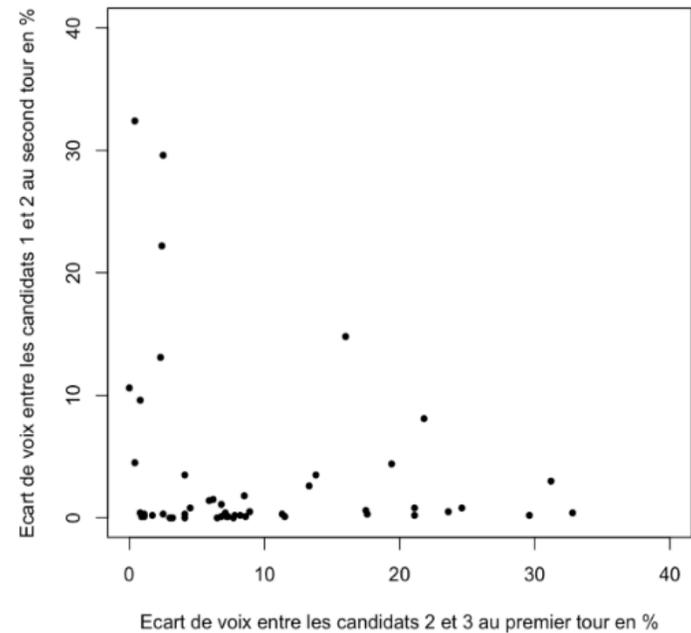
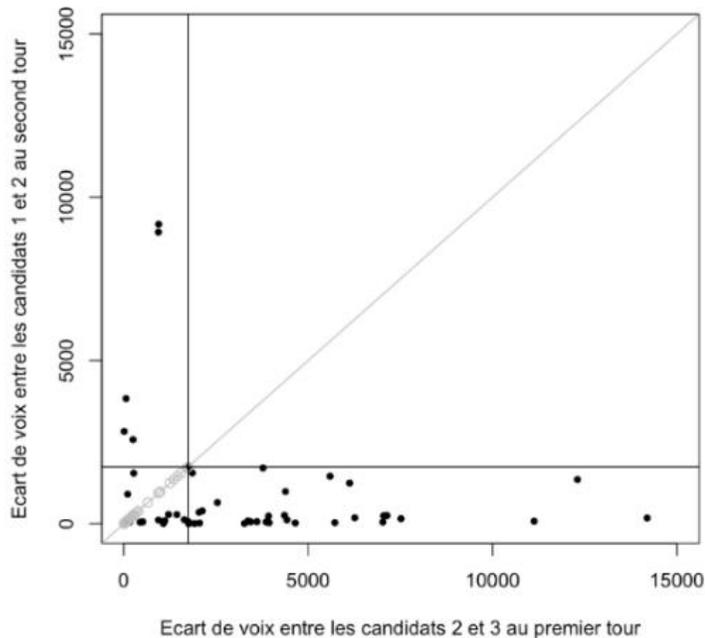
- **Conseil constitutionnel** : Les balises sémantiques suivantes permettent de trier informatiquement les décisions selon : <ECLI> (numéro unique de la décision), <URL> (lien internet), <NATURE> (AN, SEN, etc.), <TITRE> (type d'élection et circonscription concernée), <DATE_DEC> (date) et <SOLUTION> (sens de la décision). Il existe en plus sur certains décisions les « Abstracts » en JSON

- **Conseil d'Etat** :

Code PJCA	Texte Code PCJA	Total décisions	Balises solution remplies			Balises solution non remplies	
			Nombre	%	Nb Métadonnées	Nombre	%
28-005	Dispositions générales	240	98	40,8	25	142	59,2
28-01	Présidentielle	13	7	53,8	5	6	46,2
28-02	Législatives	42	23	54,8	7	19	45,2
28-023	Européennes	37	10	27,0	3	27	73,0
28-025 sauf 28-025-04	Régionales	176	29	16,5	10	147	83,5
28-03	Départementales	465	122	26,2	23	343	73,8
28-04	Municipales	1868	470	25,2	61	1398	74,8

Résultats de JADE : des statistiques descriptives...

1) La prise en compte par le juge d'un écart de voix faible entre les candidats 2 et 3 du 1^{er} tour ou entre les candidats 1 et 2 du 2nd tour



2) Analyse du faible écart de voix pris en compte par le juge de l'élection

Indicateur statistique	Valeur Absolue	% des suffrages exprimés
Minimum	5	0
Premier quartile (25%)	56	0,1
Médiane	119	0,4
Troisième quartile (75%)	281	1,4
Extrémité supérieure de la moustache	394	3,0

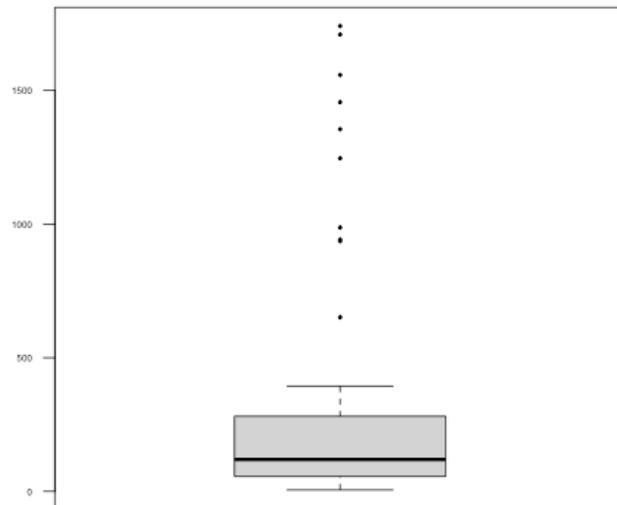


Figure 4 : Plus petits écarts de voix entre les candidats 2 et 3 du tour 1 et les candidats 1 et 2 du tour 2 en valeur absolue

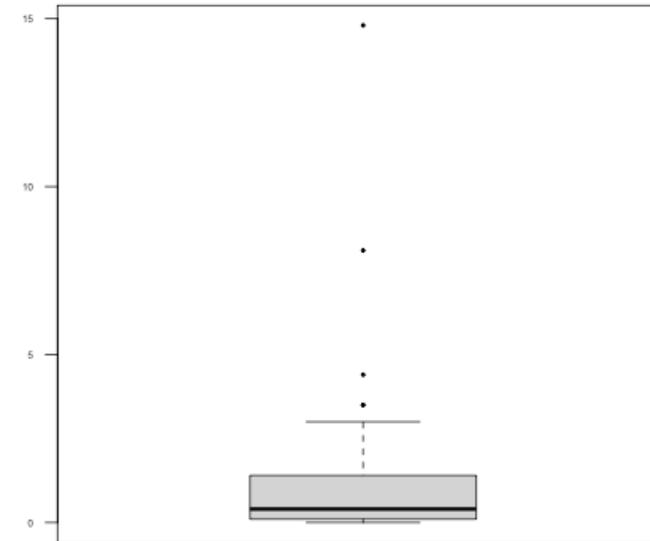


Figure 5 : Plus petits écarts de voix entre les candidats 2 et 3 du tour 1 et les candidats 1 et 2 du tour 2 en %

3) Hypothèse sur l'évolution dans le temps de l'appréciation de l'écart de voix

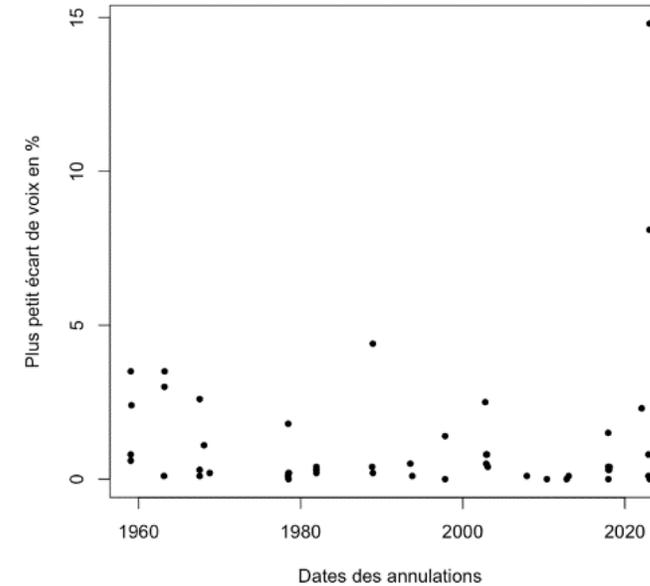
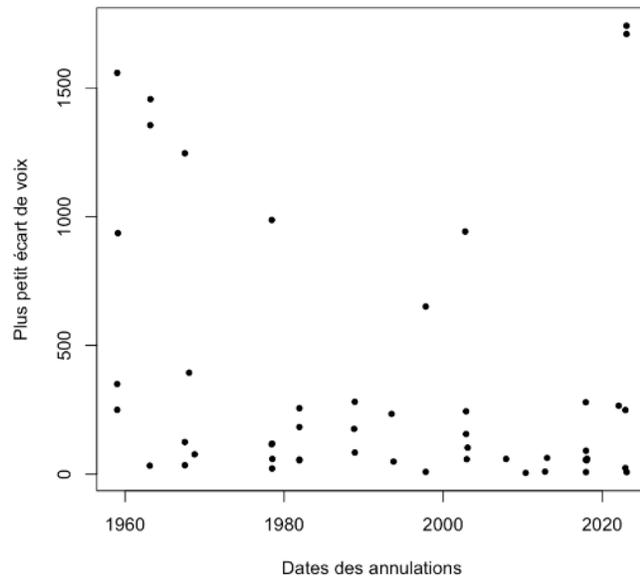


Figure 7 : Plus petit écart de voix entre les candidats 2 et 3 du tour 1 et les candidats 1 et 2 du tour 2 en pourcentage au cours du temps

Figure 6 : Plus petit écart de voix entre les candidats 2 et 3 du tour 1 et les candidats 1 et 2 du tour 2 au cours du temps

A partir de 2002, à l'exception de la décision n°2002-2725 AN (faux positif) et des deux décisions n°2022-5813/5814 AN, et n°2022-5760 AN (dysfonctionnements dans le vote électronique), toutes les annulations ont au moins un écart de voix entre les candidats 2 et 3 du tour 1 ou entre les candidats 1 et 2 du tour 2 inférieur ou égal à 279.

Pour la justice "predictive" : Enrichissement des données et annotation pour faire une analyse par griefs

- Très délicate tâche d'annotation.... L'échec de *Data Just*
- Début du travail d'annotation : quelle efficacité / coût (humain) ?

The screenshot shows the XML Author interface with the following components:

- Menu Bar:** Fichier, Éditer, Rechercher, Projet, Options, Outils, Document, Fenêtre, Aide.
- Toolbar:** Includes XPath 2.0 and 'Exécuter XPath sur "Fichier actuel"'. The XPath expression is: `cc_xml_annotation\AN\2002\rejet\Annot RR\CONSTEXT000017664439_annot - RR.xml`.
- Left Panel (Sommaire):** A tree view showing the document structure: `TEXTE_JURI_CONSTIT` (selected), `META`, and `TEXTE`.
- Main Editor:** Displays XML content with line numbers 86-98. The text includes a list of grievances with attributes like `objet` and `qualification`.

```
86 ;<br/> - SUR LES GRIEFS RELATIFS AU DÉROULEMENT DE LA CAMPAGNE ÉLECTORALE :<br/>
87 5. <Grief objet="4.Campagne_propagande_7_Lettres">Considérant que les lettres
88 mentionnées ci-dessus n'ont introduit dans la campagne électorale aucun
89 élément nouveau auquel M. CAHUZAC n'aurait pas eu le temps de répondre
90 ;</Grief><br/> 6. <Grief objet="4.Campagne_propagande_1_Affiches"
91 qualification="1.Grief_NC_2_Rejet_au_fond">Considérant que, s'il est soutenu
92 que M. MERLY se serait livré à un affichage en dehors des emplacements
93 officiels, en méconnaissance de l'article L. 51 du code électoral, il ne
94 ressort pas du procès-verbal de constat produit par le requérant que cet
95 affichage aurait revêtu un caractère massif ;</Grief><br/> 7. <Grief
96 objet="4.Campagne_propagande_16_Irrégularités_diverse"
97 qualification="1.Grief_NC_2_Rejet_au_fond">Considérant, enfin, que le grief
98 tiré d'une méconnaissance des dispositions de la loi du 6 janvier 1978,
99 relative à l'information sur les fichiers et aux libertés, n'est pas assorti
```
- Bottom Panel (Éléments):** A list of XML elements including `ANCIEN_ID`, `Annonce_conclusion`, `Application_aux_faits_de_l_espece`, `BLOC_TEXTUEL`, `CONTENU`, `DATE_DEC`, and `DECISION`.
- Bottom Panel (Attributs):** A table for attributes with columns for 'Attribut' and 'Valeur'.
- Status Bar:** Shows 'Validation échouée. Erreurs: 1.' and 'U+003C 1 : 1'.

Traitement des données

Mise en forme de données déjà annotées selon plusieurs paramètres pour permettre des comparaisons de technologies et méthodologies

Deux modalités de TAL

1- Pipeline complète, sur Bigfoot

- Nécessite beaucoup de ressources, un accès et du temps de traitement

2- Chaîne de traitement allégée, tourne en local

- Dans une perspective de comparaison de technologies

Associées à plusieurs manières de représenter l'information

Séquençage des annotations sur la base des informations qu'elles contiennent, dans la dynamique exploratoire et comparative

8 processus parallèles, autant utiles pour le projet que dans l'aspect

→ méthodologie/reproductibilité

Des annotations plurielles

Un élément est annoté simultanément sur plusieurs aspects :

- L'objet du grief, ce qui concerne la demande
- la qualification, c'est à dire l'avis du juge

Chaque aspect comprend plusieurs informations (encadrés de couleur)

```
<Grief objet='6.Financement 5 Dépenses"  
qualification='1.Grief NC 2 Rejet au fond 2 non fondé">
```

→ Réduction des annotations pour opérer sur des “super-catégories”, puis annotations des éléments dépendants sur lesdites “super-catégories”

Entraînement du modèle et annotation automatique

- Utilisation d'une pipeline orientée CPU pour qu'elle puisse tourner en local
- Pour chaque texte non-annoté :
 - Séparation des séquences à annoter par une série d'expressions régulières (“[Numéro]. “, puis par “;”)
 - Annotation automatique des séquences avec un seuil au choix de l'utilisateur
 - Reconstitution du fichier désormais annoté

Sur l'envoi des votes par correspondance aux bureaux de vote :

11. Considérant qu'il résulte de l'instruction que, contrairement aux dispositions de l'article R.87 du code électoral, 72 électeurs inscrits sur les listes électorales des communes de

ont voté par correspondance sans expédier leur vote sous pli recommandé ; que ces 72 votes sont nuls ; qu'il y a lieu, par application de la règle ci-dessus, d'opérer sur les suffrages obtenus par chacun des deux candidats la déduction de ces votes en fonction du résultat qu'ils ont obtenu dans chaque bureau de vote ; qu'en conséquence, 24 voix doivent être retranchées du nombre de suffrages recueillis par M. et 48 voix doivent être déduites du nombre de suffrages dont M. avait été crédité ;

12. Considérant que, contrairement aux allégations de la requête, il n'est pas établi que les votes par correspondance de 3 électeurs du 2e bureau de la ville de aient été acceptés sans

Post-traitement des données annotées

- Récupération des annotations automatiques pour chaque fichier (procédure similaire au pré-traitement)
- formatage des données ainsi récupérées au format .csv en vue d'une vérification humaine

Vérification humaine	Remarque / état de l'annotation	Annotation alternative proposée	Fichier	Annotation	Texte
Oui	un considérant annoté sur 3		\\AN\1963\annulation\CONSTEXT000017665174_annot.xml	7.Opérations	1. Considérant
Oui			\\AN\1967\annulation\CONSTEXT000017665356_annot.xml	7.Opérations	4. Considérant
Oui			\\AN\1967\annulation\CONSTEXT000017665356_annot.xml	7.Opérations	5. Considérant
Oui	considéran ts 3, 4, 5 ne sont pas annotés		\\AN\1968\annulation\CONSTEXT000017665568_annot.xml	7.Opérations	2. Considérant
Oui	considérant 4 n'est pas annoté		\\AN\1978\annulation\CONSTEXT000017665825_annot.xml	7.Opérations	3. Considérant
Oui	considéran ts 6, 8, 9, 10, 11, 12, 13, 16, 18 ne sont pas annotés		\\AN\1978\annulation\CONSTEXT000017665826_annot.xml	7.Opérations	1. Considérant
Oui	considéran ts 6, 8, 9, 10, 11, 12, 13, 16, 18 ne sont pas annotés		\\AN\1978\annulation\CONSTEXT000017665826_annot.xml	7.Opérations	3. Considérant
Oui	considéran ts 6, 8, 9, 10, 11, 12, 13, 16, 18 ne sont pas annotés		\\AN\1978\annulation\CONSTEXT000017665826_annot.xml	7.Opérations	4. Considérant
Oui	considéran ts 6, 8, 9, 10, 11, 12, 13, 16, 18 ne sont pas annotés		\\AN\1978\annulation\CONSTEXT000017665826_annot.xml	7.Opérations	5. Considérant
Oui	considéran ts 6, 8, 9, 10, 11, 12, 13, 16, 18 ne sont pas annotés		\\AN\1978\annulation\CONSTEXT000017665826_annot.xml	7.Opérations	7. Considérant
Oui	considéran ts 6, 8, 9, 10, 11, 12, 13, 16, 18 ne sont pas annotés		\\AN\1978\annulation\CONSTEXT000017665826_annot.xml	7.Opérations	14. Considérant
Oui	considéran ts 6, 8, 9, 10, 11, 12, 13, 16, 18 ne sont pas annotés		\\AN\1978\annulation\CONSTEXT000017665826_annot.xml	7.Opérations	15. Considérant
Oui	considéran ts 5 et 7 ne sont pas annotés		\\AN\1981\annulation\CONSTEXT000017665942_annot.xml	7.Opérations	4. Considérant
Non	considéran ts 3, 4 et 5 ne sont pas annotés	ne devrait être annoté (Rappel des faits, ici rappel des s	\\AN\1981\annulation\CONSTEXT000017665946_annot.xml	7.Opérations	1. Considérant
Oui	considéran ts 3 et 4 ne sont pas annotés		\\AN\1988\annulation\CONSTEXT000017667715_annot.xml	7.Opérations	2. Considérant
Non	considérant 7 n'est pas annoté	3. Candidature : éligibilité / inéligibilité	\\AN\1988\annulation\CONSTEXT000017667716_annot.xml	4.Campagne	5. Considérant
Oui			\\AN\1998\annulation\CONSTEXT000017667910_annot.xml	4.Campagne	1. Considérant
Oui			\\AN\1998\annulation\CONSTEXT000017667910_annot.xml	4.Campagne	2. Considérant
Oui	considérant 4 n'est pas annoté		\\AN\1999\annulation\CONSTEXT000017667949_annot.xml	7.Opérations	2. Considérant
Oui	considéran ts 2, 3 et 4 ne sont pas annotés		\\AN\2002\annulation\CONSTEXT000017664506_annot.xml	6.Financement	: 1. Considérant
Oui	considéran ts 1 et 2 ne sont pas annotés		\\AN\2002\annulation\CONSTEXT000017664522_annot.xml	6.Financement	4. Considérant
Oui	considéran ts 1 et 2 ne sont pas annotés		\\AN\2002\annulation\CONSTEXT000017664522_annot.xml	6.Financement	5. Considérant

Une solution trouvée via dépôt sur Recherche Data Gouv

- Bligny, Caroline; Rambaud, Romain, 2024, “Cross-referenced data on electoral disputes and French legislative election results”, <https://doi.org/10.57745/0BI3LB>, Recherche Data Gouv, V1.

2 "dépôts" :

- Le dépôt du Dataset jade sur recherche.data.gouv.fr, pour lequel les meta données sont visibles mais pas les données, tant que les questions rgpd ne sont pas claires.-
 - Le dépôt du code de construction et d'exploitation de la BDD jade sur gitlab (qui contient aussi le code de Louis), avec une licence ouverte mais accès restreint aussi pour la même raison
-
- Accès pour l'instant **restreint** du fait des questions RGPD & CNIL
 - Par la suite ?

Avantages et limites de la collaboration JADE / GATES

Enorme avantage pour JADE : l'aide de GATES pendant 6 mois a permis au projet JADE de commencer à expérimenter le TAL pour aller plus loin (justice prédictive)

Limites :

- Un mi-temps pendant 6 mois n'a pas été suffisant pour avoir de bons résultats, il a fallu plus de temps
- Le projet JADE a proposé qu'un article soit écrit pour conserver les résultats de cette expérimentation, mais il faut du temps laissé à l'ingénieur pour l'écrire, et il y a des débats concernant l'état de l'art (pluridisciplinarité)
- La pérennité de JADE dépend aujourd'hui de la réussite à une ANR aux très faibles taux de réussite (réponse phase 1 en février)
- Si pas d'article et/ou pas d'ANR, tout le travail effectué va se perdre définitivement

Pour en savoir plus

Blog du droit électoral :



blogdudroitelectoral.fr/justice-algorithmique-des-elections-jade/

Blog du pool d'ingénieurs GATES



gates-data-shs.gricad-pages.univ-grenoble-alpes.fr/blog/

Publications

- 1) Rambaud R., « Les "lois" de l'écart de voix », AJDA, 2020, p. 1596.
- 2) Rambaud R., Hafsaoui A., Bligny C., « Une justice algorithmique pour les élections ? », AJDA, 2023, p. 1323.
- 3) Rambaud R., C. Bligny, A. Hafsaoui, S. Cottin, « Justice algorithmique des élections (JADE): une nouvelle base de données du contentieux des élections parlementaires devant le Conseil constitutionnel », Jurimétrie, décembre 2024.
- 4) Rambaud R., C. Bligny, F. Letué, M.-J. Martinez, S. Cottin, J.-P. Camby, G. Prunier, D. Girard, A. Hafsaoui, K. Deschamps, "Projet Justice algorithmique des élections (JADE) : une analyse statistique de la jurisprudence du Conseil constitutionnel relative aux élections législatives (4 oct. 1958 – 1er avril 2024). Partie 1/2 – La structuration générale du contentieux AN, Revue française de droit constitutionnel, n°140, janvier 2025.
- 5) Rambaud R., C. Bligny, F. Letué, M.-J. Martinez, S. Cottin, J.-P. Camby, G. Prunier, D. Girard, A. Hafsaoui, K. Deschamps, "Projet Justice algorithmique des élections (JADE) : une analyse statistique de la jurisprudence du Conseil constitutionnel relative aux élections législatives (4 oct. 1958 – 1er avril 2024), Partie 2/2 – Sincérité du scrutin et critère de l'écart de voix", n°141, mars 2025.