

Réu 23 mars 2026

- Lexique
 - [Parry et al., 2025] MechIR: A Mechanistic Interpretability Framework for Information Retrieval
 - Premières mises en pratique avec MechIR
 -
 - Administratif
-

Lexique

Création d'un document ([ragright/biblio/Lexique.md](#)) pour compiler les différentes définitions et les publications associés

Continual vs Continuous learning : "Continuous learning" n'existe pas

Citation de définition des QA et variantes

[Parry et al., 2025] MechIR: A Mechanistic Interpretability Framework for Information Retrieval

Activation Patching : Observer la réaction d'un composant à une caractéristique d'une entrée

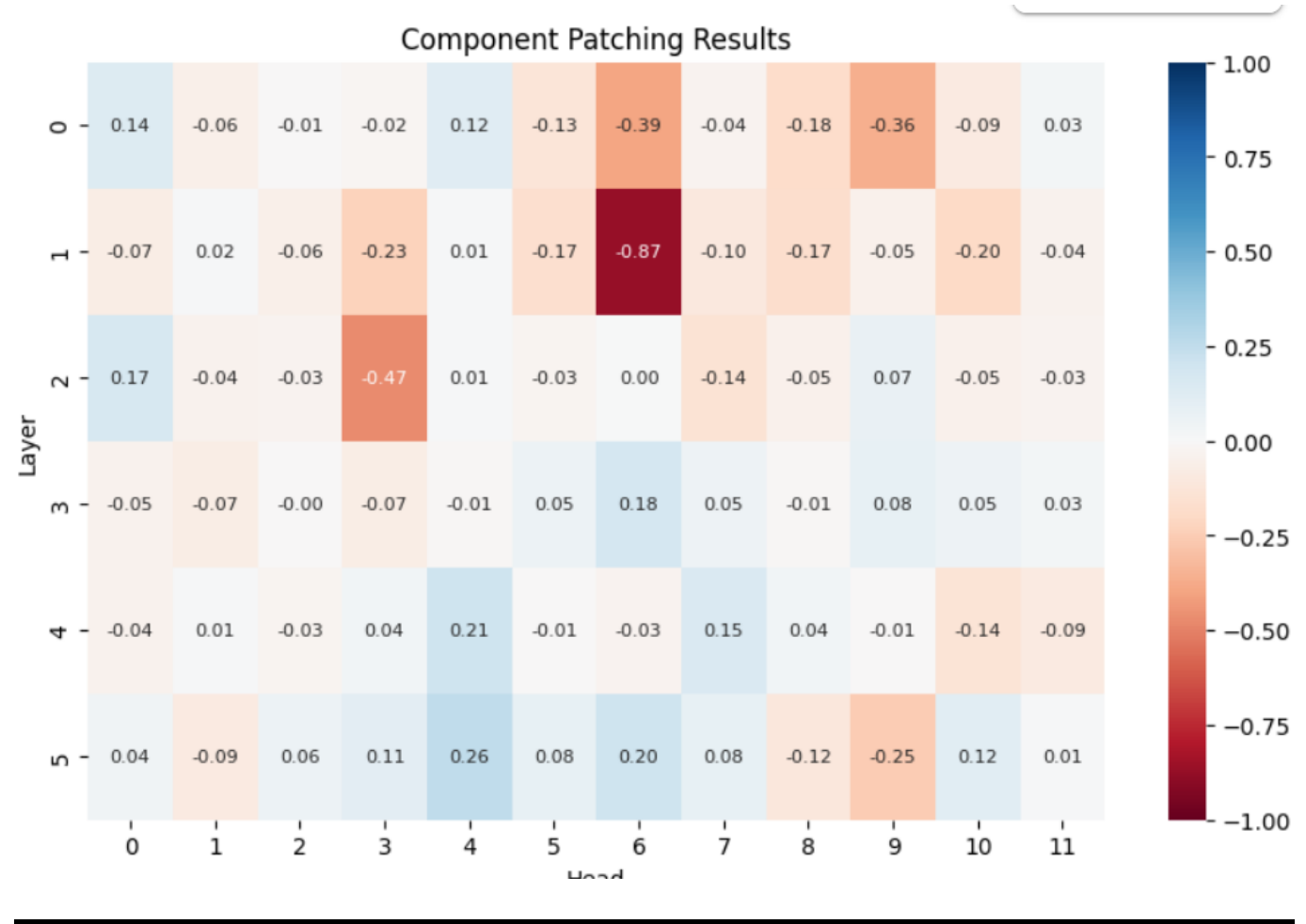
- Relevé des valeurs intermédiaires du modèle sur l'entrée $\$i\$$
 - Relevé des valeurs intermédiaires du modèle sur l'entrée perturbée $\$i'\$$
 - Fixer les valeurs du/des noeuds observés avec celles observées (meilleure perf)
 - Observer les output sur l'entrée fournie (moins bonne perf)
-

[Parry et al., 2025] MechIR: A Mechanistic Interpretability Framework for Information Retrieval

MechIR : Propose une implem pour les modèles encoder-based

Difficultés :

- Lourd en calcul
- Problèmes d'installation



Premières mises en pratique avec MechIR

TAS-B [[Hofstätter et al., 2021](#)]

Contriever:

- Des parties de l'architecture ne sont pas supportées

DPR : Retriever utilisé dans l'archi de [[Lewis et al., 2020](#)] :

- code à executer/tester

Administratif

- Lettre de recommandation pour une [Thèse à Lannion](#)
- Lettre de recommandation pour Prépa Agreg Rennes