

Réunion 12 juin 2026

- Nouveau support d'application automatique de perturbations
- Dataset vaswani - nouvelles remarques
- Dataset défense main V2 avec une question (Expé et Résultats)
- Perturbation replace *Elephant rose* sur défense main V2
- Perturbation append/prepend *Elephant rose* sur défense main V2

Nouveau support d'application automatique de perturbations

Variables utiles facilement accessible:

- modèle
- dataset : nom, moyen de récupération (ir_datasets ou csv)
- champs de csv à considérer
- fichier des perturbations
- calcul automatique de la carte (moyenne et écart-type)
- chemin de sauvegarde des graphiques
- sauvegarde des valeurs calculées

Fonctionnalités nouvelles:

- calcul de la cartographie automatique si possible
- interruption du calcul dès le premier NaN
- sauvegarde des résultats numériques

Le nombre de paires dans Vaswani (microwave)

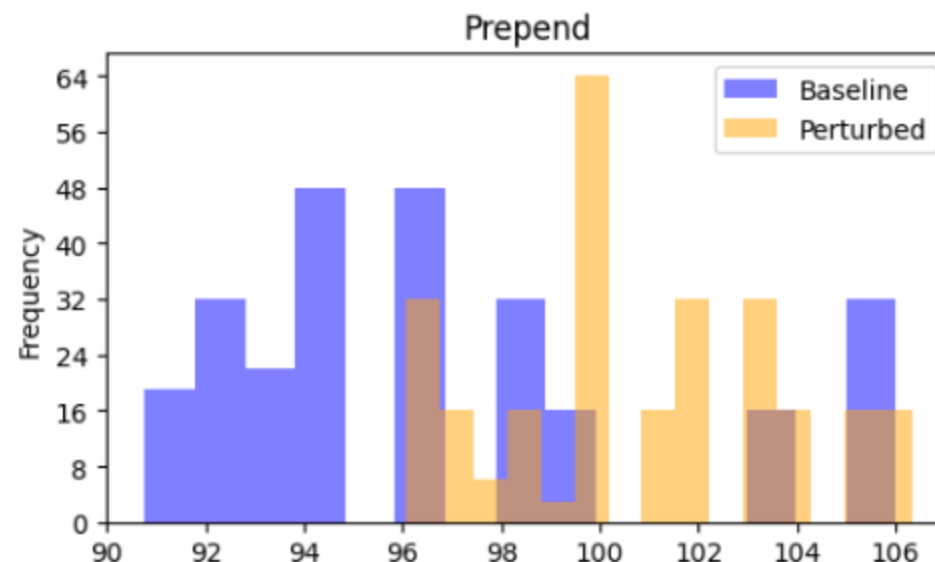
Partie utilisée dans les expés :

- 19 paires
- 1 question seulement (measurement of dielectric constant of liquids by the use of microwave techniques)
- Les paires ne semblent pas étiquetées

2 batchs (16 + 3) ~> 265

On retrouve l'effet 16 ou 3 occurrences d'un même score.

=> Conserver que la diagonale (au lieu d'aplatir)



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

Construction des perturbations:

- fréquence et TF-IDF des documents
- contenu de la 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

Dataset défense fait main V2 - Perturbations

matériel -> concret
matériel -> outil
+matériel
matériel+

système -> doctrine
système -> équation
système -> problème
système -> institution
système -> instance
+système
système+

drone+
+drone

défense -> lutte
défense -> protection
défense -> sport de combat
défense -> contre-attaque
défense -> interdiction
défense -> prohibition
défense -> corne
défense+
+défense

article -> publication
article -> journal
article -> produit
article -> revue
article -> achat
article -> sujet
article -> déterminant
article -> nom
+article
article+

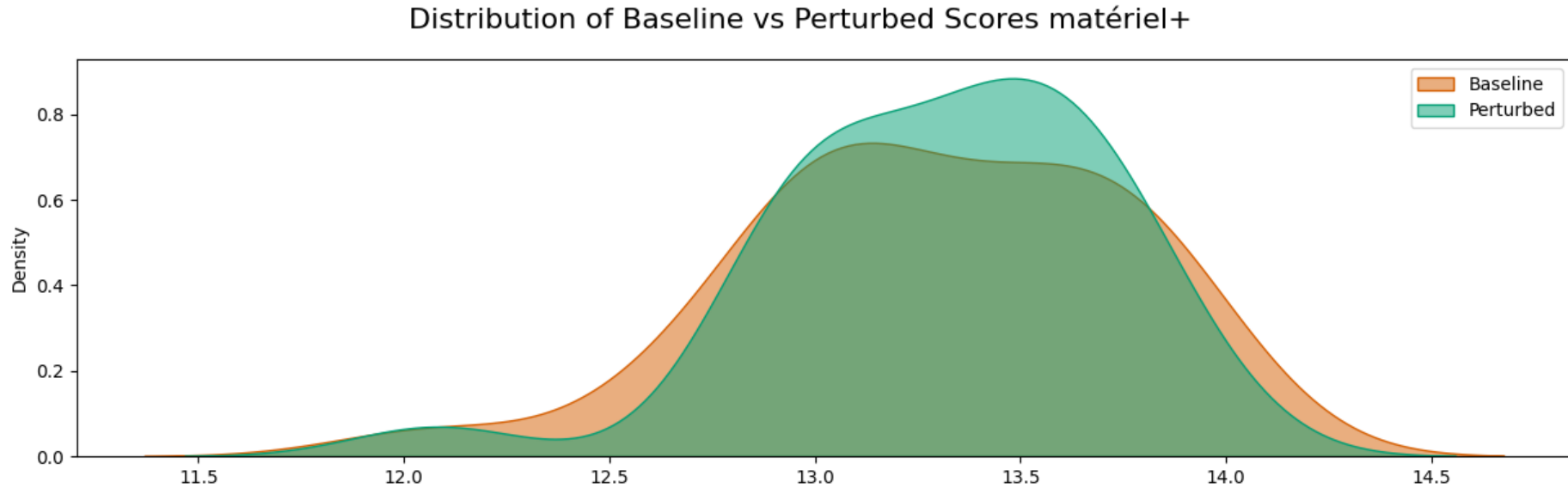
technique -> difficile
technique -> fin
technique -> procédé
+technique
technique+

aérien -> léger
aérien -> souple
+aérien
aérien+

limitation -> bornage
limitation -> démarcation
limitation -> restriction
limitation -> interdiction
+limitation
limitation+

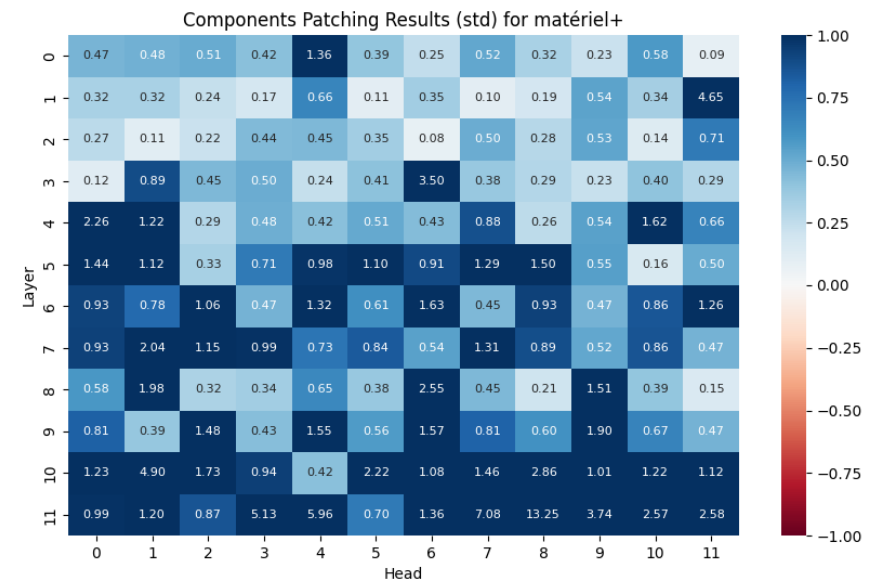
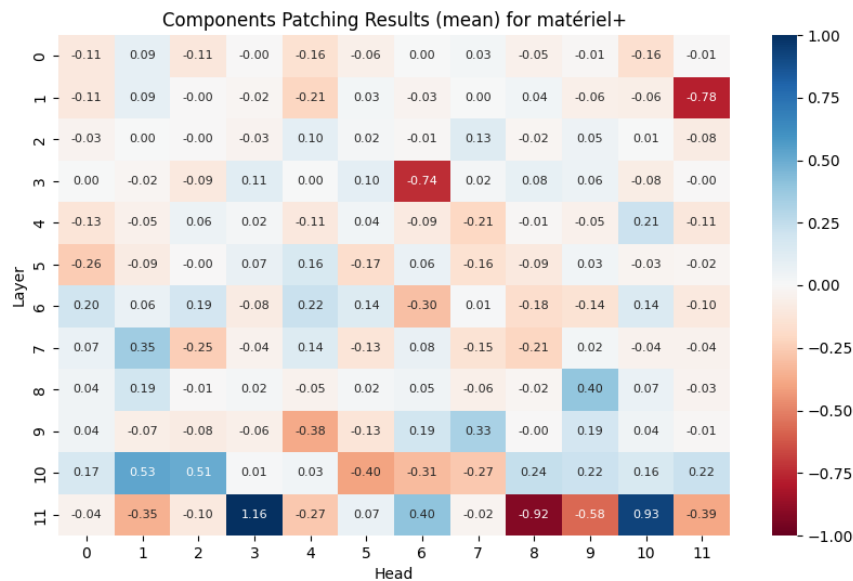
Dataset défense fait main V2 - Résultats

prepend matériel - Perturbation Score



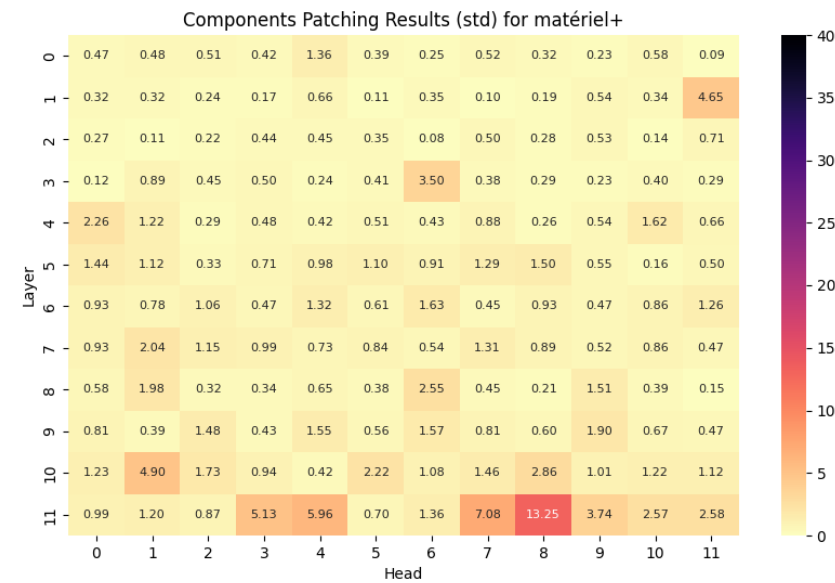
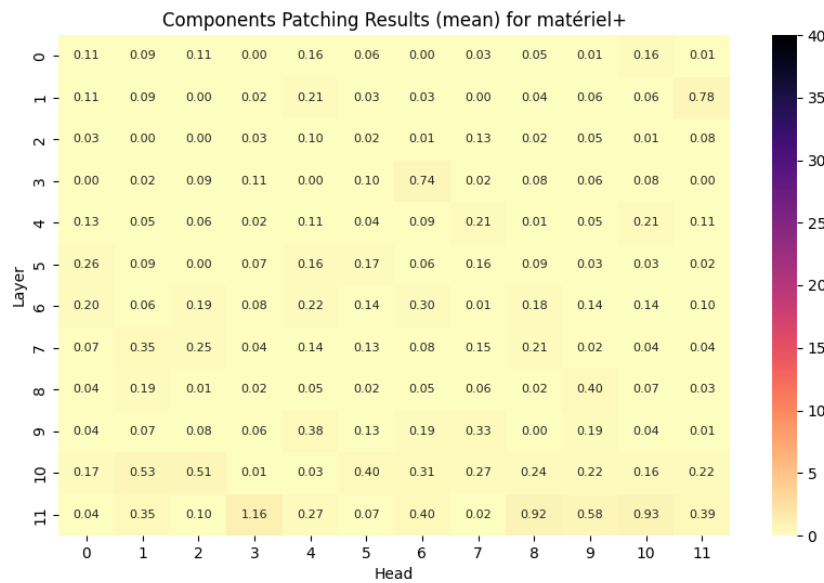
Dataset défense fait main V2 - Résultats

prepend matériel - Component Patching Results



Dataset défense fait main V2 - Résultats

prepend matériel - Component Patching Results



Dataset défense fait main V2 - Résultats

prepend `matériel` - Analyse

Graphiques

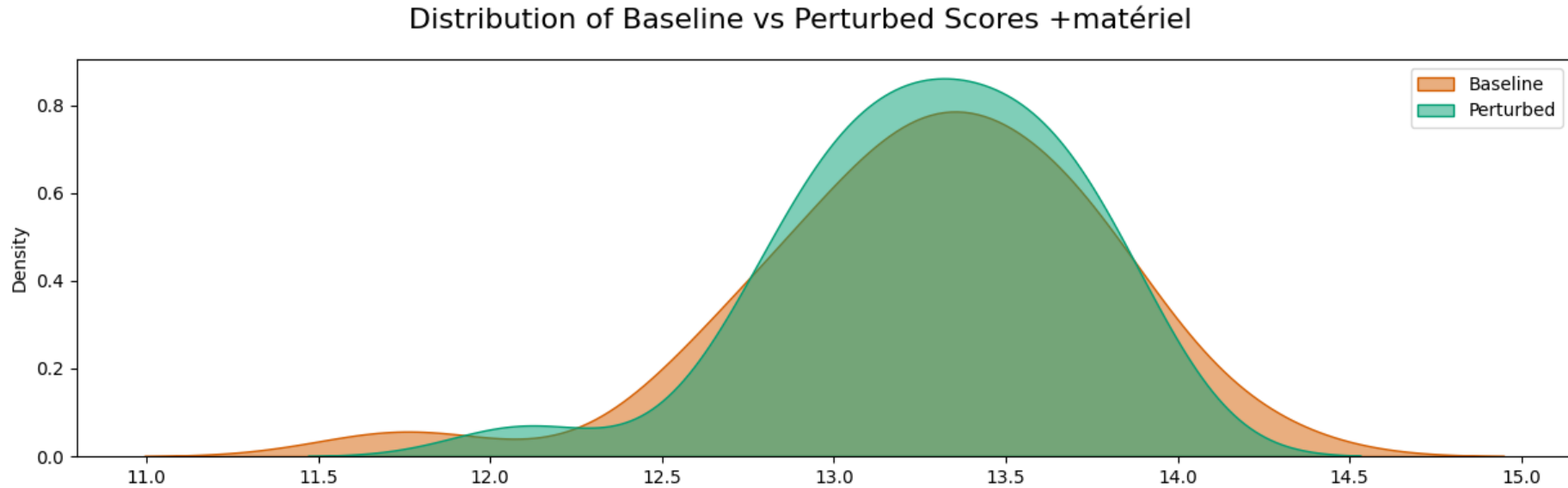
- concentre les extraits dans l'espace de représentation
- Quelques noeuds ressortent mais avec des variances assez élevés
- Des écarts-types globalement très élevés pour une bonne partie des noeuds

Vis-a-vis des données:

- *matériel* est très présent dans les documents (6 extraits issu du chapitre MATERIEL du Code de la défense)
- *matériel* n'est pas présent dans la question et n'est pas lié

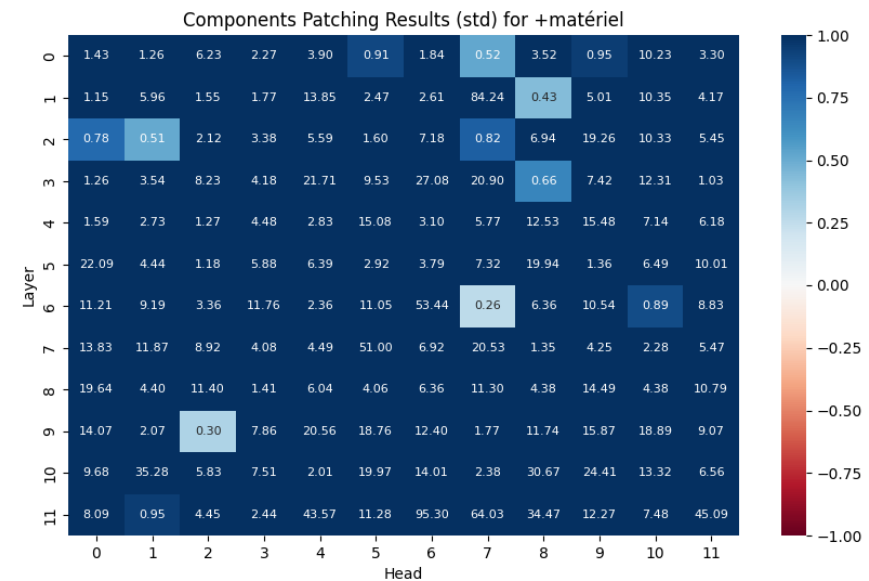
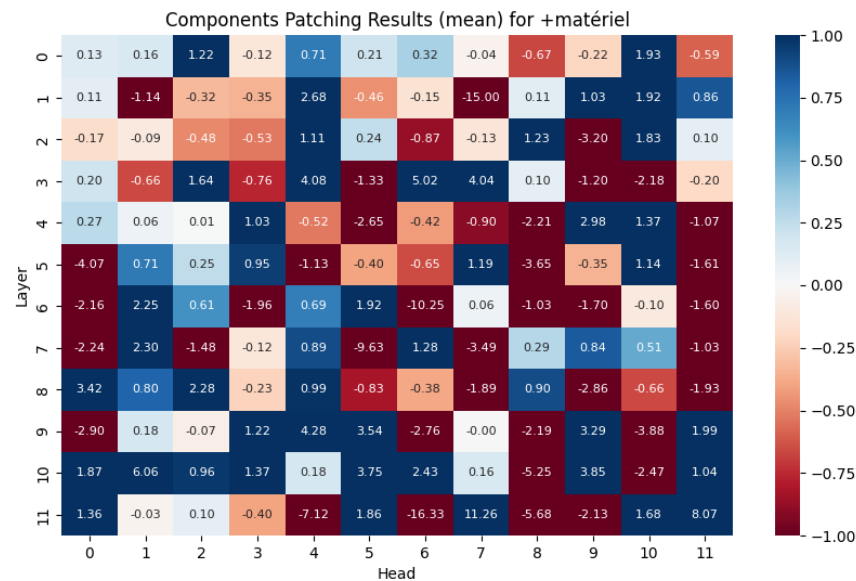
Dataset défense fait main V2 - Résultats

append matériel - Perturbation Score



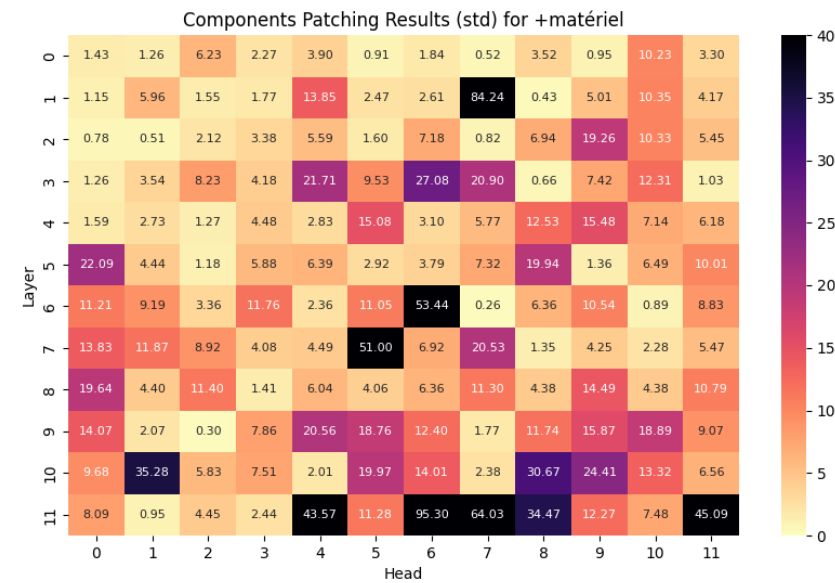
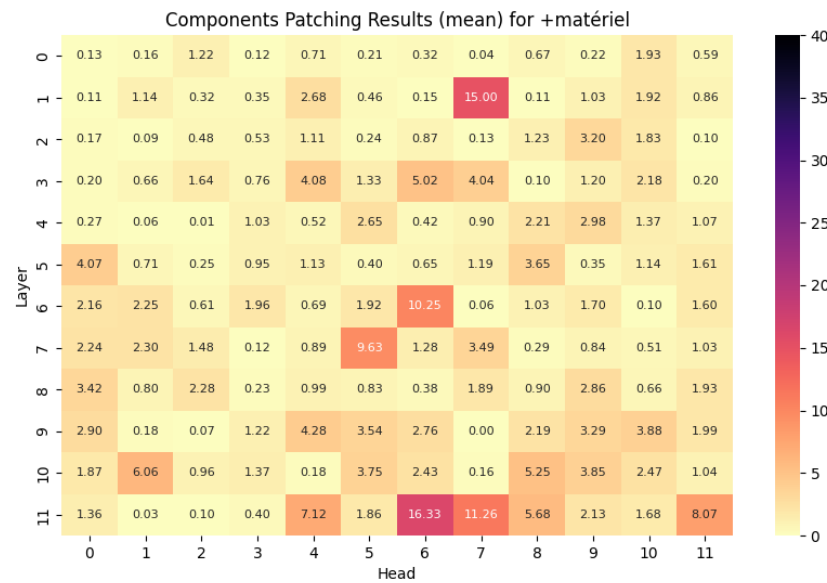
Dataset défense fait main V2 - Résultats

append matériel - Component Patching Results



Dataset défense fait main V2 - Résultats

append matériel - Component Patching Results 2e visualisation



Dataset défense fait main V2 - Résultats

append **matériel** - Analyse

Graphiques

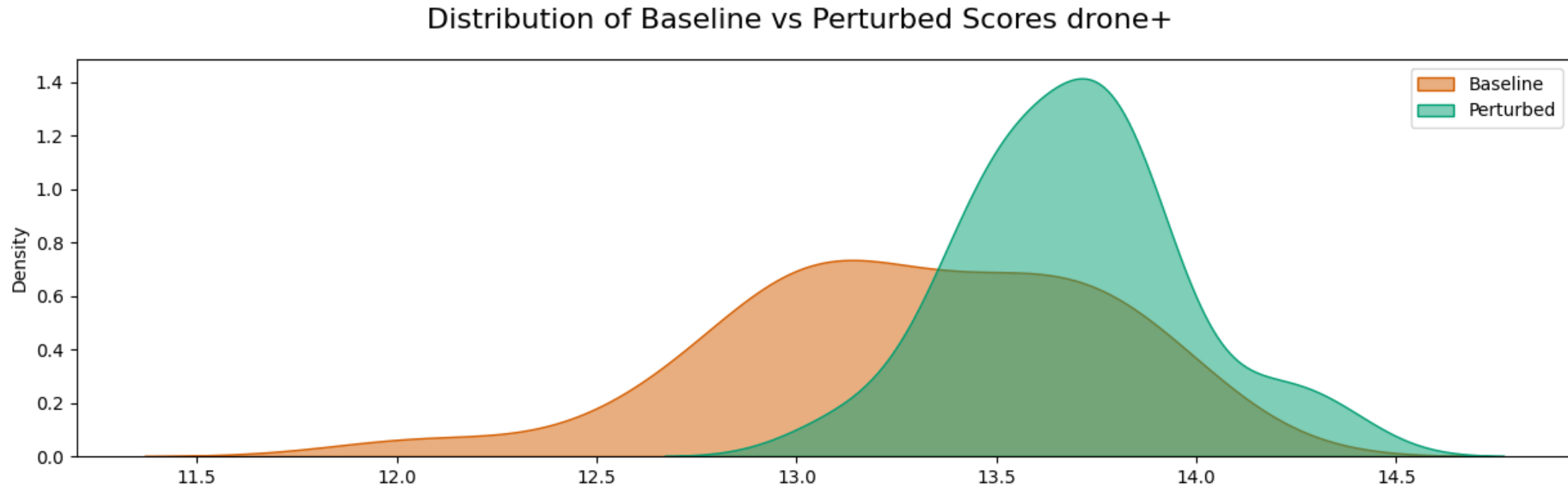
- Perturbation modérée d'après le graphique
- Réaction chaotique des noeuds + grand écart-type

Vis-a-vis des données:

- *matériel* est très présent dans les documents (6 extraits issu du chapitre MATERIEL du Code de la défense)
- *matériel* n'est pas présent dans la question et n'est pas lié

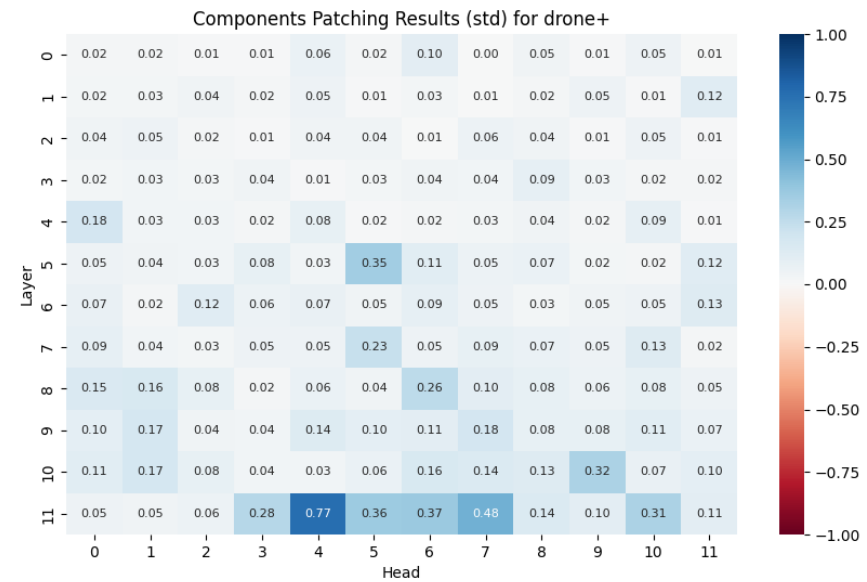
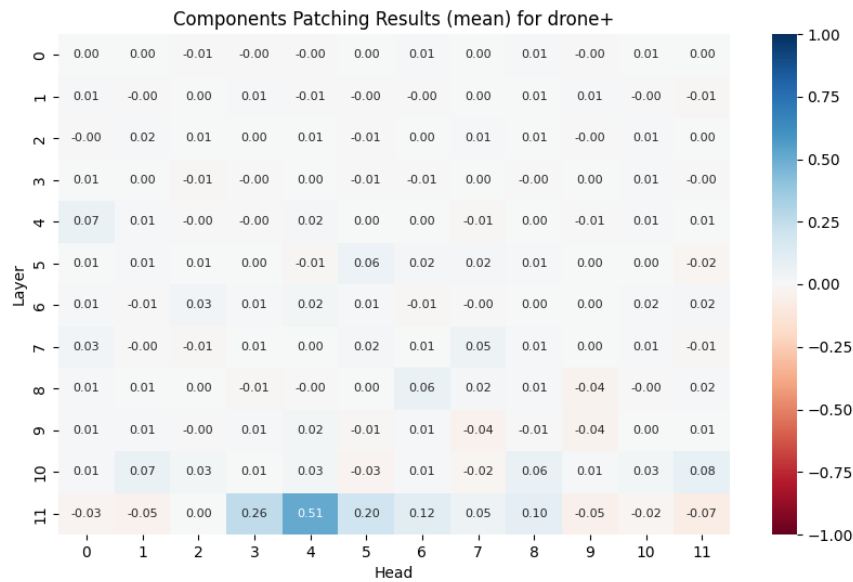
Dataset défense fait main V2 - Résultats

prepend **drone** - Perturbation Score



Dataset défense fait main V2 - Résultats

prepend **drone** - Component Patching Results



Dataset défense fait main V2 - Résultats

prepend **drone** - Analyse

Vis-à-vis des données:

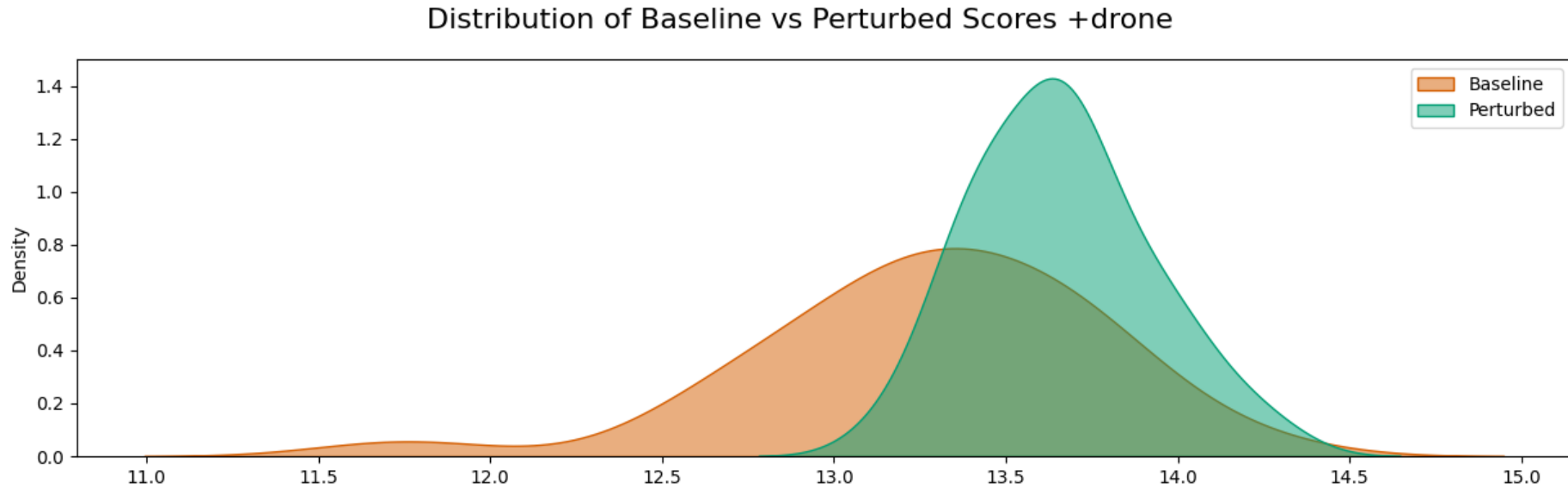
- Apparent dans la question
- Apparent dans 50% des extraits (100% extraits positifs)

Graphiques

- ajouter *drone* regroupe tous les documents
- ajouter *drone* éloigne les documents de la question
- On identifie un noeud plus sensible

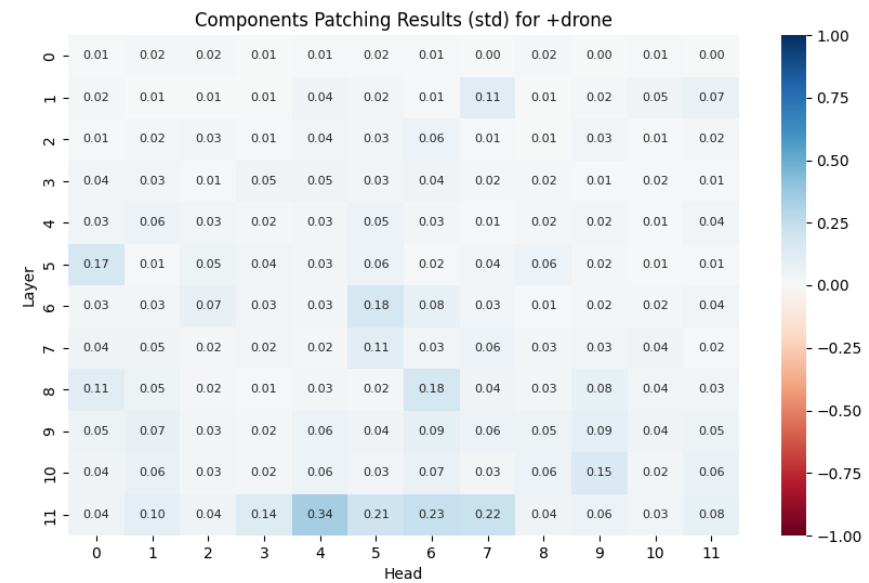
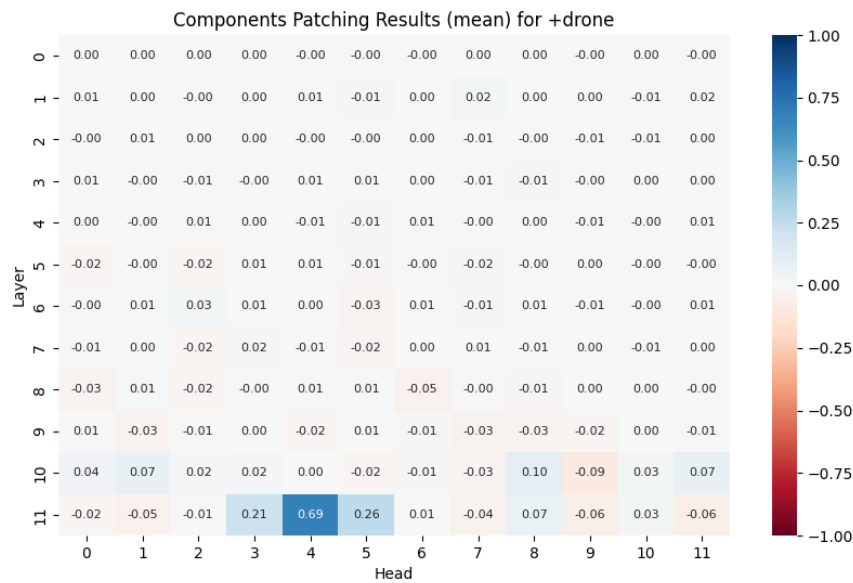
Dataset défense fait main V2 - Résultats

append `drone` - Perturbation Score



Dataset défense fait main V2 - Résultats

append **drone** - Component Patching Results



Dataset défense fait main V2 - Résultats

append **drone** - Analyse

Vis-à-vis des données:

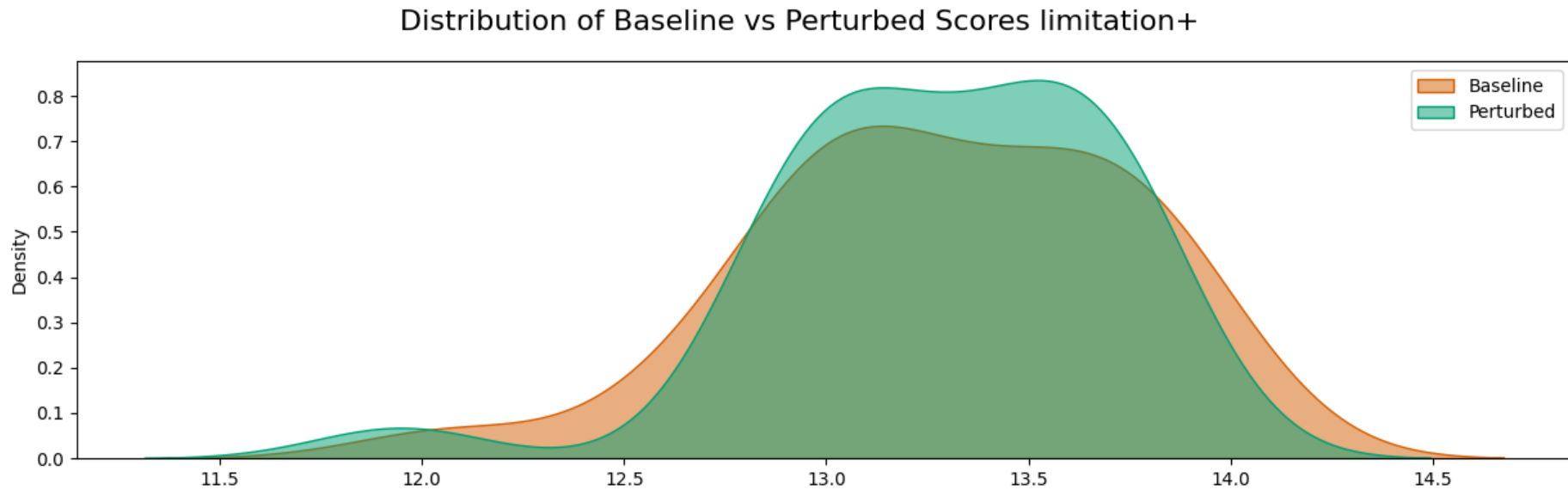
- Apparent dans la question
- Apparent dans 50% des extraits (extraits positifs)

Graphiques:

- distribution des distances similaires
- on retrouve le même noeud
- écart-type assez faible ~> résultats exploitables

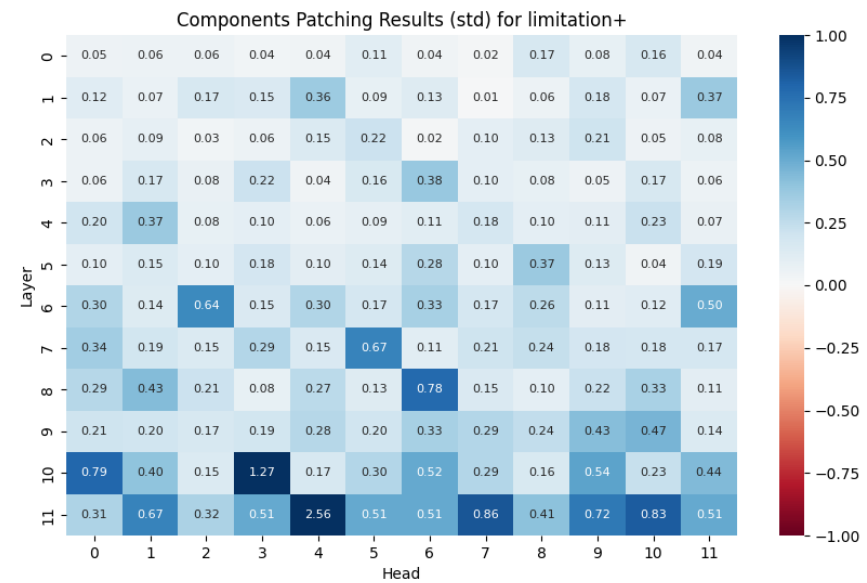
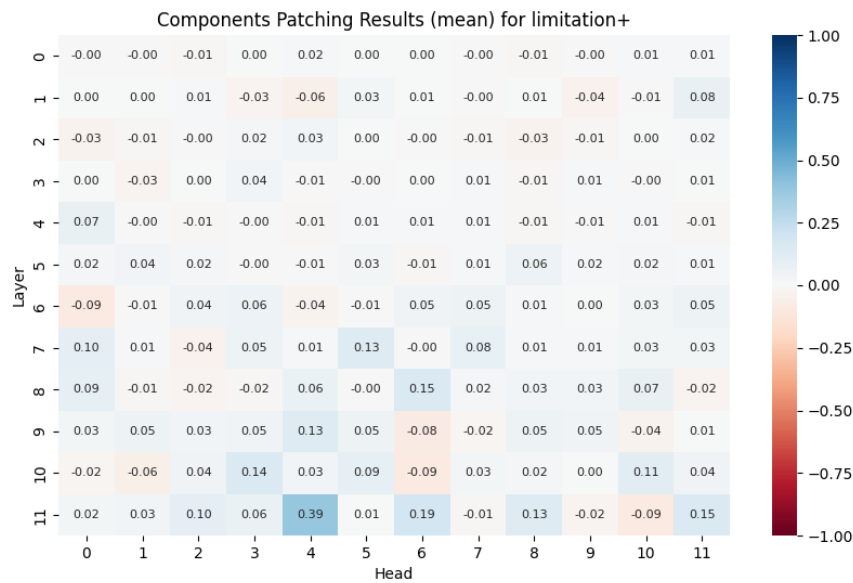
Dataset défense fait main V2 - Résultats

prepend **limitation** - Perturbation Score



Dataset défense fait main V2 - Résultats

prepend **limitation** - Component Patching Results



Dataset défense fait main V2 - Résultats

prepend **limitation** - Analyse

Vis-a-vis des données:

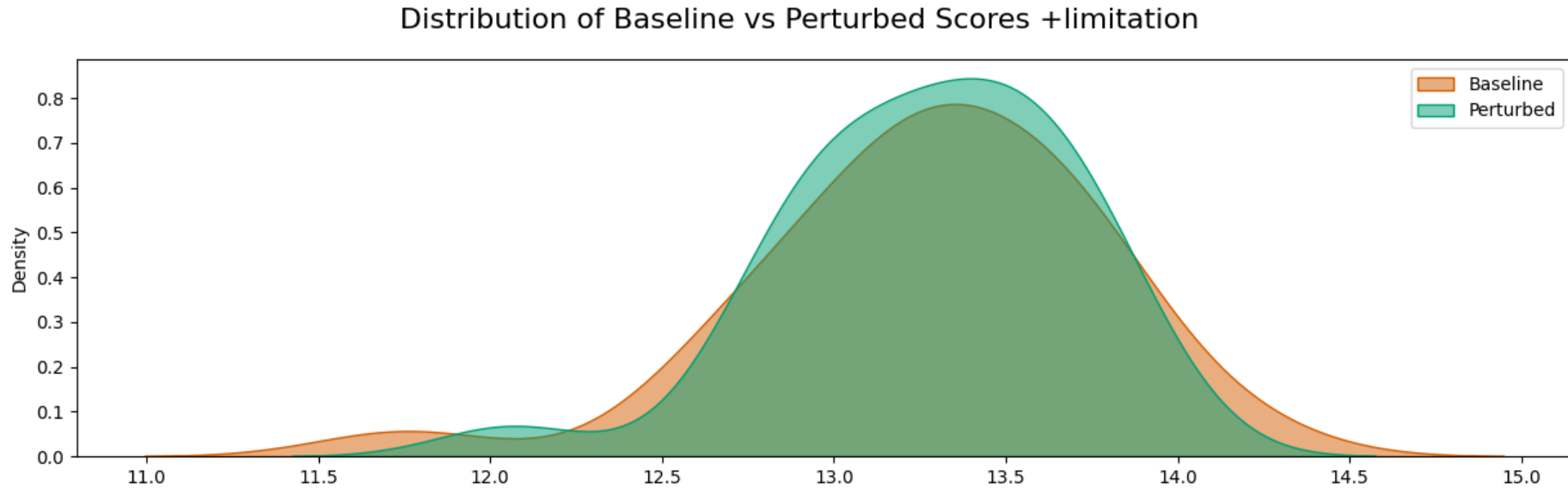
- Présent dans la question
- Présent dans les extraits (~50%)
- 2^e plus grand TF-IDF sur les extraits

Graphique:

- 1 noeud intéressant à interpréter
- des écart-types parfois élevés

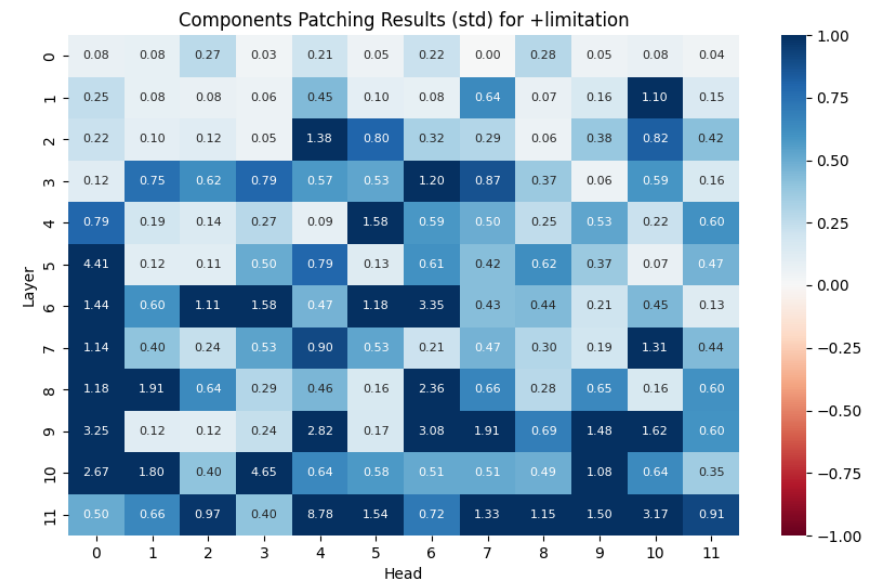
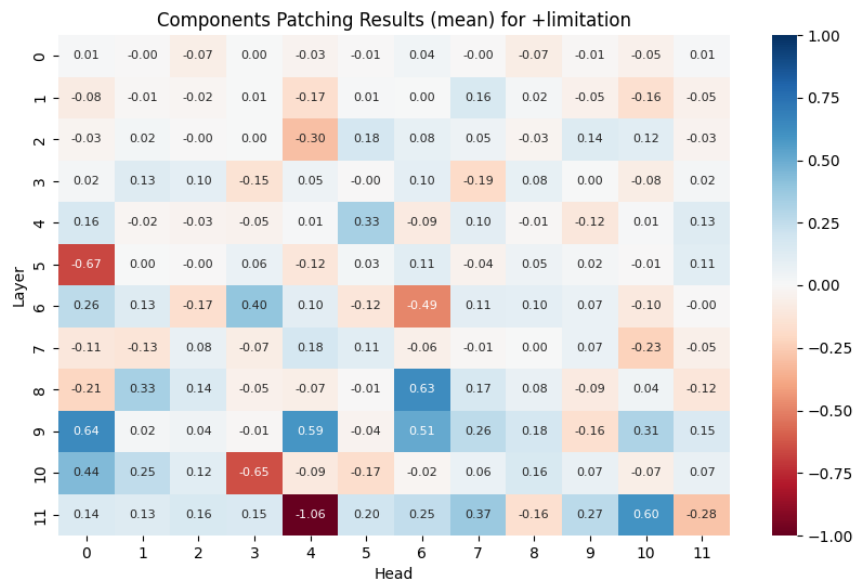
Dataset défense fait main V2 - Résultats

append **limitation** - Perturbation Score



Dataset défense fait main V2 - Résultats

append **limitation** - Component Patching Results



Dataset défense fait main V2 - Résultats

append **limitation** - Analyse

Vis-a-vis des données:

- Présent dans la question
- Présent dans les extraits (~50%)
- 2^e plus grand TF-IDF sur les extraits

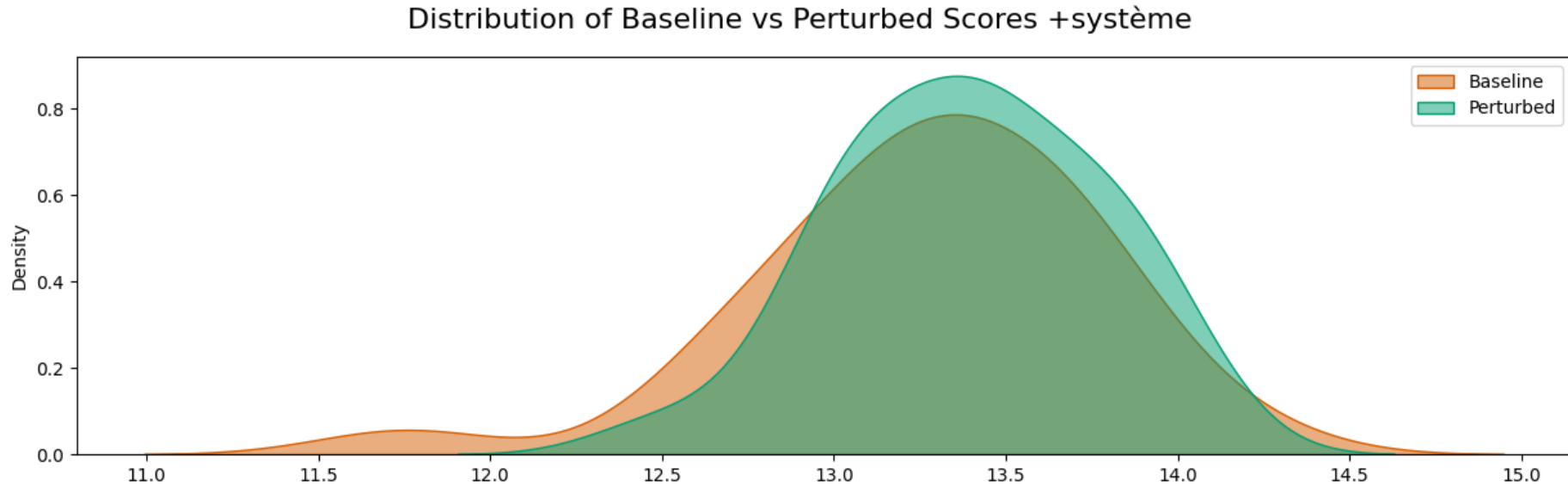
Graphique:

- plus de noeuds réactifs
- un noeud se dégage (même que pour prepend)
- MAIS des écarts-types plus importants !! ~> exploitable?

⇒ Même noeud que pour *drone*

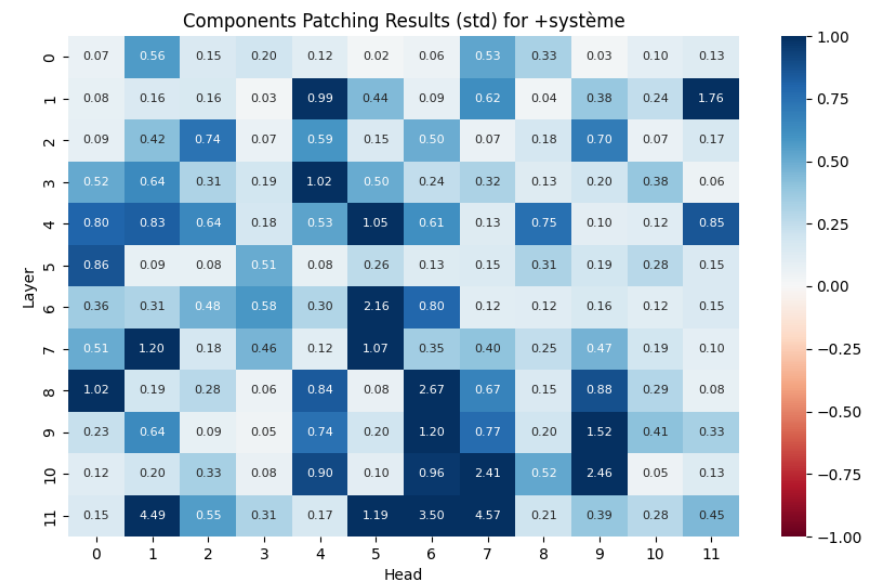
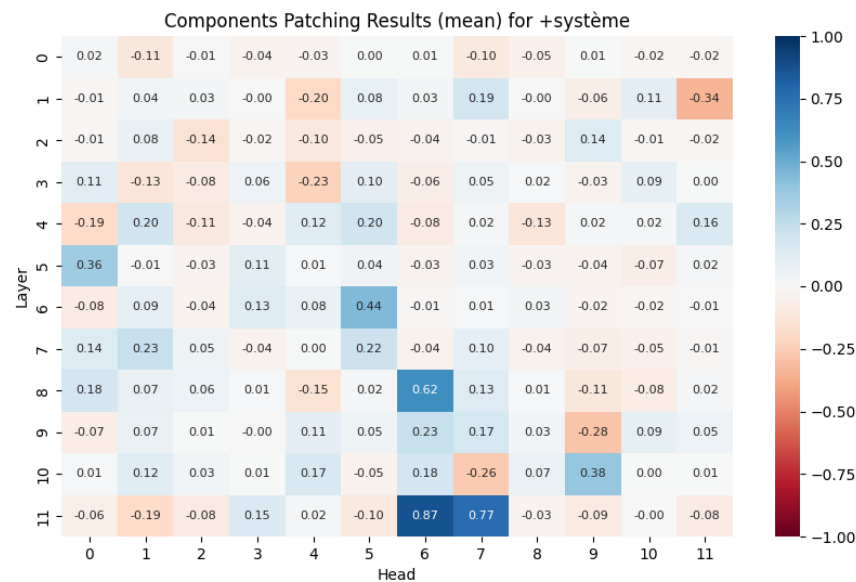
Dataset défense fait main V2 - Résultats

append système - Perturbation Score



Dataset défense fait main V2 - Résultats

append système - Component Patching Results



Dataset défense fait main V2 - Résultats

append système - Analyse

Vis-à-vis des données:

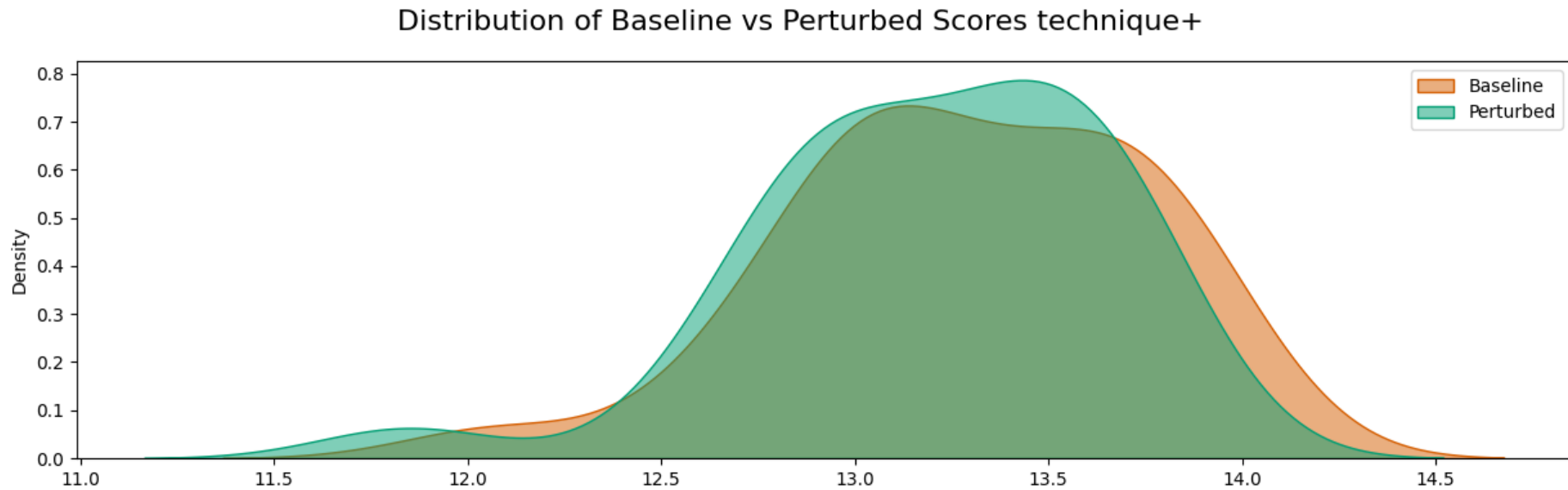
- Présent dans la question
- TF-IDF élevé

Graphique:

- perturbation qui concentre les extraits
- grand nombre de noeuds modérément sensibles
- mais grands écrans-types pour une majorité de noeuds (souvent plus grand que la moyenne) ~> Est-ce vraiment exploitable?

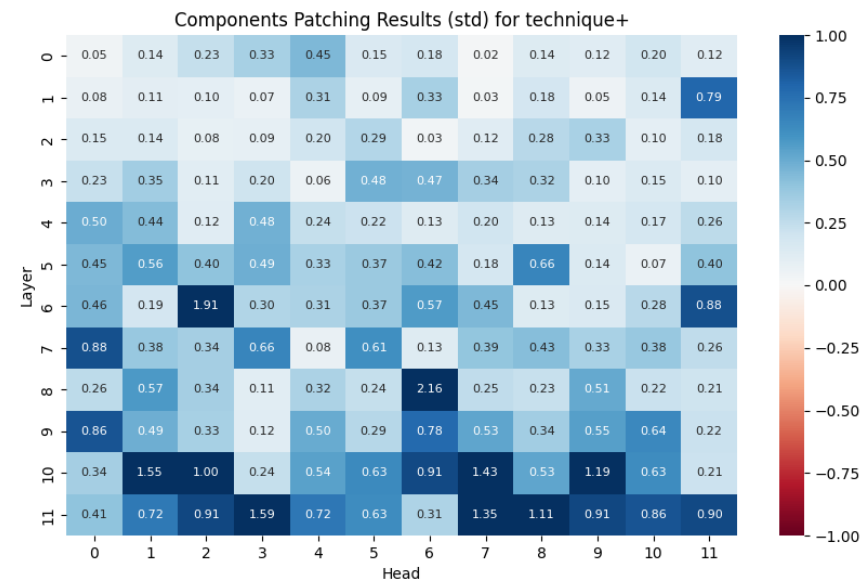
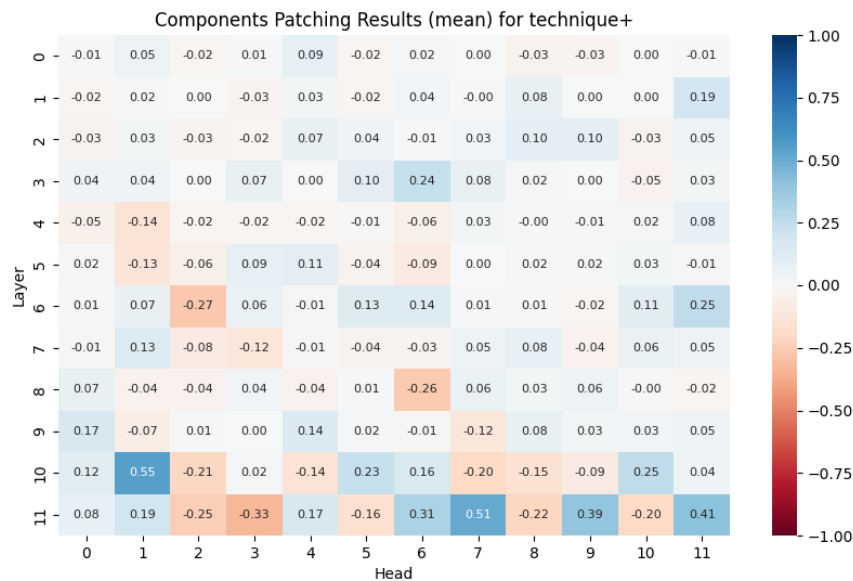
Dataset défense fait main V2 - Résultats

prepend **technique** - Perturbation Score



Dataset défense fait main V2 - Résultats

prepend technique - Component Patching Results



Dataset défense fait main V2 - Résultats

prepend **technique** - Analyse

- présent dans la question
- TF-IDF très faible (127^e, 1 occ, TF-IDF = 0.0074)
- plusieurs noeuds modérément sensibles (concentrés en fin de modèle)
- MAIS écarts-types assez importants ~> peut-être pas exploitable

Dataset défense fait main V2 - Résultats

Analyse générale

- Pas de perturbation `replace` suffisamment importante
- Certaines perturbations semblent chaotiques
- Noeuds sensibles concentrés à une extrémité du modèle
- Cartographie des noeuds parfois similaire pour *append* et *prepend* (cas avec écarts-types faibles)

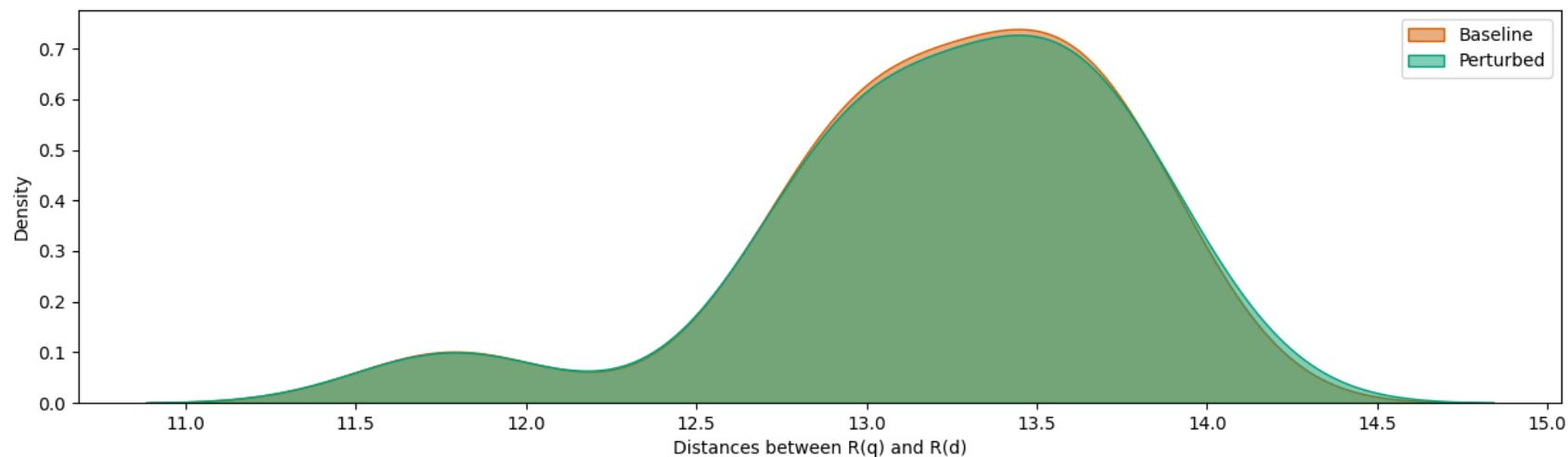
⇒ Le choix du dataset doit-il être un levier pour la cartographie?

⇒ Les ordres de grandeur sont-ils importants pour l'interprétation?

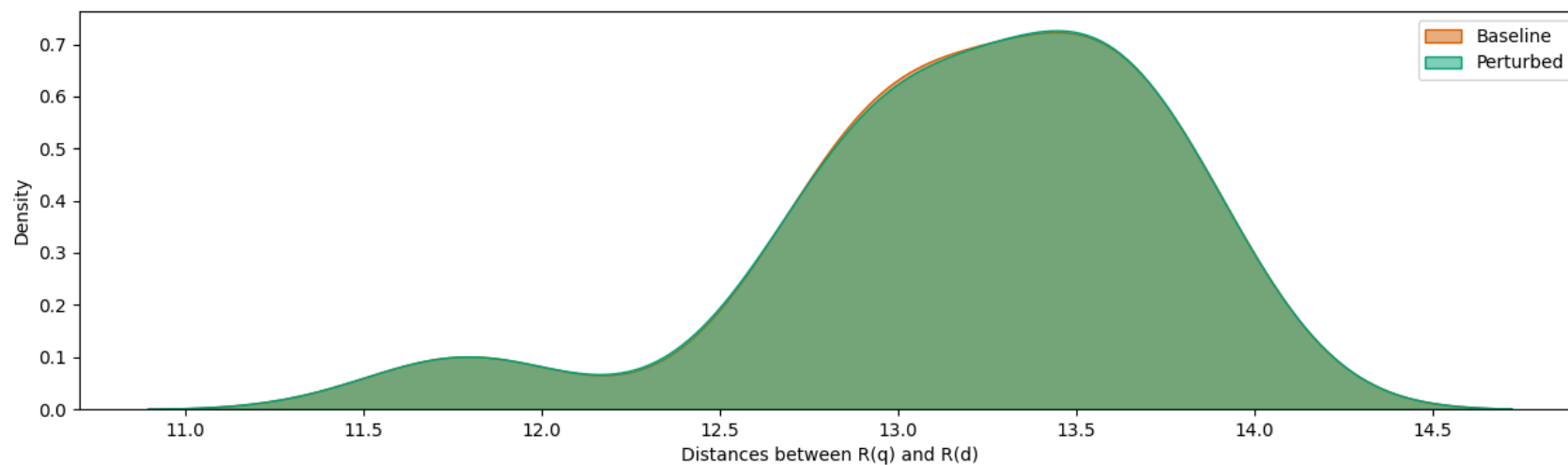
⇒ Remplacer les NaN par des 0 ?

Perturbation replace *Elephant rose* sur défense main V2

Distribution of Baseline vs Perturbed Scores article -> table

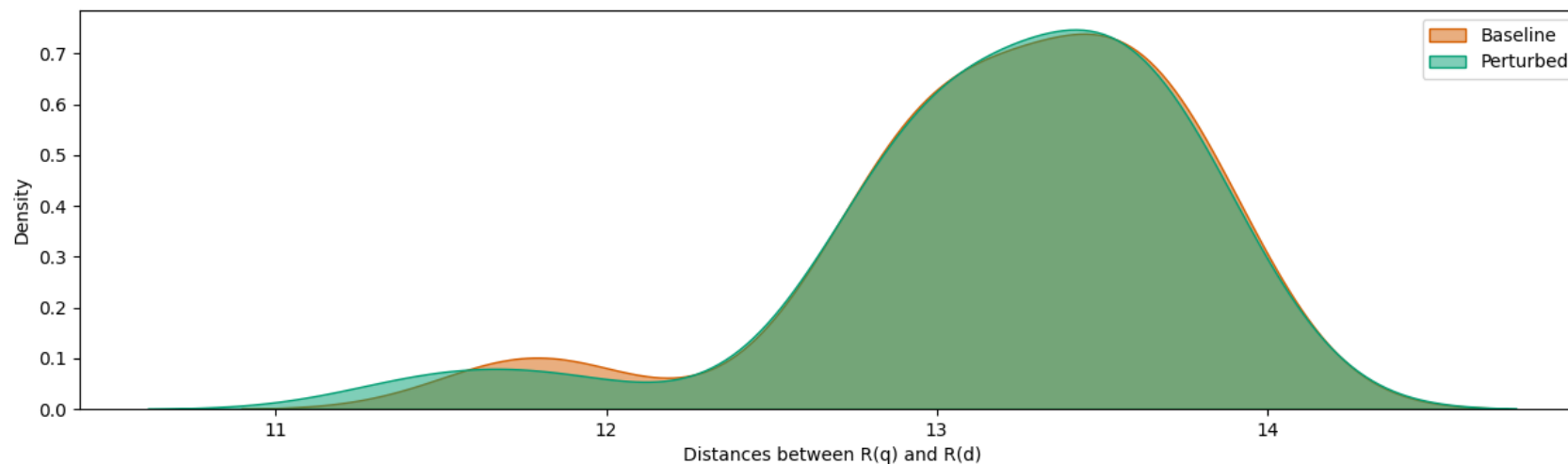


Distribution of Baseline vs Perturbed Scores défense -> fromage

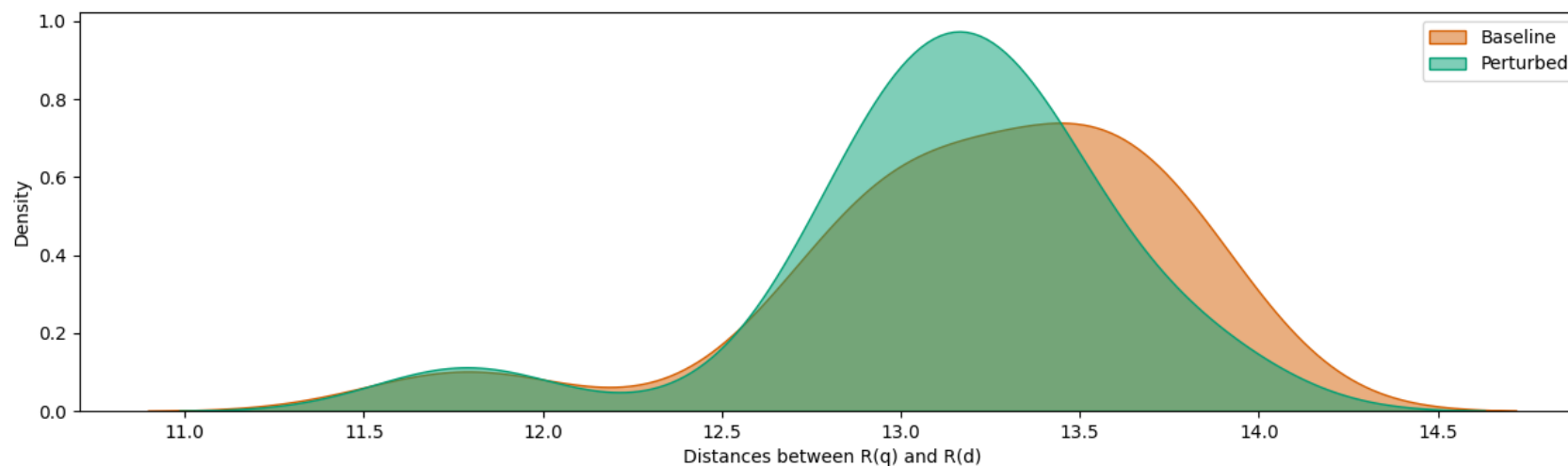


Perturbation replace *Elephant rose* sur défense main V2

Distribution of Baseline vs Perturbed Scores limitation -> animation

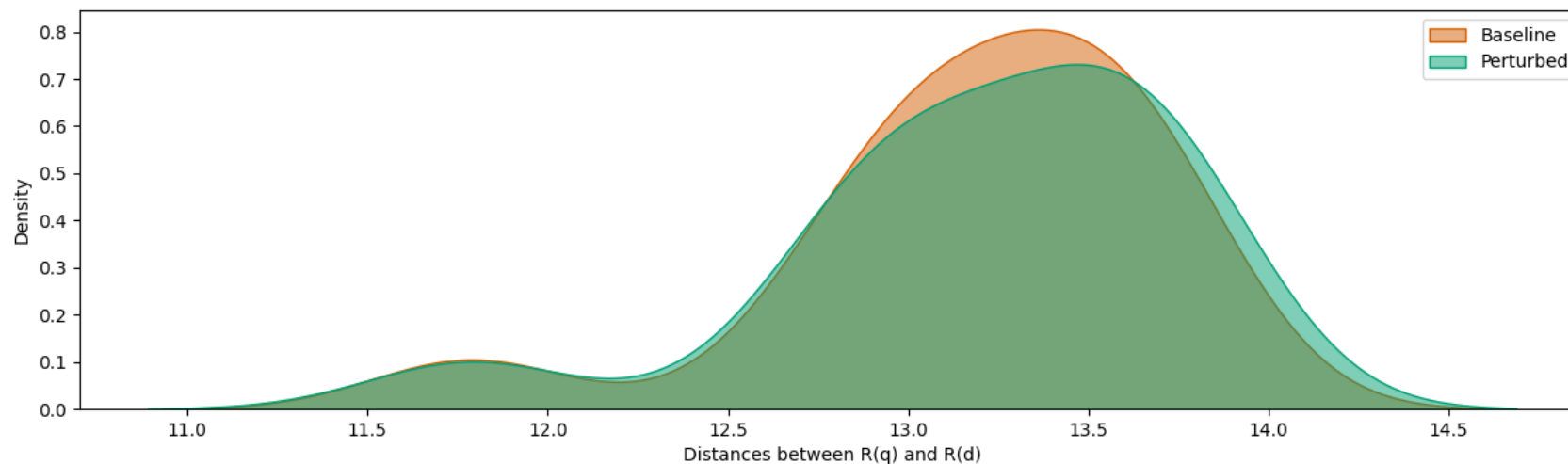


Distribution of Baseline vs Perturbed Scores drone -> gazon

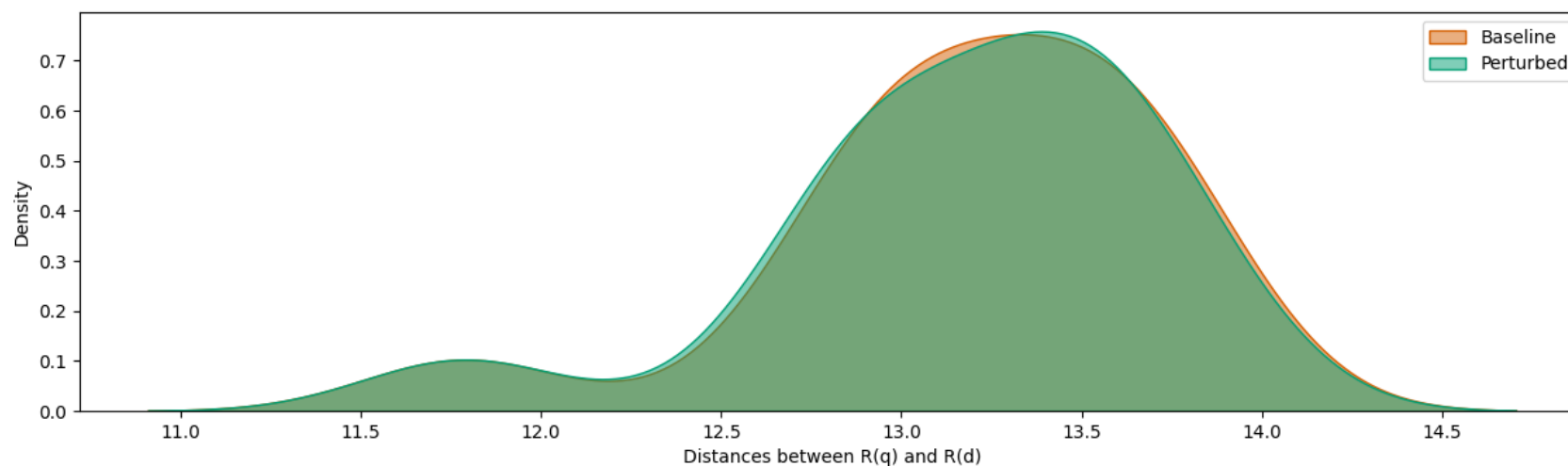


Perturbation replace *Elephant rose* sur défense main V2

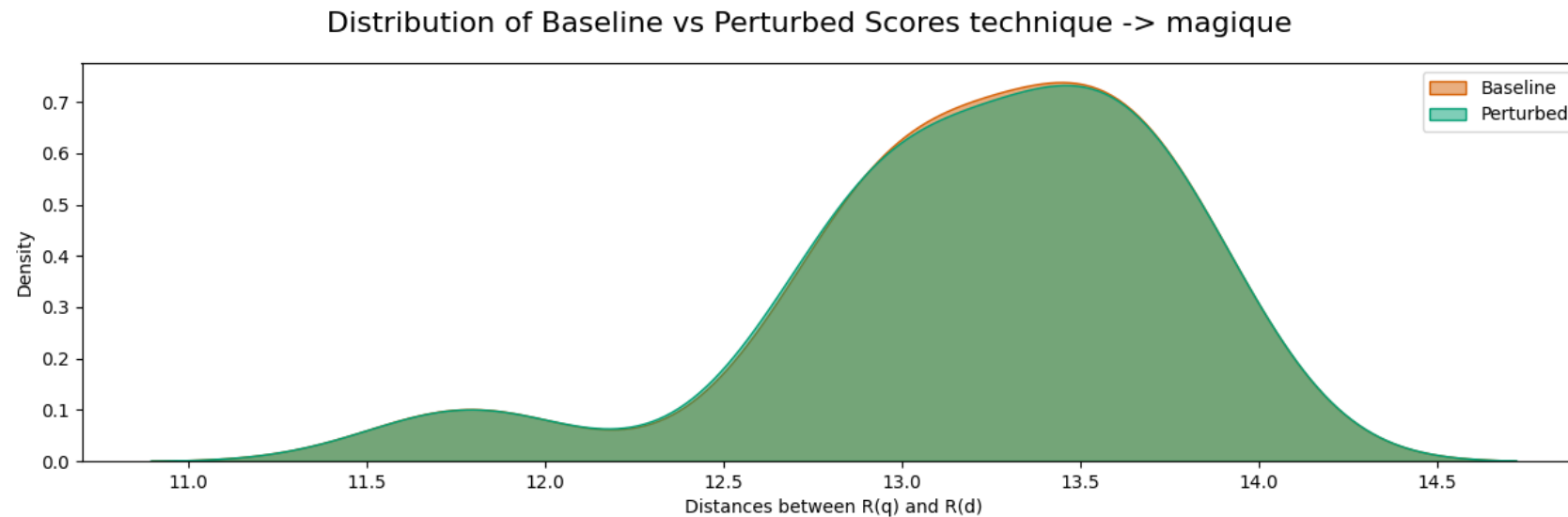
Distribution of Baseline vs Perturbed Scores matériel -> post-it



Distribution of Baseline vs Perturbed Scores système -> panneau



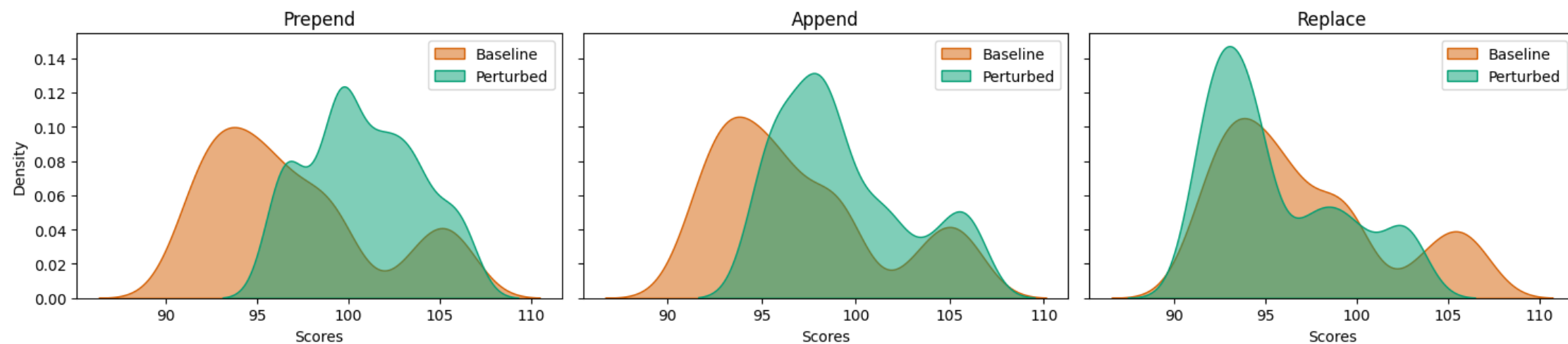
Perturbation replace *Elephant rose* sur défense main V2



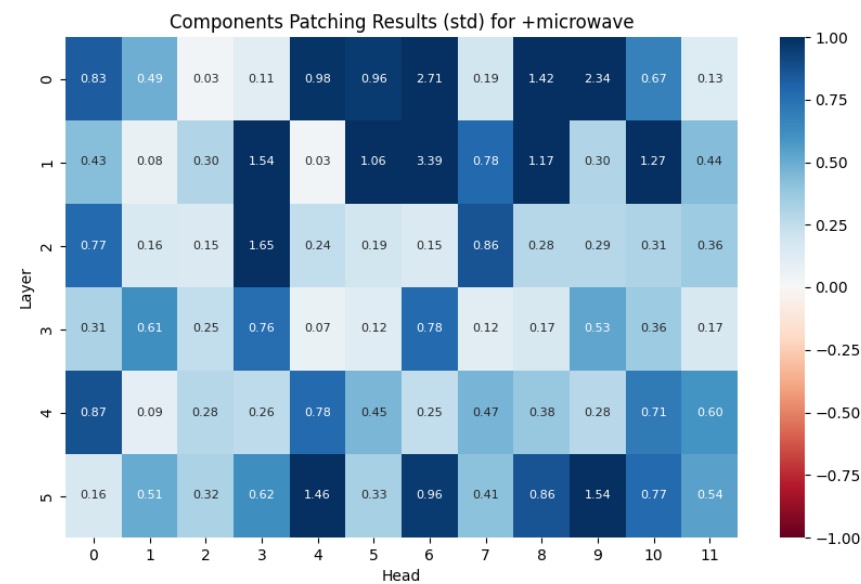
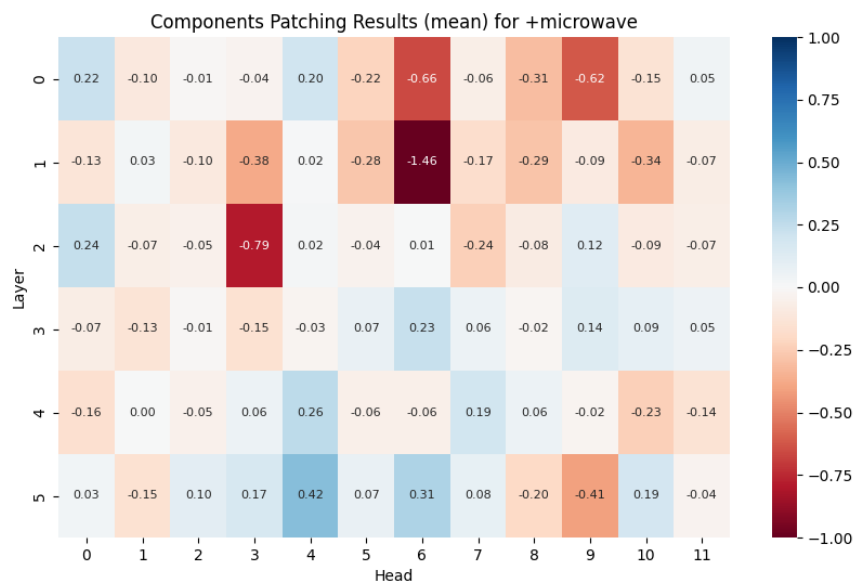
- Pas de perturbation suffisante parmi celles testées
- $\exists$? une façon de choisir les *éléphants roses* ?

Pour rappel Vaswani

Distribution of Baseline vs Perturbed Scores



Pour rappel Vaswani



Pour rappel Vaswani

