

Appliquer plusieurs perturbations replace en même temps

Contexte

début : 22 juin Objectifs : - Mettre en evidence un domaine et non simplement des mots individuellement - Réduire les effets du bruit avec des perturbations plus impactantes

Dataset défense fait main V2

Construction des données : - 1 question, 3 pdf (limitations matérielles du drone, code de la défense - matériel de guerre, fausses infos pdt les JO/JP) - 50% de textes pertinents - 28 paires

Question : > Quelles sont les trois principales catégories de limitations techniques affectant les systèmes de drones, et comment chacune influence-t-elle l'efficacité opérationnelle de ces vecteurs ?

word	freq	TF-IDF
matériel	16	0.0627430130086598
système	11	0.0542242908052481
limitation	6	0.0495565089630139
drone	10	0.0441649020535004
défense	10	0.0440955901918619
prendre	6	0.0425658179515979
compte	5	0.0404563824330286
action	8	0.0401684995180441
article	9	0.0400092952253956
ministre	8	0.03913497111578
aérien	7	0.0366033210120314
acteur	6	0.0337084867312698
guerre	8	0.0336043799517422

Perturbations

Construction des perturbations: - fréquence et TF-IDF des documents - contenu de la question

Perturbation qui visent la langue de spécialité (synonyme du sens de la langue courante) -> exploitation de la polysémie

Perturbation 1:

aérien -> poétique
technique -> difficile
défense -> corne

limitation -> bornage
matériel -> concret
système -> ensemble
article -> déterminant
drone -> transformation

Perturbation 2:

aérien -> vaporeux
technique -> procédé
défense -> justification
limitation -> démarcation
matériel -> outil
système -> doctrine
article -> achat
drone -> abeille

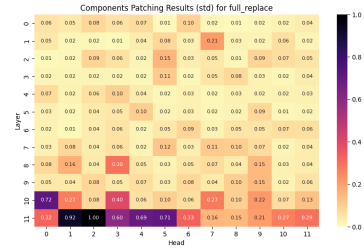
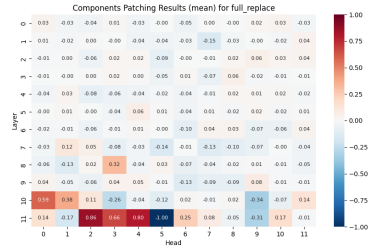
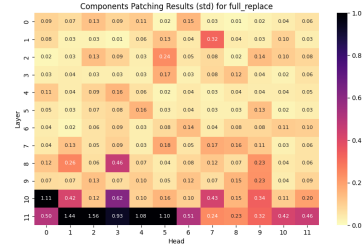
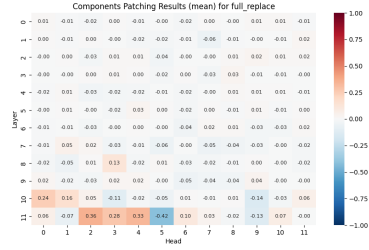
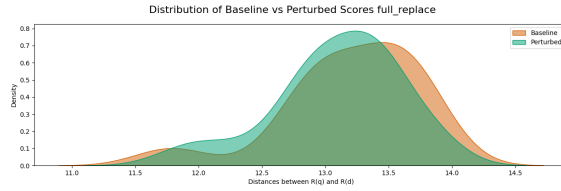
Perturbation 3:

aérien -> diaphane
technique -> fin
défense -> sport de combat
limitation -> régulation
matériel -> palpable
système -> institution
article -> publication
drone -> bourdon

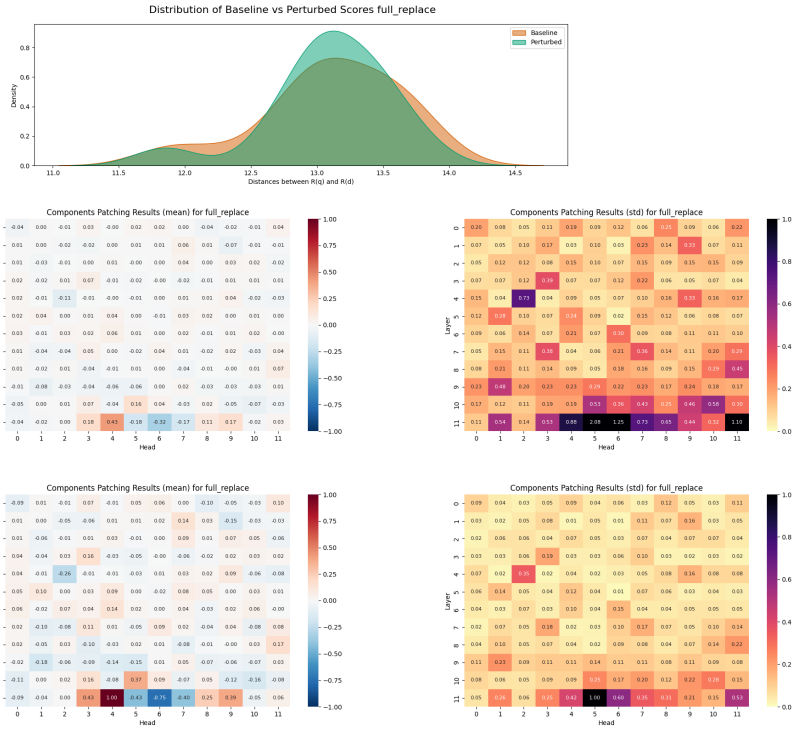
Résultats

Gestion des NaN Les trois perturbations génèrent le même nombre de NaN : 1872

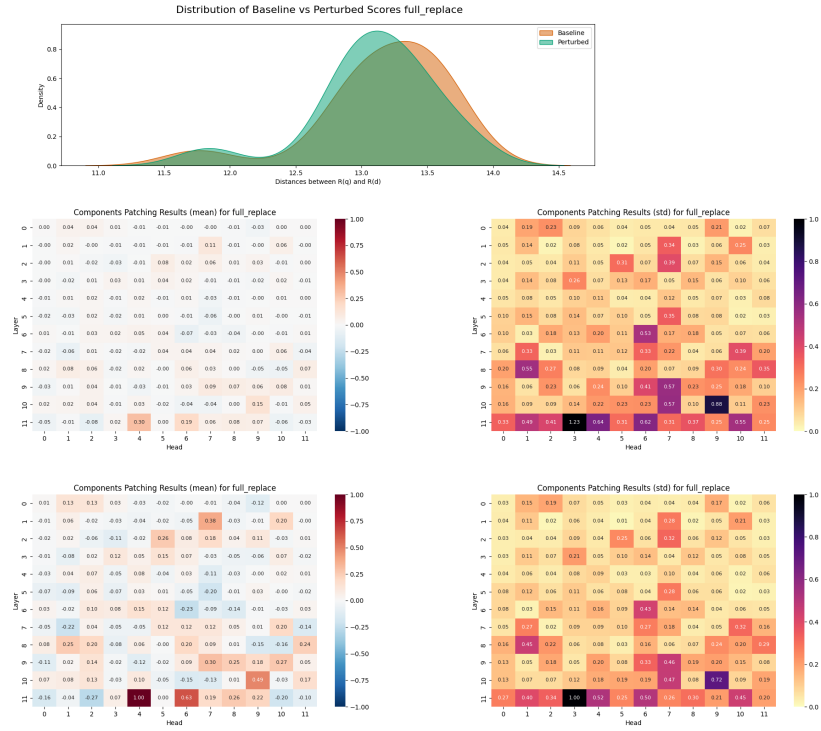
Perturbation de la représentation des documents Les trois perturbations ont un impact moyen sur la représentation des documents dans l'espace latent. Les trois perturbations sont différentes.



Perturbation 1



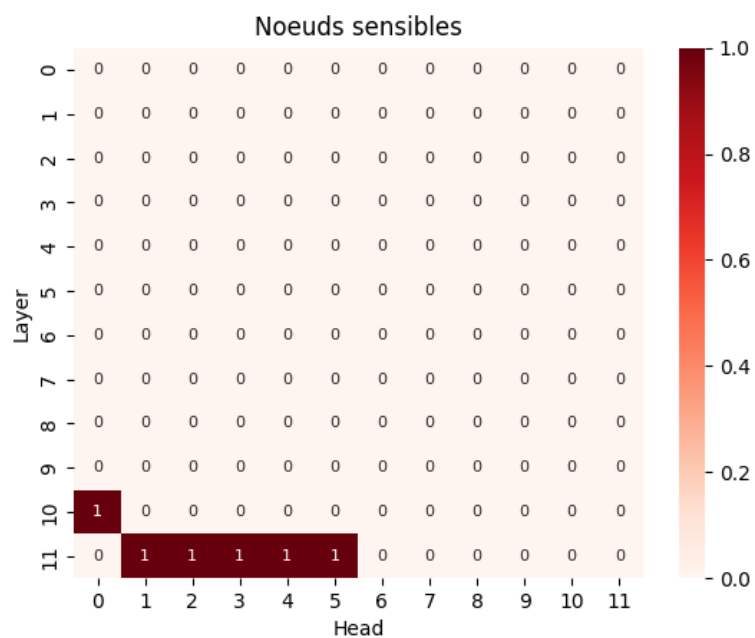
Perturbation 2



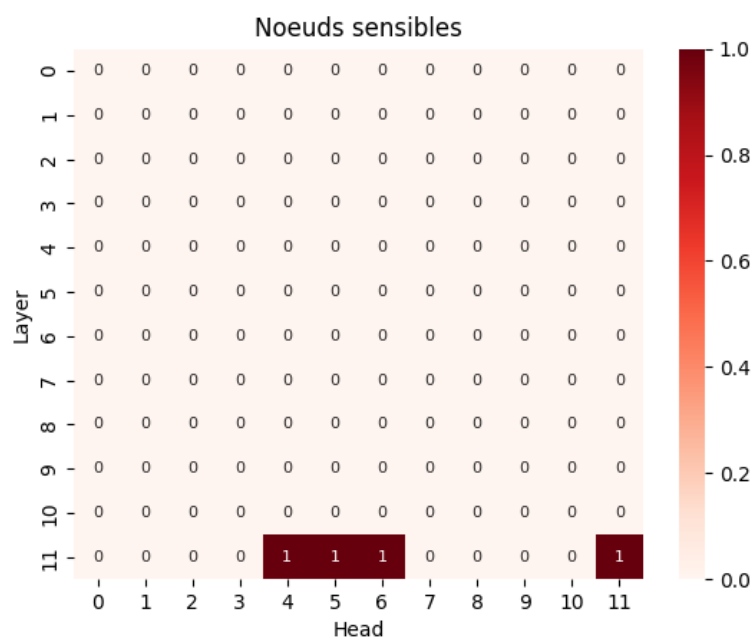
Perturbation 3

Noeuds sensibles Les trois perturbations explorées ont un impact sur des composants du modèle. Les moyennes de sensibilité les plus élevées vont de 0.43 à 0.18 en valeur absolue (ordre de grandeur satisfaisant). Pour chaque perturbation, autour de 5 noeuds se détachent à première vue. La troisième perturbation semble avoir moins de noeuds sensibles, chose confirmée par les écarts-types.

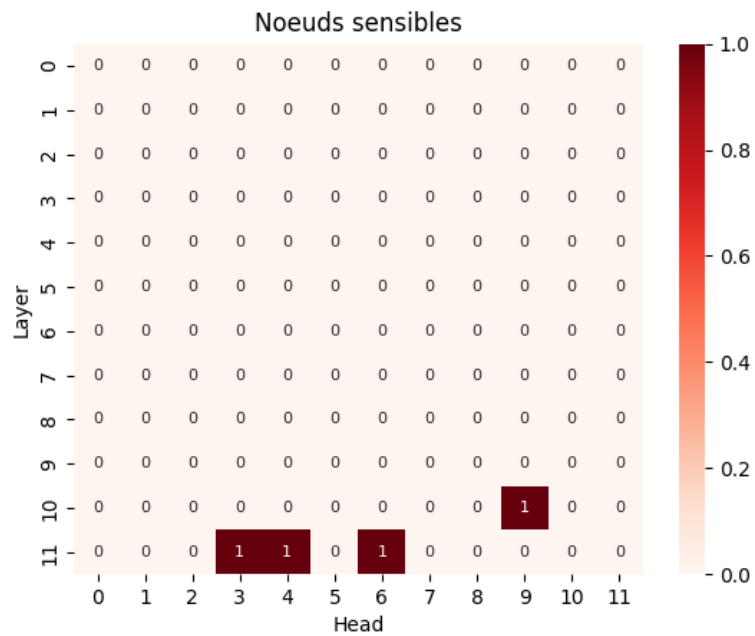
À partir des indicatrices de noeuds sensibles selon le critère (**mean** > 0.1 et **rescaled_mean** > 0.5) ou **rescaled_std** > 0.5. Seul un noeud sensible est commun aux trois perturbations explorées : (11,4). Les noeuds sensibles sont majoritairement dans la couche 11. Les deux autres sont de la couche 10.



Perturbation1



Perturbation2



Perturbation3

Analyse

Les noeuds sensibles sont majoritairement dans la dernière couche du modèle.