

UNIVERSITE D'ÉVRY PARIS SACLAY  
M2 COMPUTER & NETWORK SYSTEMS  
SYSTÈMES AUTONOMES

# Projet R&D

---

AMÉLIORER LES ENVIRONNEMENTS  
ÉDUCATIFS : APPROCHE BASÉE SUR LES LLMS

---

**Présenté par :**

BIAKO Moïse

FERNANDES DIAS Ricardo

**Encadré par :**

Mr. Massinissa HAMIDI



**Année Universitaire 2024-2025**

# Table des matières

<b>1</b>	<b>Abstract</b>	<b>3</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
<b>3</b>	<b>Etat de l’art</b>	<b>4</b>
3.1	Taxonomie : IA Générative et Éducation . . . . .	4
<b>4</b>	<b>Représentation mentale</b>	<b>7</b>
4.1	L’abstraction : Organiser les connaissances en catégories . . . . .	8
4.2	Les concepts . . . . .	9
4.3	Les graphes de concepts . . . . .	10
4.4	Capturer les relations dans les graphes . . . . .	11
4.5	Manipuler les graphes . . . . .	11
4.6	Les prérequis : Structurer les étapes de l’apprentissage . . . . .	12
<b>5</b>	<b>Alignement de la représentation</b>	<b>13</b>
5.1	LLMs et Graphes . . . . .	14
5.2	Alignement avec les graphes . . . . .	14
5.3	Mesurer l’alignement . . . . .	15
5.4	Performance de l’alignement . . . . .	15
<b>6</b>	<b>Conclusion</b>	<b>16</b>
<b>7</b>	<b>Acronymes</b>	<b>21</b>

# 1 Abstract

Dans le domaine de l'éducation, la diversité des capacités des élèves crée des défis uniques en termes de compréhension et de concentration. Ce projet explore l'utilisation des grands modèles de langage (LLMs) pour améliorer l'expérience éducative des étudiants en proposant des contenus pédagogiques adaptés à leurs besoins. L'objectif est de mettre en œuvre un environnement intelligent efficace qui favorise la personnalisation et la création d'un parcours pédagogique sur mesure pour chaque personne.

L'un des principaux défis de cette approche réside dans la représentation des connaissances cognitives et des prérequis des étudiants et dans la personnalisation du contenu éducatif généré par les LLM.

Pour valider cette approche, nous avons mené une expérimentation sur un échantillon représentatif d'étudiants en générant des QCM basés sur plusieurs matières du cursus. Les résultats obtenus sont satisfaisants et ils ouvrent des perspectives prometteuses pour l'élaboration d'un agent capable d'accompagner les étudiants tout au long de leur cursus en s'adaptant de manière dynamique.

## 2 Introduction

Nous vivons dans une société où les avancées technologiques connaissent de nombreux changements, notamment dans le domaine de l'éducation. Ces avancées offrent de nouvelles opportunités pour améliorer les méthodes d'apprentissage en cours. Les grands modèles de langages sont au cœur de ces transformations.

Les LLMs présentent des limites, notamment dans leur tendance à produire des contenus incohérents ou infondés, les hallucinations dans les environnements éducatifs. [1] Ces modèles rencontrent des problèmes lorsqu'il faut personnaliser du contenu en fonction du besoin spécifique d'un étudiant. [2] Parmi leurs autres lacunes, les LLMs ne peuvent pas intégrer de nouvelles informations sans réentraînement, ils peuvent perdre en cohérence lorsque les tâches nécessitent une compréhension ou un raisonnement avancé. [3]

Chaque individu possède une représentation mentale unique, développée par ses expériences personnelles et ses connaissances acquises, et chacun a sa propre manière d'organiser l'information. Cette variabilité représente un défi : comment aligner le contenu éducatif généré par les grands modèles de langage avec la représentation spécifique de chaque étudiant ? Et comment personnaliser le contenu qui est généré par la suite ?

La représentation mentale a été étudiée de différentes manières dans des domaines dont la psychologie cognitive comme illustré dans l'article de la Revue internationale de psychosociologie [4] . Elle se compose d'une multitude de concepts, mais dans le cadre de notre approche, nous nous concentrerons sur la représentation mentale que nous capturons dans un graphe de concepts (ou graphe de connaissance). Nous ferons ensuite l'alignement entre les représenta-

tions mentales et les contenus éducatifs générés, afin de faciliter un apprentissage collaboratif avec les étudiants et de mettre en œuvre un enseignement personnalisé basé sur les besoins des étudiants et d’enrichir les environnements éducatifs intelligents.

Notre projet vise à développer un LLM capable de capturer une représentation mentale de chacun des étudiants sous la forme d’un graphe de concept, puis de l’exploiter pour encoder les schémas cognitifs, les prérequis maîtrisés, les lacunes, et les préférences d’apprentissage propres à chaque étudiant. À partir de ces informations, le modèle sera en mesure de générer des questions, des exercices et des explications adaptés. L’objectif est donc de concevoir une solution capable d’aligner efficacement les contenus éducatifs avec les besoins spécifiques des étudiants facilitant ainsi leur évolution pendant leur cursus scolaire.

Bien que notre contribution soit modeste, nous espérons néanmoins apporter des éléments pertinents pour l’élaboration d’environnements éducatifs intelligents, capables de personnaliser et d’enrichir l’expérience d’apprentissage de manière significative.

Dans notre état de l’art d’abord nous ferons la taxonomie des approches d’intelligence artificielle générative appliquées à l’éducation, en s’appuyant sur les discussions provenant du NeurIPS’23 Workshop : Generative AI for Education (GAIED). Ensuite, nous analyserons la notion de représentation mentale. Enfin, nous nous concentrerons sur la question de l’alignement.

## 3 Etat de l’art

### 3.1 Taxonomie : IA Générative et Éducation

Dans un premier temps, nous allons nous intéresser à des solutions existantes. Parmi elles, les travaux récents présentés lors du *NeurIPS’23 Workshop : Generative AI for Education (GAIED)* [5] offrent des pistes intéressantes. Ces recherches mettent en avant des approches innovantes utilisant l’intelligence artificielle générative pour concevoir des contenus éducatifs adaptés. Nous nous appuierons sur ces solutions pour orienter notre réflexion et apporter notre propre contribution.

Nous allons aborder les principales solutions basées sur l’IA générative présentées lors de la conférence NeurIPS’23, en mettant en avant celles qui convergent avec notre objectif dans les environnements éducatifs. À partir d’articles sélectionnés pour leurs pertinences, nous avons identifié plusieurs thématiques principales : la personnalisation du contenu, l’évaluation des connaissances, la génération de questions, l’amélioration de l’apprentissage et l’organisation des ressources.

Pour faire la taxonomie, nous avons donc étudié les différents articles du NeurIPS’23 Workshop et avons construit une taxonomie qui nous aidera dans notre étude en utilisant les thématiques

que nous avons repéré.

- *La personnalisation du contenu* permet d'adapter les supports d'apprentissage aux besoins individuels des étudiants.

Personalization and Contextualization of Large Language Models for Improving Early Forecasting of Student Performance [2] : Cet article propose l'utilisation des LLM dans le but de prédire les performances des étudiants. Cela se fait en intégrant les données académiques et sociales, ensuite les auteurs nous montrent que la personnalisation et le contexte permettent d'améliorer la précision des prévisions.

Ruffle Ripley : Towards the Automated Induction of Conversational Tutoring Systems [6] : Cet article nous présente le système de tutorat Ruffle&Ripley qui utilise GPT-4 pour générer des scripts pédagogiques à partir de cours. Nous nous sommes dit que cela améliore l'expérience d'apprentissage et que cela a été testé dans une étude utilisateur.

- *L'évaluation des connaissances* vise à modéliser le niveau de chaque étudiant ,à détecter leurs lacunes et identifier leurs besoins

Generative Agent for Teacher Training : Designing Educational Problem-Solving Simulations [7] : Cet article nous montre une approche pour la formation des enseignants et utilisant des agents génératifs sur GPT-4 dans des simulations sur Roblox. Ce sont des agents ayant des personnalités qui sont réalistes et qui aident les futurs enseignants à régler des problèmes éducatifs ce qui aide à réduire l'écart entre théorie et pratique. Les résultats de ces tests montrent une bonne immersion.

Angel : A New Generation Tool for Learning Material based Questions and Answers [8] : Cet article nous présente l'outil Angel basé sur l'IA qui génère des questions éducatives adaptées aux différents niveaux, en s'appuyant sur la taxonomie de Bloom. Cet outil améliore la qualité et la diversité des questions.

- *La génération de questions* permet de proposer des exercices sur mesures pour l'accompagnement et avoir des retours pertinents.

Automated Distractor and Feedback Generation for Math Multiple-choice Questions via In-context Learning [9] : Cet article nous propose l'utilisation de LLM (GPT-3.5, GPT-4) pour générer exprès de mauvaises réponses et des explications qui sont adaptées dans des questions à choix multiple en maths. Tout ça dans le but de mieux comprendre les erreurs des élèves.

Small Generative Language Models for Educational Question Generation [10] : Cet article compare des petits modèles de langage à des LLM comme ChatGPT pour générer des questions éducatives. Les résultats de cette comparaison montrent que les petits modèles de langage génèrent

des questions presque autant efficaces avec un coût moindre.

Evaluating ChatGPT-generated Textbook Questions using IRT [11] : Dans cet article les auteurs évaluent la capacité de génération de questions de ChatGPT à partir de manuels scolaires résumés. Les résultats montrent une similarité entre les questions générées et celles des manuels en termes de difficultés et sur la capacité de différenciation des compétences des étudiants. Cela montre que des systèmes de tutorat automatisés sont possibles.

- *L'amélioration de l'apprentissage* favorise les interactions avec l'étudiant plus immersives et engageantes.

Beyond Hallucination : Building a Reliable Question Answering & Explanation System with GPTs [1] : Cet article un système qui utilise GPT-4 afin de répondre aux questions et de donner des explications détaillées. Ce système grâce à des explications étape par étape et à des références améliore la satisfaction de l'utilisateur.

- *L'organisation des ressources* facilite la conception des cours pour l'enseignant et l'étudiant.

Field experiences and reflections on using LLMs to generate comprehensive lecture metadata [12] : Cet article nous décrit l'utilisation de GPT-4 pour générer des métadonnées éducatives comme de résumés et des questions à partir de vidéos de cours en ligne dans le but de personnaliser l'apprentissage et d'apporter un soutien aux enseignants.

Nous commencerons par présenter la représentation mentale qui est le socle sur lequel se base notre projet, puis nous présenterons les graphes de connaissances qui permettent de structurer la représentation mentale d'un individu.

## 4 Représentation mentale

*“ La cognition peut être définie par l’ensemble des activités qui découlent du fonctionnement cérébral chez l’homme et chez l’animal : sensori-motricité, perception, langage, apprentissage, mémoire, représentation des connaissances, décision et raisonnement, planification, coordination motrice... ” [13]*

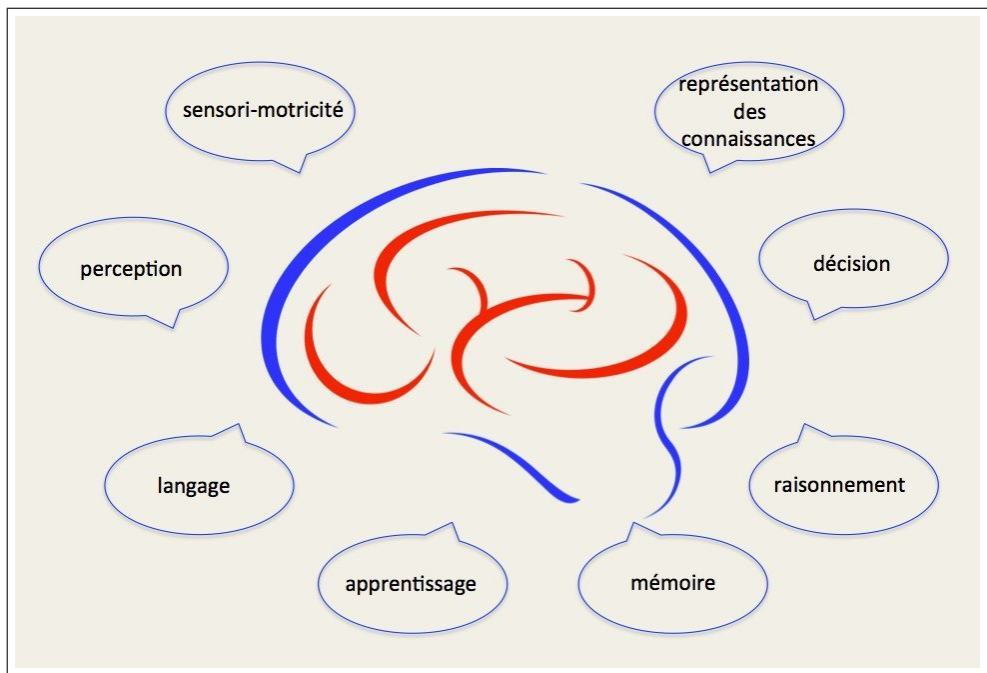


Figure 1 : Domaines d’études impliquant la cognition [13]

Notre objectif est de capturer la représentation mentale d’un étudiant puis de la traduire sous la forme d’un graphe pour ensuite aligner le LLM, nous allons donc nous concentrer sur le concept de représentation mentale uniquement en omettant les autres (mémoire, langage, apprentissage etc).

La représentation mentale est un des concepts centraux de notre approche. Elle est définie de différentes manières :

- *“les représentations mentales sont définies comme des constructions intellectuelles produites par l’entendement, qui permettent à un individu de se rapporter à des objets, des idées, des événements ou des expériences” [14]*
- *“objet cognitif qui "rend présent" un élément absent de la perception directe, en permettant de traiter et d’organiser les connaissances” [15]*

Chaque individu possède sa propre représentation mentale qui est façonnée par ses multiples expériences, cela influence directement la manière dont il perçoit et assimile le contenu éducatif. [16]

Pour pouvoir aligner un LLM avec les besoins des étudiants, il est nécessaire de “capturer” cette représentation unique.

Utiliser la représentation mentale dans notre projet a pour but de transformer des concepts abstraits/complexes en quelque chose d’exploitable pour le LLM. En le traduisant sous forme de graphe de concepts/connaissances on pourra représenter les connaissances de l’étudiant, leurs relations, et leurs dépendances . Cela constitue une base solide pour personnaliser les contenus éducatifs et proposer des exercices adaptés.

Pour aller plus loin, nous proposons dans la prochaine partie de montrer comment les connaissances sont regroupées et organisées. L’abstraction a un rôle central dans cette approche car elle va permettre la classification des connaissances en catégorie pour créer les graphes de concepts.

## 4.1 L’abstraction : Organiser les connaissances en catégories

Les humains ont une mémoire limitée mais sont capables de prendre des décisions efficaces dans des environnements complexes. Cette capacité est basée sur l’abstraction qui consiste à simplifier des problèmes en éliminant les détails inutiles tout en conservant les informations les plus utiles. Par exemple, au cours d’une randonnée, les éléments qui composent l’environnement sont très nombreux (pierres, arbres, nuages etc), une indication du style “il faut aller à l’Ouest de la rivière” est suffisante pour nous guider dans la forêt plutôt que de chaque pas ou arbre rencontré avant d’arriver dans le lieu souhaité. [17]

L’*agent* (personne ou LLM) va se concentrer sur l’aspect le plus significatif de l’environnement. L’abstraction compresse les informations en réduisant les détails les plus superflus pour faciliter l’apprentissage.

Chez un individu, le développement de la pensée se fait progressivement de l’enfance à l’âge adulte. Les capacités cognitives se construisent par étapes, l’individu apprend à manipuler des objets physiques puis des concepts abstraits, Jean Piaget le décrit très bien. [18]

Les concepts se forment avec l’expérience, un enfant aura tendance à dire de tous les véhicules qu’il aperçoit “c’est une voiture”. Pour acquérir les concepts, il aura besoin d’interactions extérieures (un enseignant par exemple), il pourra ainsi distinguer le camion du bus.

Appliqués à un graphe, l’abstraction va nous permettre d’éliminer les détails superflus et de suivre une hiérarchie progressive allant de notions concrètes vers des notions plus abstraites. [19] Nous pourrions enrichir via retours humains pour mieux personnaliser l’apprentissage.

Nous proposons à présent une rapide présentation de la notion de concepts afin d’introduire les graphes de connaissances ou concepts et cela dans le but que les plus néophytes d’entre vous aient un niveau de compréhension minimal.



## 4.2 Les concepts

Un objet est une entité du monde réel. Un concept est une catégorie d'objets ayant des propriétés en communs. Un objet est donc une instance de ce concept. Les propriétés en communs des objets d'un concept sont représentés par des attributs : *“Les attributs permettent par exemple de modéliser que toute instance du concept voiture a une couleur, roule à une certaine vitesse, et possède 4 roues. On dira alors que le concept voiture a comme attributs couleur, vitesse, et roue, qui sont eux même des concepts.”* [20] Nous nous intéressons à la façon dont les concepts sont organisés, il faut donc utiliser une relation de spécialisation pour hiérarchiser les concepts “les voitures sont une sous catégorie des véhicules”. [20]

Pour acquérir un concept, il y a 3 critères d'évaluation qui sont représentés dans le tableau ci-dessous :

3 niveaux	Reproduction	Abstraction	Généralisation, transfert
Définitions	Savoir reconnaître le concept et le nommer	Savoir justifier cette reconnaissance en nommant les attributs essentiels	Savoir générer ses propres exemples du concept en les justifiant
Exemples	C'est un carré.	C'est un carré parce que ses côtés sont de même longueur et il a quatre angles droits.	Tracer un carré puis expliquer pourquoi c'est un carré.

Figure 2 : Critères d'évaluation [21]

Nous nous baserons sur ces critères afin d'établir les QCM nécessaires à l'évaluation du niveau de l'élève et établir sa représentation mentale.

### 4.3 Les graphes de concepts

Par définition “Un graphe conceptuel est un graphe biparti connexe orienté. Les deux types de nœuds sont les concepts et les relations. Un arc relie toujours 2 nœuds de type différent”. [22]

L'idée des graphes conceptuels est de représenter les connaissances par des nœuds et les relations par des flèches. Il existe 2 types de nœuds : [22]

- nœud conceptuel : un rectangle représente une entité
- nœud relationnel : une ellipse représente une propriété ou une relation

On identifie le type d'un concept par un identificateur. Un concept est concret si on arrive à former une image mentale comme chien, maison, etc , et il est abstrait sinon comme l'amour, la réussite etc. [22]

Il est possible de représenter sous forme textuelle ou linéaire mais cette version ne nous intéresse pas.

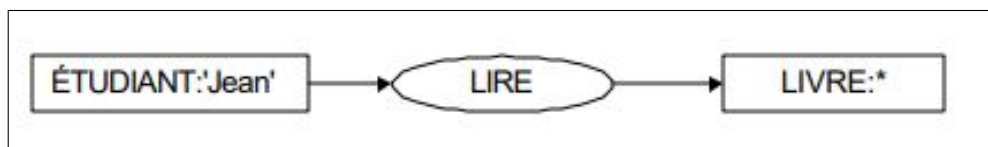


Figure 3 : Schéma d'une relation entre deux concepts [22]

La hiérarchie des types , c'est organiser des concepts par niveaux, où chaque niveau représente un certain degré de complexité. Elle permet de structurer les connaissances en faisant des relations en fonction de l'inclusion ou la dépendance des concepts représentés.

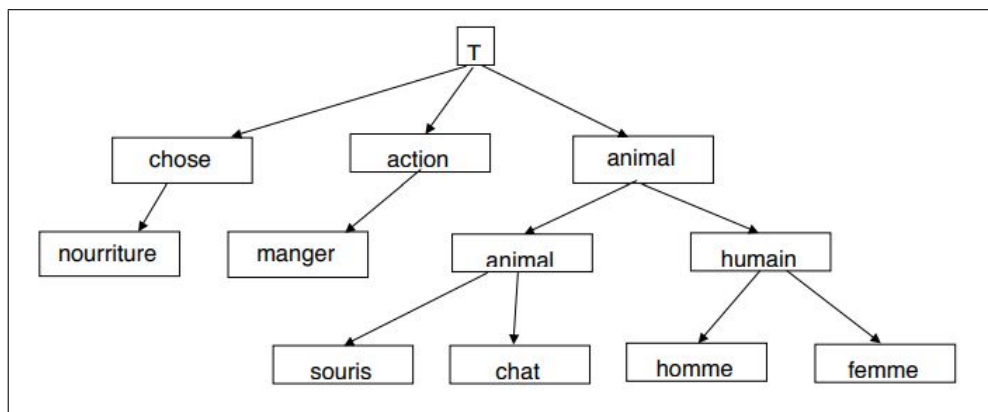


Figure 4 : Représentation d'une hiérarchie des types [23]

## 4.4 Capturer les relations dans les graphes

Les graphes organisent les concepts et leurs relations offrant ainsi une vision structurée des connaissances. Cela permet de préciser les connaissances déjà acquises par les étudiants, les apprentissages à effectuer et l'intégration des nouvelles notions dans leur parcours.

En incluant dans les graphes les relations de prérequis, d'abstraction et d'inférence, on peut identifier :

- Les points faibles d'un étudiant en suivant les éléments qu'il n'a pas encore assimilés.
- Les contenus sont ajustés en fonction de son niveau actuel en se basant sur les prérequis.
- Les parcours sur mesure, en créant des chemins précis dans le graphe afin de guider l'étudiant à travers des concepts structurés en fonction de son avancement.

## 4.5 Manipuler les graphes

L'approche la plus évidente est l'approche manuelle qui consiste à faire à la main le graphe cible. L'outil d'édition manuelle de KG le plus célèbre est Protégé. [24]

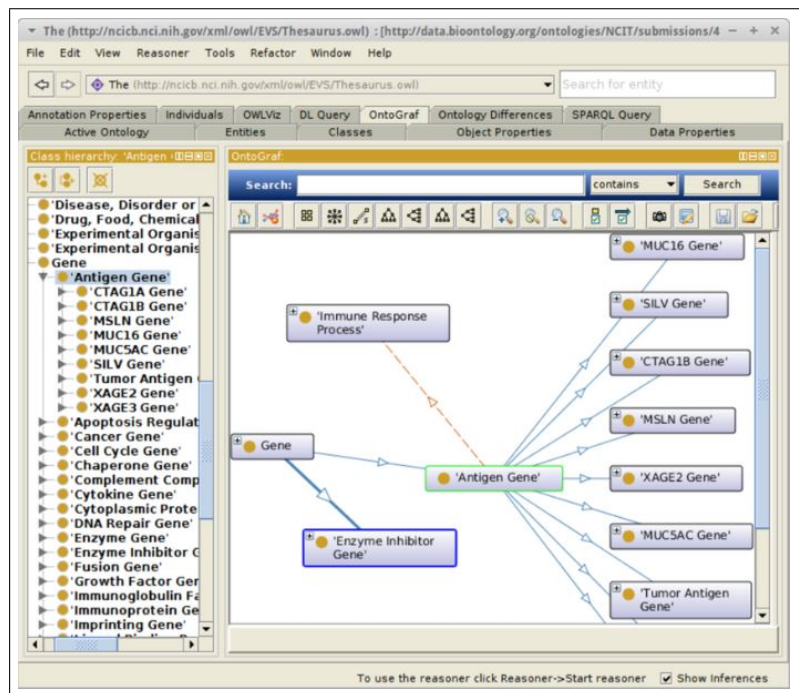


Figure 5 : Outil Protégé [24]

Logiciel développé par l'Université de Stanford depuis les années 1990 et spécialisé pour la création et l'édition d'ontologies. Aujourd'hui, certains graphes de concepts (KGs) massifs tel que Wikidata [25] reposent toujours sur l'édition manuelle par ses utilisateurs, à travers une interface prenant la forme d'un formulaire. D'autres KGs, tels que YAGO [26] ou le Google

Knowledge Graph [27], reposent sur l'extraction massive de faits à partir de pages Web structurées, telles que les infobox de Wikipedia ou les sites de e-commerce. Parmi les outils les plus expérimentaux, on peut citer [28] qui propose une interface d'édition interactive de KG se basant sur un moteur de recommandation.

## 4.6 Les prérequis : Structurer les étapes de l'apprentissage

On entend par prérequis *“toute connaissance ou compétence qui répond simultanément aux deux traits essentiels suivants : elle s'avère, d'une part, cruciale pour la maîtrise d'un cours, d'une discipline, d'un programme ou plus généralement, pour l'affiliation aux études universitaires et elle est, d'autre part, considérée par les enseignants, explicitement ou implicitement, comme devant être acquise préalablement à l'entrée de ces études ou comme devant être acquise au cours de ces études indépendamment d'un enseignement systématique et explicite”* [29]

Le document [30] nous est très utile pour comprendre les relations de prérequis entre concepts éducatifs. Il met en évidence l'importance d'assimiler certains concepts en premier pour faciliter l'assimilation de concepts plus complexes.

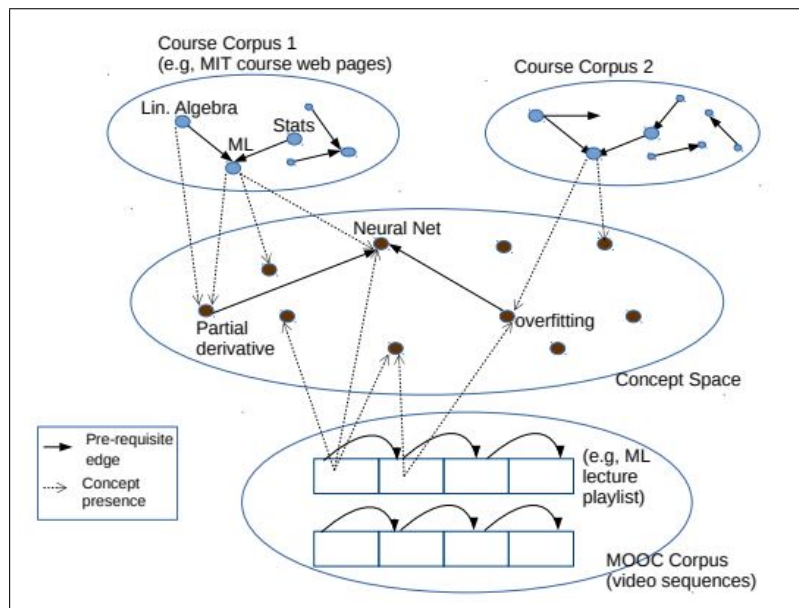


Figure 5 : Graphe de concept basé sur l'approche PREREQ [30]

Le modèle PREREQ s'appuie sur des données éducatives de cours en ligne ou d'universités en représentant les concepts sous forme de graphes orientés. Nous pouvons voir notamment que des prérequis pour le NEURAL NET sont les STATS. Les objectifs du modèle sont clairs : identifier les concepts qui doivent être appris avant d'en aborder d'autres et construire un graphe de relations de prérequis pour les cours ou les sujets.

Dans un graphe, on peut représenter les relations de prérequis en utilisant des arêtes directionnelles entre les concepts. Par exemple, l'équation linéaire serait une condition préalable

à l'équation quadratique. Selon la direction de l'arête, il est essentiel de maîtriser le premier concept avant d'aborder le second.

## 5 Alignement de la représentation

Pour personnaliser le contenu éducatif de façon efficiente, nous devons aligner les LLMs en fonction du besoin de l'étudiant. L'article [31] explore des approches permettant de mesurer cette correspondance. Il est nécessaire d'adopter des méthodes d'entraînements basées sur des feedbacks humains, ce qui nous conduirait à utiliser des prompts personnalisés. Cela améliorera la pertinence des contenus pédagogiques.

L'alignement des LLMs avec les représentations est un sujet très actuel. Il est défini comme *"Aligning large language models (LLMs) involves ensuring that their behaviors are consistent with human values and intentions, thereby mitigating potential risks associated with their deployment."* [32]. Un manque d'alignement entre LLMs et les valeurs humaines (non-discrimination et d'égalité) peut aboutir à des décisions automatisées injustes, affectant l'accès aux droits fondamentaux tels que l'emploi, l'éducation, la santé et la justice. [33] Des LLMs comme Chatgpt que l'on utilise actuellement ont été alignées pour qu'ils soient lisibles par l'humain. Ils doivent être structuré sous forme d'items au lieu de générer du texte brut (que des paragraphes). [34]

L'alignement peut se faire via le *Fine-tuning* (Ajustement fin) [35] dont le plus populaire est le RLHF (Reinforcement Learning with Human Feedback). De nombreuses méthodes pour améliorer le RLHF existent parmi lesquels on trouve le DPO(Direct Preference Optimization) [36], le PPO(Proximal Policy Optimization) [37] et le CPO( Contrastive Preference Optimization) [38]

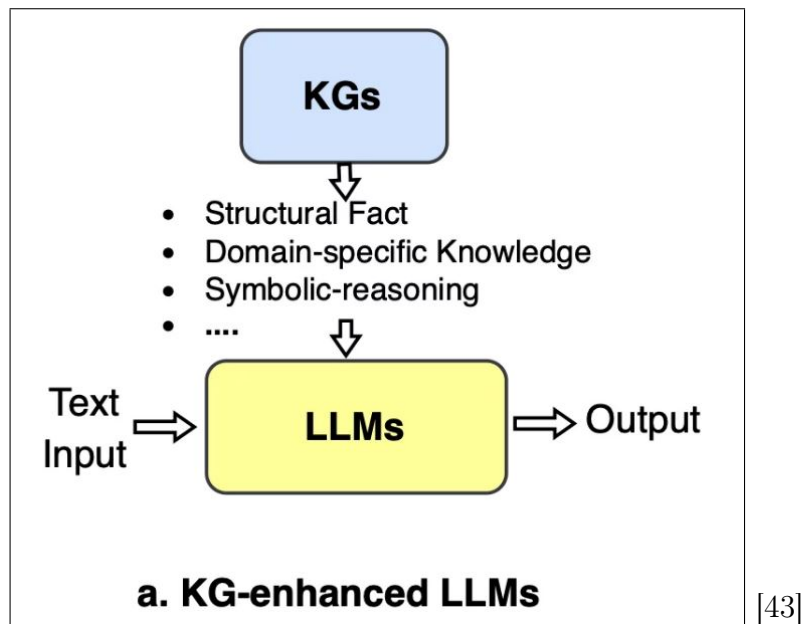
L'autre méthode d'alignement est le *Test Time Alignment* (Inférence) [35] dont la manière la plus à utiliser se base sur des prompts, plusieurs travaux ont été réalisés pour améliorer l'honnêteté des réponses fournies par les LLMs. [39] [40] [41]

Nous préconisons une approche basée sur les prompts pour l'élaboration de notre solution.

## 5.1 LLMs et Graphes

Les LLMs peuvent parfois être victime d'hallucinations. Pour contrer ce phénomène, le graphe de connaissance intervient en encadrant la réponse *“En intégrant dans le prompt un cadre de réponse strict, le graphe de connaissance guide la génération de texte en alignant la réponse sur la question de l'utilisateur, tout en respectant les contraintes du graphe.”* [42]

Les graphes de connaissances donnent une structure et un contexte aux informations tandis que les LLMs apportent une nuance. La combinaison des deux crée une synergie qui permet de répondre à des questions de façon adaptée.



## 5.2 Alignement avec les graphes

Les capacités émergentes des LLMs (vision par ordinateur [44], systèmes de recommandation [45] et reconnaissance vocale [46]) incitent les chercheurs à les appliquer aux graphes afin de développer un modèle pour les graphes qui fonctionne dans différents scénarios. [47] La capacité à utiliser les LLMs à partir de données fournies par les graphes réside dans l'alignement avec ces graphes pour qu'une compréhension puisse avoir lieu. Plusieurs efforts ont été réalisés dans ce domaine, on peut les classer en deux catégories *graph to-text*. [48] [49] [50] [51] et *graph-to-token* [52] [53] [54]

### 5.3 Mesurer l’alignement

Notre objectif est de pouvoir travailler sur l’alignement des LLMs pour la génération de contenus éducatifs. Pour constater les résultats de cet alignement, il est essentiel de trouver un moyen de mesurer et d’analyser les données obtenues.

Il existe de nombreux modèles de mesures étudiés dans la littérature en neurosciences cognitives comme *la méthode d’analyse de similarité représentationnelle (RSA)* via Turing RSA [55] qui compare la similarité des résultats entre différents LLMs, ou encore *la méthode de mesure de similarité fonctionnelle* qui évalue les représentations à différentes couches avant la sortie. [56]

### 5.4 Performance de l’alignement

L’objectif de cette partie est de comprendre si l’alignement est bénéfique pour la performance des modèles.

L’alignement des représentations est souvent associé à une meilleure robustesse et capacité de généralisation dans les tâches. L’Article [57] nous démontre qu’il y a une complexité dans les différents niveaux d’alignement :

- *un alignement moyen* se révèle parfois moins pertinent qu’un alignement faible. L’alignement doit donc être utilisé avec précaution car il n’est pas toujours bénéfique et doit être adapté aux tâches ou du domaine d’application “*Models with either high or low alignment with humans are better at few-shot learning (FSL) than models with medium alignment, even correcting for pre-training performance*”
- *une contrainte de ressources*, atteindre un alignement élevé nécessite de grandes quantités de données annotées par des humains “*Achieving high alignment often requires substantial amounts of human-labeled data, which may not always be practical or cost-effective*”

Dans le contexte de l’enseignement entre machines ou entre machines et humains, un alignement peut améliorer significativement l’apprentissage, notamment lorsqu’il est associé à une capacité d’adaptation des enseignants pour comprendre les représentations des élèves. Cela souligne l’importance de concevoir des systèmes qui peuvent s’adapter aux spécificités de leurs utilisateurs.

## 6 Conclusion

En étudiant l'alignement, nous avons trouvé du potentiel pour améliorer la personnalisation du contenu pour les étudiants. Nous avons une idée plus claire sur la marche à suivre de notre projet. Cela nous réconforte dans le fait qu'il s'agisse d'un gap scientifique et que personne n'ait essayé de faire l'alignement avec les représentations mentales pour les environnements éducatifs. Nous avons identifié des axes d'implémentation : représenter la cognition (les ensembles de concepts), identifier des approches pour aligner LLMs avec représentation mentale sous forme de graphe. Nous sommes prêts à tester tout cela mais il se peut que l'alignement ne soit pas aussi efficace que ce que l'on attend car il y a des études démontrant que l'alignement n'est pas forcément quelque chose de positif, notamment via des benchmarks car comme vu dans la partie performance, certains alignements peuvent s'avérer moins optimales . Si nous prenons un LLM sans fine tuning ou bien sans attente humaine celui-ci sera simplement un générateur de texte, pour éviter cela on trouve des pistes que l'on testera lors de l'implémentation. Ainsi, si nos pistes sont viables, nous pourrions obtenir l'environnement éducation intelligent que l'on souhaite.



# Références

- [1] Kazem Jahanbakhsh, Mahdi Hajiabadi, Vipul Gagrani, Jennifer Louie, and Saurabh Khanwalkar. Beyond hallucination : Building a reliable question answering explanation system with gpts. In *NeurIPS Workshop on Generative AI for Education (GAIED)*, 2023.
- [2] Ahatsham Hayat and Mohammad Rashedul Hasan. Personalization and contextualization of large language models for improving early forecasting of student performance. In *NeurIPS Workshop on Generative AI for Education (GAIED)*, 2023.
- [3] Ziwei Xu, Sanjay Jain, and Mohan Kankanhalli. Hallucination is inevitable : An innate limitation of large language models. *arXiv preprint arXiv :2401.11817*, 2024.
- [4] Christian Bourion. Le concept de représentation mentale. *Revue internationale de psychosociologie*, (25) :21–38, 2005.
- [5] Several. NeurIPS’23 Workshop : Generative AI for Education (GAIED). <https://gaied.org/neurips2023/index.html>, 2023. [Online ; accessed 15-October-2024].
- [6] Robin Schmucker, Meng Xia, Amos Azaria, and Tom Mitchell. Ruffle&riley : Towards the automated induction of conversational tutoring systems. *arXiv preprint arXiv :2310.01420*, 2023.
- [7] Unggi Lee, Sanghyeok Lee, Junbo Koh, Yeil Jeong, Haewon Jung, Gyuri Byun, Yunseo Lee, Jewoong Moon, Jieun Lim, and Hyeoncheol Kim. Generative agent for teacher training : Designing educational problem-solving simulations with large language model-based agents for pre-service teachers.
- [8] Ariel Blobstein, Daniel Izmaylov, Tal Yifat, Michal Levy, and Avi Segal. Angel : A new generation tool for learning material based questions and answers.
- [9] Hunter McNichols, Wanyong Feng, Jaewook Lee, Alexander Scarlatos, Digory Smith, Simon Woodhead, and Andrew Lan. Automated distractor and feedback generation for math multiple-choice questions via in-context learning, 2023.
- [10] Fares Fawzi, Sadie Amini, and Sahan Bulathwela. Small generative language models for educational question generation. In *Proceedings of the NeurIPS Workshop on Generative Artificial Intelligence for Education (GAIED), New Orleans, LA, USA*, volume 15, 2023.
- [11] Shreya Bhandari, Yunting Liu, and Zachary A Pardos. Evaluating chatgpt-generated textbook questions using irt. In *Proceedings of the Generative AI for Education Workshop (GAIED) at the Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS). New Orleans, LA*, 2023.
- [12] Sumit Asthana, Taimoor Arif, and Kevyn Collins Thompson. Field experiences and reflections on using llms to generate comprehensive lecture metadata. In *NeurIPS’23 Workshop on Generative AI for Education (GAIED)*, 2023.
- [13] Michel Billières. L’avènement de la cognition. <https://www.verbotonale-phonetique.com/avenement-cognition/>, 2016. [Online ; accessed 26-November-2024].

- [14] Claude Meyer. Les représentations mentales. entre «res» et «flatus vocis». *Communication. Information médias théories pratiques*, 21(1) :9–31, 2001.
- [15] Séverine Casalis. Les représentations mentales. [https://pro.univ-lille.fr/fileadmin/user\\_upload/pages\\_pros/severine\\_casalis/Cours\\_les\\_representations\\_L1\\_cognitive\\_CASALIS.pdf](https://pro.univ-lille.fr/fileadmin/user_upload/pages_pros/severine_casalis/Cours_les_representations_L1_cognitive_CASALIS.pdf), 2023-2024. [Online; accessed 27-November-2024].
- [16] Andrée Archambault and Michèle Venet. Le développement de l’imagination selon piaget et vygotsky : d’un acte spontané à une activité consciente. *Revue des sciences de l’éducation*, 33(1) :5–24, 2007.
- [17] Mark K Ho, David Abel, Thomas L Griffiths, and Michael L Littman. The value of abstraction. *Current opinion in behavioral sciences*, 29 :111–116, 2019.
- [18] Carla Bugosen. Développement cognitif de l’enfant : les stades chez J.Piaget. <https://psy-enfant.fr/stade-developpement-jean-piaget/>, 2018. [Online; accessed 28-November-2024].
- [19] Iovka B Boneva, Arend Rensink, Marcos E Kurban, and Jörg Bauer. Graph abstraction and abstract graph transformation. 2007.
- [20] Jean-Louis Ermine, Mathias Chaillot, Philippe Bigeon, Boris Charreton, and D—MKSM Malavieille. Méthode pour la gestion des connaissances. *Ingénierie des systèmes d’information, AFCET-Hermès*, 4(4) :541–575, 1996.
- [21] Cécile. Vers l’apprentissage de l’abstraction (Partie 1). <https://www.lecoledemesreves.com/vers-lapprentissage-de-labstraction-partie-1/>. [Online; accessed 27-November-2024].
- [22] Tourigny et Capus. Représentation des connaissances. <http://www2.ift.ulaval.ca/~ericae/ift17586/e2003/diapos/RC.pdf>, 2001. [Online; accessed 01-December-2024].
- [23] Bernard ESPINASSE. Représentation des connaissances : Introduction aux Graphes Conceptuels. <https://pageperso.lis-lab.fr/bernard.espinasse/wp-content/uploads/2021/12/GraphesConceptuels-oct08-4p.pdf>, 2021. [Online; accessed 28-November-2024].
- [24] Mark A Musen. The protégé project : a look back and a look forward. *AI matters*, 1(4) :4–12, 2015.
- [25] DENNY VRANDECIC and MARKUS KRÖTZSCH. Wikidata : A free. *COMMUNICATIONS OF THE ACM*, 57(10), 2014.
- [26] J. Gabarró Thomas Steiner, Raphaël Troncy. Adding Realtime Coverage to the Google Knowledge Graph. [https://www.researchgate.net/publication/233985837\\_Adding\\_Realtime\\_Coverage\\_to\\_the\\_Google\\_Knowledge\\_Graph](https://www.researchgate.net/publication/233985837_Adding_Realtime_Coverage_to_the_Google_Knowledge_Graph), 2012. [Online; accessed 02-November-2024].
- [27] Thomas Pellissier Tanon, Gerhard Weikum, and Fabian Suchanek. Yago 4 : A reasonable knowledge base. In *The Semantic Web : 17th International Conference, ESWC 2020*,

*Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings 17*, pages 583–596. Springer, 2020.

- [28] Pierre Maillot, Sébastien Ferré, Peggy Cellier, Mireille Ducassé, and Franck Partouche. Nested forms with dynamic suggestions for quality rdf authoring. In *Database and Expert Systems Applications : 28th International Conference, DEXA 2017, Lyon, France, August 28-31, 2017, Proceedings, Part I* 28, pages 35–45. Springer, 2017.
- [29] Sandrine Vieillevoys, Valérie Wathelet, and Marc Romainville. Maîtrise des prérequis et réussite à l’université. In *Réussite, échec et abandon dans l’enseignement supérieur*, pages 221–250. De Boeck, 2012.
- [30] Sudeshna Roy, Meghana Madhyastha, Sheril Lawrence, and Vaibhav Rajan. Inferring concept prerequisite relations from online educational resources. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 9589–9594, 2019.
- [31] Joy He-Yueya, Wanjing Anya Ma, Kanishk Gandhi, Benjamin W Domingue, Emma Bruns-kill, and Noah D Goodman. Psychometric alignment : Capturing human knowledge distributions via language models. *arXiv preprint arXiv :2407.15645*, 2024.
- [32] Tianhao Shen, Renren Jin, Yufei Huang, Chuang Liu, Weilong Dong, Zishan Guo, Xinwei Wu, Yan Liu, and Deyi Xiong. Large language model alignment : A survey. *arXiv preprint arXiv :2309.15025*, 2023.
- [33] Christophe Lacroix. Prévenir les discriminations résultant de l’utilisation de l’intelligence artificielle. <https://assembly.coe.int/LifeRay/EGA/Pdf/TextesProvisaires/2020/20200915-PreventingDiscriminationAI-FR.pdf>, 2020. [Online ; accessed 01-November-2024].
- [34] Dominic Ligot. AI Alignment Learning from how ChatGPT Learns. [https://www.researchgate.net/publication/370264910\\_AI\\_Alignment\\_Learning\\_from\\_how\\_ChatGPT\\_Learns](https://www.researchgate.net/publication/370264910_AI_Alignment_Learning_from_how_ChatGPT_Learns), 2023. [Online ; accessed 01-November-2024].
- [35] Ling kai Kong, Haorui Wang, Wenhao Mu, Yuanqi Du, Yuchen Zhuang, Yifei Zhou, Yue Song, Rongzhi Zhang, Kai Wang, and Chao Zhang. Aligning large language models with representation editing : A control perspective. *arXiv preprint arXiv :2406.05954*, 2024.
- [36] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization : Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.
- [37] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv :1707.06347*, 2017.
- [38] Haoran Xu, Amr Sharaf, Yunmo Chen, Weiting Tan, Lingfeng Shen, Benjamin Van Durme, Kenton Murray, and Young Jin Kim. Contrastive preference optimization : Pushing the boundaries of llm performance in machine translation. *arXiv preprint arXiv :2401.08417*, 2024.

- [39] Amanda Askill, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Ben Mann, Nova DasSarma, et al. A general language assistant as a laboratory for alignment. *arXiv preprint arXiv :2112.00861*, 2021.
- [40] Zhexin Zhang, Junxiao Yang, Pei Ke, Fei Mi, Hongning Wang, and Minlie Huang. Defending large language models against jailbreaking attacks through goal prioritization. *arXiv preprint arXiv :2311.09096*, 2023.
- [41] Bill Yuchen Lin, Abhilasha Ravichander, Ximing Lu, Nouha Dziri, Melanie Sclar, Khyathi Chandu, Chandra Bhagavatula, and Yejin Choi. The unlocking spell on base llms : Rethinking alignment via in-context learning. *arXiv preprint arXiv :2312.01552*, 2023.
- [42] Unknown. L’IA générative et les graphes de connaissance au service de la productivité. <https://www.clemenceconsulting.fr/fr/llm-et-graphe-de-connaissance-une-synergie-pour-lia-au-service-de-la-productivite/>. [Online; accessed 26-November-2024].
- [43] Ludo Louis. Les graphes de connaissances (KG) et les modèles de langage (LLM) : une fusion puissante. <https://ludo-louis.fr/kg-graphe-connaissance-et-llm-une-fusion-puissante/>, 2023. [Online; accessed 28-November-2024].
- [44] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. A survey on multimodal large language models. *National Science Review*, page nwae403, 2024.
- [45] Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. Tallrec : An effective and efficient tuning framework to align large language model with recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1007–1014, 2023.
- [46] Yassir Fathullah, Chunyang Wu, Egor Lakomkin, Junteng Jia, Yuan Shangguan, Ke Li, Jinxi Guo, Wenhan Xiong, Jay Mahadeokar, Ozlem Kalinli, et al. Prompting large language models with speech recognition abilities. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 13351–13355. IEEE, 2024.
- [47] Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S Yu, et al. Towards graph foundation models : A survey and beyond. *arXiv preprint arXiv :2310.11829*, 2023.
- [48] Zhikai Chen, Haitao Mao, Hang Li, Wei Jin, Hongzhi Wen, Xiaochi Wei, Shuaiqiang Wang, Dawei Yin, Wenqi Fan, Hui Liu, et al. Exploring the potential of large language models (llms) in learning on graphs. *ACM SIGKDD Explorations Newsletter*, 25(2) :42–61, 2024.
- [49] Bahare Fatemi, Jonathan Halcrow, and Bryan Perozzi. Talk like a graph : Encoding graphs for large language models. *arXiv preprint arXiv :2310.04560*, 2023.

- [50] Jiayan Guo, Lun Du, Hengyu Liu, Mengyu Zhou, Xinyi He, and Shi Han. Gpt4graph : Can large language models understand graph structured data? an empirical evaluation and benchmarking. *arXiv preprint arXiv :2305.15066*, 2023.
- [51] Chang Liu and Bo Wu. Evaluating large language models on graphs : Performance insights and comparative analysis. *arXiv preprint arXiv :2308.11224*, 2023.
- [52] Runjin Chen, Tong Zhao, Ajay Jaiswal, Neil Shah, and Zhangyang Wang. Llaga : Large language and graph assistant. *arXiv preprint arXiv :2402.08170*, 2024.
- [53] Lecheng Kong, Jiarui Feng, Hao Liu, Chengsong Huang, Jiaxin Huang, Yixin Chen, and Muhan Zhang. Gofa : A generative one-for-all model for joint graph language modeling. *arXiv preprint arXiv :2407.09709*, 2024.
- [54] Jiabin Tang, Yuhao Yang, Wei Wei, Lei Shi, Lixin Su, Suqi Cheng, Dawei Yin, and Chao Huang. Graphgpt : Graph instruction tuning for large language models. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 491–500, 2024.
- [55] Michael Wolmetz Mattson Ogg. Measuring Alignment between Human and Artificial Intelligence with Representational Similarity Analysis. [https://2024.ccneuro.org/pdf/522\\_Paper\\_authored\\_Ogg\\_et\\_al\\_CCNeuro\\_Proxy\\_RSA\\_WordTemplate.pdf](https://2024.ccneuro.org/pdf/522_Paper_authored_Ogg_et_al_CCNeuro_Proxy_RSA_WordTemplate.pdf). [Online; accessed 20-November-2024].
- [56] Max Klabunde, Tobias Schumacher, Markus Strohmaier, and Florian Lemmerich. Similarity of neural network models : A survey of functional and representational measures. *arXiv preprint arXiv :2305.06329*, 2023.
- [57] Ilia Sucholutsky and Tom Griffiths. Alignment with human representations supports robust few-shot learning. *Advances in Neural Information Processing Systems*, 36, 2024.

## 7 Acronymes

*LLMs* Grands modèles de langages