

PROEDUCA

Guide pour l'utilisation responsable de l'IA générative dans les tâches de recherche

17/07/2024

Index

1. Résumé Analytique - 3
 - Points clés d'application générale
 - Guide pour cas pratiques
2. Introduction - 5
 - Utilisation basique de l'IA générative
3. Utilisation responsable de l'IA - 7
 - Édition de texte
 - Production de contenu, y compris audiovisuel
 - Thèses de doctorat
 - Mémoires de Projets Concurrentiels et Subventions de Recherche
 - Révision et Évaluation de textes et de Productions Scientifiques
 - Recherche Appliquée et Mise en Œuvre de Projets Concurrentiels
4. Conception & Attribution - 10
 - Reconnaissance de l'Utilisation des Outils d'IA Générative
5. Crédits - 11
6. Droits d'auteur et références - 11
7. Références - 11
8. ANNEXE 1. Exemples de Conversations et de Tâches avec l'IA Générative - 13

1. Résumé Analytique

Proeduca établit des lignes directrices pour l'utilisation responsable des outils d'Intelligence Artificielle (IA) Générative dans le domaine de la recherche, qui sont développés ou déployés par toutes les universités du groupe éducatif, telles que l'Université Internationale de La Rioja (UNIR), UNIR Mexico, MIU City University Miami, UNIPRO et d'autres.

1.1. Points clés d'application générale

- **Éthique** : en général, les mêmes principes éthiques s'appliquent à l'utilisation de l'IA générative dans la recherche qu'à tout autre processus de production, d'édition, de révision et de diffusion, avec un soin particulier accordé à l'auteur et à son identification, à la confidentialité et à la sécurité de ses propres données et de celles des autres, et à l'utilisation responsable en tout scénario.
- **Conception & Attribution** : l'IA générative ne peut pas être considérée comme auteur d'un article de recherche. Seules les personnes qui peuvent en assumer la responsabilité et consentir à leur inclusion doivent être reconnues comme auteurs.
- **Reconnaissance** : il est obligatoire de documenter l'utilisation de l'IA générative à n'importe quelle étape du processus de recherche. Les chercheurs doivent vérifier l'originalité du contenu, éviter le plagiat, respecter les droits d'auteur et s'assurer qu'ils n'introduisent pas de biais par l'utilisation des outils d'IA ou des produits qu'elles génèrent. Selon l'utilisation spécifique, des notes de bas de page, des remerciements ou des citations peuvent être utilisés pour reconnaître l'aide de l'IA.
- **Utilisation responsable** : une distinction doit être faite entre la correction de contenu et l'aide à la création ou la modification substantielle de celle-ci. Si le premier cas ne nécessite pas de citation spécifique, toute reformulation ou aide significative de l'AI doit être citée.
- **Confidentialité et sécurité des données** : il est nécessaire d'être prudent avant de partager des données avec des outils commerciaux d'IA générative, en évitant de divulguer du matériel sensible ou protégé à la fois par l'auteur ou par l'institution à laquelle il appartient. Il existe des alternatives qui permettent d'utiliser localement l'IA générative dans le but de protéger la vie privée.

1.2 Guide pour cas pratiques

- **Édition de texte** : la réécriture substantielle d'un texte devra reconnaître l'utilisation des outils d'IA générative. L'utilisation de tout type d'outil pour vérifier l'orthographe, la construction grammaticale ou les expressions dans des langues non maternelles est autorisée.
- **Production de contenu, y compris audiovisuel** : l'utilisation de l'IA générative pour créer ou modifier des images, des vidéos, des audios et, en général, tout produit multimédia, destinés à faire partie d'une thèse de doctorat ou d'un article scientifique, n'est pas recommandée, sauf si la génération de l'image elle-même par l'IA est pertinente pour le sujet de recherche. À l'inverse, cette recommandation autorise les outils d'édition d'images, tels que la modification du contraste ou de la luminosité, ainsi que d'autres aspects de la retouche d'images sans modification substantielle de l'original. Dans le cas d'une production audiovisuelle ou de tout type d'utilisation associée à un projet de recherche (révision systématique, contenu éducatif, code de programmation, référentiel de compétences, etc.), l'utilisation de l'IA est permise seulement à condition de recevoir l'autorisation du vice-rectorat correspondant et, en outre, de respecter les points clés vus dans la section précédente.

- **Thèses de doctorat** : étant donné l'importance de l'apprentissage autonome, l'utilisation de l'IA générative pour la rédaction de thèses doctorales, y compris pour les états de l'art, n'est pas autorisée. Dans tous les autres cas (par exemple comme outil d'aide à l'analyse quantitative), le doctorant doit avoir l'autorisation expresse de son directeur de thèse et des instances universitaires nécessaires.
- **Propositions de projets concurrentiels et de subventions de recherche** : en règle générale, il n'est pas permis d'utiliser l'IA générative dans la rédaction de propositions de subventions. Si cette ressource doit être utilisée pour des paragraphes ou des sections complémentaires, l'autorisation du vice-rectorat dont dépend la convocation est requise.
- **Examen et évaluation de la production scientifique** : pour l'examen ou l'évaluation de mémoires de projets concurrentiels, d'articles scientifiques, de livres, de thèses, de rapports techniques ou de tout autre type de production scientifique, qui sont commandés par n'importe quelle université du groupe éducatif Proeduca, l'utilisation d'outils d'IA générative n'est pas autorisée. Dans le cas d'une commande d'un organisme (public, privé), d'une institution ou d'une communauté scientifique, la réalisation personnelle est recommandée, mais les directives de l'organisme demandeur seront suivies.
- **Recherche appliquée et mise en œuvre de projets concurrentiels** : l'utilisation de l'IA générative pour des tâches de recherche appliquée est autorisée en tant qu'outil d'amélioration de la productivité. Dans tous les cas, la confidentialité des données doit toujours être prise en compte et seules les données qui pourraient également être partagées avec des organisations externes peuvent être partagées avec des outils d'IA générative selon la politique de confidentialité. Une fois partagée avec un outil d'IA générative, toute donnée ou document est ouvertement exposé. Pour la rédaction de rapports techniques de n'importe quel type de projet de recherche financé ou supervisé en interne ou en externe, l'utilisation de l'IA générative n'est pas autorisée.

Les universités du groupe éducatif Proeduca s'engagent à surveiller le développement et la mise en place de ces technologies et à ajuster cette politique si nécessaire, en s'alignant sur les pratiques adoptées par d'autres institutions académiques au niveau national et international. Tout chercheur d'une université du groupe éducatif Proeduca qui utilise l'IA générative dans un ou plusieurs des cas envisagés dans cette politique, ou dans tout autre cas actuel ou futur, le fait sous sa propre responsabilité et attribution, et ni Proeduca ni ses universités ne sont responsables de toute utilisation inappropriée de celle-ci.

2. Introduction

Les outils d'IA générative ont la capacité de créer du texte, une image, une voix ou une vidéo en réponse à une indication (un « prompt » en anglais). Des exemples de ces outils sont : ChatGPT-3.5, ChatGPT-4.0, Gemini ou Grammarly, pour la génération de textes, ou DALL-E, pour la génération d'images.

Les outils d'IA générative peuvent être utilisés de différentes manières au cours du processus de recherche : rédaction, traduction et révision de textes, découverte et analyse de documentation, génération automatique de code, génération et modification d'images, analyse de données, etc. Également pour la production de ressources associées comme la vidéo, l'image, le texte, la voix (pour générer, éditer ou cloner, par exemple). Il est important de souligner que l'exactitude et l'exhaustivité des résultats donnés par l'IA générative ne peuvent pas être assurées par défaut et qu'il faut toujours procéder à des vérifications.

Ce document vise à fournir un guide aux chercheurs sur le sujet de l'IA générative et des technologies assistées par l'IA. Proeduca suivra leur développement et ajustera ou affinera cette politique le cas échéant. Les chercheurs peuvent légitimement utiliser des outils d'IA générative dans le cadre des politiques et directives associées, cela s'applique également aux doctorants.

Cette politique de Proeduca mise en place par ses universités s'inscrit dans la lignée de celles déjà mises en place par d'autres universités et centres de recherche nationaux (Universidad Carlos III de Madrid [1], Universidad de Granada [2], Universidad Autónoma de Madrid [3]) et internationaux (Deakin University [4], The University of Melbourne [5, 6], Australian Research Council [7], University of California, Los Angeles [8], Aalto University [9]).

2.1 Utilisation rudimentaire de l'IA générative

L'IA générative fait référence à un volet de l'IA centré sur la création de modèles capables de générer de nouveaux contenus, qu'il s'agisse de textes, d'images, de vidéos, de voix ou de musique, entre autres formats. Si nous nous concentrons sur la génération de texte, un grand modèle de langage (ou LLM, Large Language Model) est un élément essentiel de l'IA générative. Il s'agit d'un algorithme entraîné pour comprendre et générer un texte cohérent dans une ou plusieurs langues. Ces modèles utilisent des réseaux neuronaux pour identifier des schémas linguistiques à partir de vastes ensembles de données textuelles d'entraînement. Ils diffèrent d'un LLM de base, qui consistent en des modèles statistiques capturant les schémas du langage, sans aucune application spécifique, et des modèles adaptés à des applications concrètes. Ceux-ci sont des modèles de base qui ont été adaptés à des tâches particulières, telles que la traduction de textes ou la tenue de conversations. Ainsi, GPT 4.0 est un modèle de base et ChatGPT 4.0 est un modèle adapté et entraîné pour avoir des conversations avec des humains.

La manière la plus simple d'interagir avec les outils d'IA générative, qu'ils génèrent du texte, de l'image ou de la voix, consiste à utiliser des prompts ou indications. Un prompt consiste en une courte instruction ou question envoyée au modèle d'IA générative pour lancer le processus de génération de contenu. Il est essentiel de formuler des prompts clairs et concis, car ils déterminent en grande partie la qualité et la pertinence des réponses générées. Un prompt efficace doit présenter les caractéristiques suivantes :

- Clarté : il doit être compréhensible et spécifique pour guider le modèle vers la génération du type de contenu souhaité.
- Concision : il doit être bref et aller droit au but, en évitant les ambiguïtés ou les redondances dans les informations fournies.
- Contextualisation : il doit fournir le contexte pertinent pour aider le modèle à générer des réponses plus précises et cohérentes.

Par exemple, le prompt : « Parle-moi du changement climatique » n'est ni concis, ni clair, et produira des réponses qui peuvent être vagues ou ne pas correspondre au résultat attendu. Voici un exemple qui apporte de la clarté et du contexte :

« Je suis un professeur enseignant à des enfants de 14 ans à qui je veux présenter le changement climatique. Donne-moi un aperçu complet et détaillé de l'évolution historique du changement climatique, en incluant ses causes, ses répercussions et la compréhension scientifique de ce phénomène ».

Le prompting est généralement un processus itératif dans lequel l'utilisateur saisit un prompt clair, concis et contextuel, puis évalue les résultats produits par le système. En fonction de ce résultat, il redéfinit le prompt pour continuer à fournir le contexte nécessaire à la production d'un résultat satisfaisant. C'est-à-dire qu'un résultat satisfaisant ne s'obtient pas avec un seul prompt, mais par des conversations avec l'IA générative. Ce sont ces conversations qui fournissent de plus en plus de contexte et de conseils à l'outil pour qu'il produise les résultats souhaités.

Dans le cas particulier d'une IA générative de texte, les principales fonctions sont de résumer, déduire, développer et de transformer un texte. Les exemples de transformation comprennent la traduction du texte, sa correction et son amélioration, le changement de style (formel, informel, scientifique, etc.) et le changement de format (prose, vers, etc.) ou de sortie (HTML, JSON, Markdown, etc.) Ces principales fonctions ont plusieurs utilités dans le domaine de la recherche, comme le traitement de documents scientifiques, l'amélioration et l'expansion des idées de projet, l'autocorrection et le passage au style scientifique, la génération de code et l'analyse de données, le tuteur et le partenaire d'étude, etc.

Il existe également une grande variété de stratégies pour élaborer des prompts [10][11]. L'une des plus courantes est le RTF (Role-Task-Format), dans lequel le prompt indique le rôle que doit jouer le LLM, la tâche qu'il doit accomplir et le format dans lequel la réponse doit se faire. Par exemple :

« Comporte-toi comme un professeur d'histoire de l'art. Crée un programme d'études pour la matière Introduction à l'art moderne en Espagne. Le cours dure 14 semaines à raison de deux leçons de 90 minutes par semaine. Génère également deux devoirs, en semaine 5 et en semaine 12, pour un travail à rendre. Génère le résultat en HTML pour qu'il puisse être téléchargé sur une page web. »

Ce prompt utilise une stratégie RTF qui fournit également le contexte nécessaire pour effectuer la tâche (le nombre de classes, la durée, format demandé pour le résultat...). Évidemment, le premier prompt ne produira pas le résultat souhaité, les prochains prompts devront donc indiquer les modifications à apporter jusqu'à l'obtention d'un résultat satisfaisant. Plusieurs exemples de conversations génératives avec l'IA ont été insérés dans l'Annexe 1.

Enfin, comme nous l'avons évoqué dans l'introduction, il est important de souligner que l'exactitude et l'exhaustivité des résultats de l'IA générative ne peuvent pas être assurées par défaut et qu'il faut toujours procéder à des vérifications. L'un des phénomènes à prendre en compte dans le cadre de l'IA générative est celui des hallucinations. Ce terme décrit les situations dans lesquelles l'IA génère des informations ou des données incorrectes, invérifiables ou complètement fictives, en les présentant comme si elles étaient avérées. L'existence de ces hallucinations souligne l'importance cruciale de vérifier et de faire valider les résultats produits par des algorithmes d'IA générative. La section 4 présente plus en détail les étapes à suivre avant d'utiliser les résultats produits par une IA générative.

3. Utilisation responsable de l'IA

Cette politique distingue cinq domaines d'application de l'IA générative : l'édition de textes, les images, les thèses doctorales, les mémoires de projets concurrentiels et les tâches de recherche.

3.1. Édition de texte

L'édition de textes comprend toute modification de texte, par exemple la réécriture, le développement, la traduction et/ou le filtrage. On peut distinguer deux niveaux : (1) la correction et l'amélioration du contenu, qui comprendraient, par exemple, des changements grammaticaux et orthographiques (2) la reformulation, le développement, la réécriture et le filtrage du texte, qui se concentrent sur la génération/réécriture du texte à partir d'une entrée, que ce soit une question ou une idée simple à partir de laquelle reformuler, filtrer ou développer. Un exemple du premier cas serait, par exemple, l'aide fournie par Microsoft Word ou Google Docs, qui se limite (pour l'instant) à la correction orthographique et aux améliorations grammaticales. Les outils les plus récents offrent une aide supplémentaire à l'édition. Grammarly, par exemple, sert à vérifier l'expression et la grammaire, mais peut aussi aider à la reformulation, au développement, au filtrage, à la réécriture et à la traduction.

Pour le premier cas, il n'est pas nécessaire de citer l'utilisation de l'outil, mais pour tout texte répondant au deuxième cas (y compris la traduction entre différentes langues), l'utilisation de l'outil correspondant doit être citée. En général, une réécriture substantielle d'un texte sera toujours considérée comme plus importante qu'une correction d'un texte, et l'utilisation d'outils d'IA générative dans le cadre de sa rédaction devra être citée.

3.2. Production de contenu, y compris audiovisuel

L'utilisation de l'IA générative pour créer ou modifier des images, des vidéos, des audios et, en général, tout produit multimédia, destinés à faire partie d'une thèse de doctorat ou d'un article scientifique, n'est pas recommandée, sauf si la génération de l'image par l'IA est elle-même pertinente pour le sujet de recherche. **À l'inverse, cette recommandation autorise les outils d'édition d'images,** tels que la modification du contraste ou de la luminosité, ainsi que d'autres aspects de la retouche d'images sans modification substantielle de l'original.

Dans le cas d'une production audiovisuelle ou de tout type d'utilisation associée à un projet de recherche (révision systématique, contenu éducatif, code de programmation, référentiel de compétences, etc.), **l'utilisation de l'IA est permise seulement à condition de recevoir l'autorisation** du vice-rectorat correspondant et, en outre, de respecter les points clés vus dans la section précédente.

3.3. Thèses de doctorat

Les thèses de doctorat sont un cas particulier en raison du processus d'apprentissage de l'étudiant. On attend d'un doctorant qu'il soit capable de mener à bien toutes les étapes de la recherche, de l'élaboration des concepts à la communication des résultats, ce qui inclut non seulement la rédaction de textes, mais aussi la structuration d'arguments, l'élaboration de descriptions complexes, etc.

En général, l'un des principaux objectifs d'une thèse de doctorat est que l'étudiant apprenne à exprimer efficacement des idées complexes à l'écrit, de sorte que **l'utilisation de l'IA générative n'est pas autorisée pour la rédaction ou la réécriture substantielle de tout élément d'une thèse de doctorat, y compris pour les états de l'art ou l'analyse méthodique de documentation.**

Si le directeur de thèse du programme doctoral et l'école doctorale de l'université le jugent approprié, l'étudiant pourrait être autorisé à utiliser l'IA générative pour des tâches appliquées spécifiques (voir la section correspondante) et les points évoqués dans les sections Conception & Attribution et Reconnaissance de ce document s'appliqueraient alors.

Il existe des outils d'IA générative pour effectuer des états de l'art et comprendre la littérature scientifique. L'utilisation d'outils d'IA générative pour l'analyse de documentation et les recherches sur l'état de l'art n'est pas conseillée. Cela s'explique par l'importance du processus de connaissance de l'état de l'art dans la réalisation d'une thèse, et par le fait que les recueils d'articles de recherche préexistants identifiés par les outils d'IA générative ne sont généralement ni exhaustifs, ni nécessairement les plus pertinents.

De même, l'utilisation d'outils d'IA générative n'est pas recommandée pour comprendre des articles scientifiques ou pour en faire des résumés. Lire un article implique de le comprendre et c'est la clé du processus d'apprentissage. Par ailleurs, les résumés produits par ces outils peuvent sembler efficaces, mais le contenu peut pourtant être incohérent ou incorrect.

3.4. Mémoires de Projets Concurrentiels et Subventions de Recherche

Lors de la rédaction des demandes de subvention, il faut toujours garder en tête que l'utilisation de l'IA générative peut poser des problèmes de propriété intellectuelle et de droits d'auteur, entre autres, en plus du risque d'être factuellement incorrecte [7].

En règle général, l'utilisation de l'IA générative pour la rédaction de mémoires de projets concurrentiels n'est pas autorisée. Pour un usage ponctuel (paragraphe unique, sections complémentaires), les candidats doivent suivre les points indiqués dans les sections Conception & Attribution et Reconnaissance détaillée ci-dessous, en plus de demander l'autorisation au vice-rectorat responsable de la gestion de la convocation. Il est également recommandé de s'assurer que l'organisme de financement n'impose pas de limites à l'utilisation de l'IA générative.

3.5. Révision et Évaluation de Textes et de Productions Scientifiques

Pour l'examen ou l'évaluation de mémoires de projets concurrentiels, d'articles scientifiques, de livres, de thèses, de rapports techniques ou de tout autre type de production scientifique, qui sont commandés par n'importe quelle université du groupe éducatif Proeduca, l'utilisation d'outils d'IA générative n'est pas autorisée, en raison, entre autres, de la nécessité de fournir une critique juste et impartiale et de la nécessité de maintenir la confidentialité de l'information. Dans le cas d'une commande d'un organisme (public, privé), d'une institution ou d'une communauté scientifique, la réalisation personnelle est recommandée, mais les directives de l'organisme demandeur seront suivies.

3.6. Recherche Appliquée et Mise en Œuvre de Projets Concurrentiels

Les outils d'IA générative ont de nombreuses applications dans la recherche, au-delà de la génération de texte et de la structuration de document. Certaines de ces applications comprennent l'analyse de données, l'analyse de texte ou la génération de recommandations. **L'utilisation de l'IA générative est autorisée pour ces tâches en tant qu'outil d'amélioration de la productivité.** Dans tous les cas, la confidentialité des données doit toujours être prise en compte et seules les données qui pourraient également être partagées avec des organisations externes peuvent être partagées avec des outils d'IA générative selon la politique de confidentialité. Une fois partagée avec un outil d'IA générative, toute donnée ou document est ouvertement exposé. [12].

En uploadant des données vers un outil commercial, on en perd le contrôle et, bien que les entreprises puissent fournir des options aux utilisateurs pour limiter l'utilisation de ce contenu, il y a toujours un risque de réutilisation des informations envoyées. En particulier, le contenu protégé par droits d'auteur, le contenu confidentiel, les informations personnelles et/ou les données de recherche basées sur l'être humain ne doivent pas être partagés avec des outils commerciaux. En général, le règlement général sur la protection des données (RGPD, ou GDPR selon l'acronyme anglais) doit toujours être pris en compte avant d'envoyer des données à une IA générative [9].

Compte tenu des indications ci-dessus, il est également déconseillé de partager des produits scientifiques (articles, thèses, rapports techniques, mémoires de projet, etc.) pour les soumettre à l'examen d'outils d'IA générative commerciaux, car cela pourrait entraîner le partage de certaines des informations énumérées ci-dessus et donc la perte de contrôle des données et des résultats associés. Certains éditeurs, comme SAGE [13], interdisent explicitement l'utilisation de l'IA générative dans le processus de révision des articles scientifiques.

Pour la rédaction de rapports techniques ou de produits résultant de n'importe quel type de projet de recherche financé ou supervisé en interne ou en externe, l'utilisation de l'IA générative n'est pas autorisée.

Si l'utilisation de l'IA générative est souhaitée pour un des cas cités plus haut, il existe des alternatives aux outils commerciaux qui peuvent être exécutées localement sans avoir besoin de matériel spécifique [14][15]. Dans ce cas, bien que le temps de réponse soit généralement plus long, la confidentialité des données est préservée, car les données ne quittent jamais l'environnement local.

4. Conception & Attribution

Seuls les auteurs qui donnent leur accord et qui ont la responsabilité de leur travail [16] peuvent être inclus dans la conception d'un travail. Tous sont responsables des résultats présentés dans le fruit du travail, y compris ceux pour lesquels des outils d'IA générative ont été utilisés.

Les outils d'IA générative ne peuvent pas être nommés comme auteurs d'un article de recherche, car ils ne peuvent donner leur consentement pour faire partie de l'article ni en être responsables. Dans tous les cas, les droits de propriété intellectuelle sont personnels, inaliénables et incessibles. C'est-à-dire qu'il ne peuvent pas être indûment appropriés, ni être cédés ou vendus.

4.1. Reconnaissance de l'Utilisation des Outils d'IA Générative

Si le contenu a été créé, révisé ou amélioré avec un outil d'IA générative à n'importe quelle étape du processus de recherche, ce fait doit être reconnu et documenté. De plus, il est à nouveau souligné que, comme l'exactitude et l'exhaustivité des résultats de l'IA générative ne peuvent pas être assurées par défaut, tous les auteurs sont responsables de garantir l'exactitude du travail publié et de reconnaître l'utilisation de l'IA générative. Cela inclut les mesures suivantes :

- Il incombe à tous les auteurs de s'assurer de la véracité du contenu généré.
- Il incombe à tous les auteurs de s'assurer que le contenu généré ne commet pas de plagiat, car il peut arriver que le résultat de l'IA générative soit une répétition d'une autre source.
- Il incombe à tous les auteurs de soumettre le contenu à l'examen des droits d'auteur, car l'IA générative peut produire des résultats protégés par la loi sur les droits d'auteur.
- Il incombe à tous les auteurs de s'assurer que les outils d'IA générative n'ont pas introduit de biais dans le contenu généré. Comme tout système basé sur des données, les outils d'IA générative peuvent introduire des biais dans les informations fournies, il faut donc faire un effort conscient pour les identifier et y remédier.
- Les auteurs doivent être explicites quant au contenu qui a été créé ou modifié par des outils d'IA, tel que des passages spécifiques du texte ou des images.

Si un article comprend un texte généré et/ou édité par l'IA, et qu'il remplit les conditions détaillées ci-dessus, il appartient alors aux auteurs d'établir et de reconnaître l'aide fournie pour le contenu de certaines sections du travail [16]. La façon de procéder dépendra de la tâche pour laquelle l'IA a été utilisée. Voici quelques options possibles :

- **Note de bas de page ou entrée pour référence** : par exemple, dans les cas d'édition et d'amélioration de texte ou de découverte de documentation, une note de bas de page est utilisée à la place d'une citation dans le texte, indiquant l'outil utilisé, la version (si disponible), la date à laquelle il a été utilisé et le texte du prompt. Étant donné que les conversations avec l'outil d'IA générative se déroulent généralement jusqu'à ce que le résultat souhaité soit atteint, **le lien vers cette conversation doit être joint**. La plupart des outils commerciaux ont l'option de générer un lien à partir de la conversation.
- **Citation** : il est particulièrement conseillé d'utiliser une citation traditionnelle lorsque l'IA générative a été utilisée pour l'analyse de données, ou pour la mise en œuvre de tâches telles que le traitement du langage ou la classification. En suivant le style APA, cela donnerait :
 - OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. <https://chat.openai.com/chat>

- **Section spécifique** : certains éditeurs peuvent exiger l'insertion d'une section spécifique détaillant l'utilisation de l'IA générative, précisant l'outil, pourquoi et dans quel but elle a été utilisée, et reconnaissant la contribution et la responsabilité sur le contenu de la publication. L'éditeur Elsevier, par exemple, exige l'utilisation de cette section spécifique [17].
- **Intégration automatique** : certains outils, comme Grammarly, disposent d'un bouton permettant de reconnaître l'utilisation de l'IA générative dans les textes et d'en inclure automatiquement la mention [18]. Le chercheur doit être conscient du degré de fiabilité de ces outils, ainsi que de leur niveau d'actualisation constant.
- **Autres cas** : certaines revues peuvent avoir leur propre politique sur l'utilisation de l'IA générative pour certaines des tâches ci-dessus, il est donc important de s'assurer pour chacun des cas que les exigences de la revue sont respectées avant l'envoi de l'article scientifique. Exemples : Elsevier [17], SAGE [13], Science [19] ou Wiley [20].

5. Crédits

- Éditeur : Daniel Burgos, Universidad Internacional de La Rioja (UNIR), MIU City University Miami
- Auteurs : Enrique Frías, Dai Griffiths, Pablo Moreno, Rubén González, José Carlos San José, Natalia Padilla, Alberto Corbí, Joaquín Alonso, Shila Ganguly, Stefania Aceto, Sérgio Vilela (UNIR)

6. Droits d'auteur et références

- © Universidad Internacional de La Rioja (UNIR) / Proeduca. Sous licence Creative Commons CC-BY-NC : <https://www.creativecommons.org/licenses/by-nc/4.0/deed.es>
- Citer comme suit : « Universidad Internacional de La Rioja (2024). Guide pour l'utilisation responsable de l'IA générative dans les tâches de recherche. Proeduca. DOI: <https://doi.org/10.5281/zenodo.11609187>”

7. Références

- [1] “Recomendaciones para la docencia con inteligencias artificiales generativas”, Universidad Carlos III de Madrid, Vicerrectorado de Postgrado y Formación Permanente (PDF; Accedido 5 marzo 2024), <https://e-archivo.uc3m.es/rest/api/core/bitstreams/456229c0-3670-4198-b86d-ff2762a3e1f9/content>
- [2] “Inteligencia Artificial en la universidad”, Centro de Producción de Recursos para la Universidad Digital, Universidad de Granada (PDF, Accedido 5 marzo 2024), <https://ceprud.ugr.es/formacion-tic/inteligencia-artificial>
- [3] “Guía básica sobre el uso de la Inteligencia artificial para docentes y estudiantes”, Universidad Autónoma de Madrid, <https://www.uam.es/uam/media/doc/1606941290988/guia-visual-iagen.pdf>
- [4] “Responsible use of GenAI in Research”, Deakin University, <https://deakin.libguides.com/generative-ai-research/begin>
- [5] “Statement on responsible use of digital assistance tools in research”, The University of Melbourne, <https://research.unimelb.edu.au/strengths/ethics/research-integrity-and-digital-assistance-tools>

- [6] “Graduate researchers and digital assistance tools”, The University of Melbourne (Accedido el 3 de abril de 2024), <https://gradresearch.unimelb.edu.au/preparing-my-thesis/graduate-researchers-and-digital-assistance-tools>
- [7] “Policy on Use of Generative Artificial Intelligence in the ARCs grants programs”, Australian Research Council (version 2023.1, 7 July 2023), <https://www.arc.gov.au/sites/default/files/2023-07/Policy%20on%20Use%20of%20Generative%20Artificial%20Intelligence%20in%20the%20ARCs%20grants%20programs%202023.pdf>
- [8] “Guidance for the use of Generative AI”, UCLA Center for the Advancement of Teaching, https://teaching.ucla.edu/resources/ai_guidance/
- [9] “Responsible use of Artificial Intelligence in the research process”, Aalto University, <https://www.aalto.fi/en/services/responsible-use-of-artificial-intelligence-in-the-research-process>
- [10] “Prompt engineering for educators – making generative AI work for you”, D. Liu, The University of Sidney, <https://educational-innovation.sydney.edu.au/teaching@sydney/prompt-engineering-for-educators-making-generative-ai-work-for-you/>
- [11] “Must-Read Best Of Practical Prompt Engineering Strategies To Become A Skillful Prompting Wizard In Generative AI”, L. Eliot, Forbes, Dec. 28 2023, <https://www.forbes.com/sites/lanceeliot/2023/12/28/must-read-best-of-practical-prompt-engineering-strategies-to-become-a-skillful-tprompting-wizard-in-generative-ai/?sh=1e3369d019cd>
- [12] “Generative Artificial Intelligence (AI). Thinking about using AI in your research? Read this first”, Deakin University, <https://www.deakin.edu.au/research/support-for-researchers/research-integrity/generative-artificial-intelligence-ai>
- [13] “ChatGPT and Generative AI”, Sage Editorial Policies, <https://us.sagepub.com/en-us/nam/chatgpt-and-generative-ai>
- [14] GPT4ALL, <https://gpt4all.io/index.html>
- [15] “6 Ways For Running A Local LLM (how to use HuggingFace)”, Tomas Fernandez, <https://semaphoreci.com/blog/local-llm>
- [16] “Authorship and AI Tools”, COPE Committee on Publications Ethics, <https://publicationethics.org/cope-position-statements/ai-author>
- [17] “The use of generative AI and AI-assisted technologies in writing for Elsevier”, Elsevier, <https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier>
- [18] “Back to School in the Generative AI Era”, Grammarly, <https://www.grammarly.com/blog/back-to-school-in-the-generative-ai-era/>
- [19] “Editorial Policies, Image and Text Integrity “, Science Journals, <https://www.science.org/content/page/science-journals-editorial-policies#image-and-text-integrity>
- [20] “Best Practice Guidelines on research Integrity and Publishing Ethics”, Artificial Intelligence Generated Content, Wiley Author Services, <https://authorservices.wiley.com/ethics-guidelines/index.html>

ANNEXE 1. Exemples de Conversations et de Tâches avec l'IA Générative

Les exemples détaillés dans cette annexe présentent des applications de l'IA générative dans le milieu universitaire. Ces conversations ont été générées par ChatGPT3.5 et, dans certains cas, par ChatGPT4.0. Dans les conversations, « Utilisateur » représente le prompt saisi par l'humain (surligné en bleu), et ChatGPT le résultat produit par l'outil d'IA générative. **Les données générées n'ont pas été validées, vérifiées ou contrôlées.**

Exemple 1 : Tuteur d'étudiants

Utilisateur (prompt). Je suis étudiant en première année d'Informatique. J'étudie la matière Introduction aux systèmes d'exploitation. J'en suis actuellement au chapitre consacré à la gestion du processeur. Je ne comprends pas la méthode FIFO. Explique-la-moi.

ChatGPT

La gestion du processeur à l'aide de l'algorithme FIFO (First-In-First-Out) est une méthode simple de programmation des opérations dans un système d'exploitation. En fait, cet algorithme traite les opérations dans la file d'attente de manière à ce que la première arrivée soit la première exécutée. Voici une explication plus détaillée du fonctionnement de la gestion du processeur avec la méthode FIFO :

9. File d'attente des opérations : Lorsque plusieurs opérations attendent d'être exécutées, elles sont placées dans une file d'attente. Cette file d'attente est une structure de données de type « premier entré, premier sorti » (FIFO).
10. Sélection des opérations : Lorsque le système d'exploitation doit sélectionner une opération à exécuter, il prend celle qui se trouve en tête de la file d'attente. Cette opération est celle qui est dans la file d'attente depuis le plus longtemps et qui est donc considérée comme la prochaine devant être exécutée.
11. Exécution des opérations : L'opération sélectionnée s'exécute sur le CPU jusqu'à ce que son exécution soit terminée ou interrompue pour une raison quelconque (par exemple, lorsqu'il se bloque en attente d'une entrée/sortie). Une fois que l'opération en cours se termine ou se bloque, l'opération suivante dans la file d'attente FIFO est sélectionnée pour s'exécuter.
12. Avantages et inconvénients : Le principal avantage de la méthode FIFO est sa simplicité. Elle est facile à mettre en œuvre et à comprendre. Cependant, ce n'est pas forcément la meilleure option en termes de performances, car elle ne tient pas compte de la priorité des opérations ni de leur utilisation en ressources.

Utilisateur. Qu'est-ce que la file d'attente ?

ChatGPT

La « file d'attente des opérations » fait référence à une structure de données utilisée dans les systèmes d'exploitation et la programmation simultanée pour garder une trace des opérations qui attendent d'être exécutées ou traitées par le système. En termes simples, il s'agit d'une liste ordonnée d'opérations qui attendent d'être traitées par l'unité centrale de calcul (CPU) ou d'autres ressources du système.

Chaque opération dans la file d'attente a un certain nombre de caractéristiques associées, comme son état actuel, sa priorité, son identifiant, son heure d'arrivée, son temps d'exécution restant et d'autres données pertinentes. La file d'attente des opérations est gérée par le système d'exploitation et est utilisée pour décider de quelle opération exécuter ensuite et de comment allouer le temps CPU.

Il existe différents algorithmes d'ordonnement des opérations qui déterminent comment une opération est choisie dans la file d'attente pour être exécutée. Certains de ces algorithmes intègrent la méthode FIFO

(First-In-First-Out), l'ordonnancement selon la priorité ou l'ordonnancement round-robin, entre autres. Chaque algorithme d'ordonnancement a ses propres règles et critères pour sélectionner l'opération suivante dans la file d'attente et lui allouer du temps CPU.

Utilisateur. Peux-tu me donner un exemple pratique d'une file d'attente FIFO ?

ChatGPT

Bien sûr, voici un exemple pratique du fonctionnement d'une file d'attente FIFO dans la gestion des opérations d'un système d'exploitation :

Supposons que tu aies trois opérations, A, B et C, qui veulent utiliser le CPU, et qu'elles soient mises en file d'attente pour être exécutées dans l'ordre dans lequel elles sont arrivées :

1. L'opération A arrive en premier et est placée dans la file d'attente.
2. L'opération B arrive ensuite et est placée dans la file d'attente derrière A.
3. L'opération C arrive en dernier et est placée dans la file d'attente derrière B.

À ce stade, la file d'attente des opérations FIFO ressemblerait à ceci :

File d'attente FIFO : A -> B -> C

Maintenant, le système d'exploitation doit décider de quelle opération exécuter ensuite en utilisant l'algorithme FIFO. Dans ce cas, comme la méthode FIFO suit le principe du « premier entré, premier sorti », l'opération A, qui est arrivée en premier, sera sélectionnée pour s'exécuter. Le système d'exploitation alloue du temps CPU à A pour qu'elle s'exécute.

Pendant que l'opération A s'exécute, les autres processus (B et C) restent dans la file d'attente. Une fois que l'opération A a terminé son exécution ou a été bloquée pour une raison quelconque (par exemple, attente d'une entrée/sortie), l'opération suivante dans la file d'attente FIFO, la B, sera sélectionnée pour être exécutée. L'opération C est exécuté après que l'opération B se soit terminée ou bloquée.

Exemple 2 : Générer des idées pour des projets de recherche

Utilisateur. Pour la matière Histoire de l'Espagne moderne, je dois faire un devoir. Je m'intéresse particulièrement aux relations de l'Espagne avec l'Asie, quels sujets me proposerais-tu pour le devoir ?

ChatGPT

Les relations de l'Espagne avec l'Asie pendant la période moderne sont un sujet fascinant et offrent une variété de sous-thèmes que tu pourrais explorer dans ton devoir. Voici quelques propositions de sujets que tu pourrais envisager :

1. Les Philippines sous la domination espagnole (1565-1898) :
 - La conquête et la colonisation des Philippines.
 - Les révoltes et la résistance des Philippins contre la domination espagnole.
 - L'influence culturelle et religieuse espagnole aux Philippines.
2. La Route des Galions de Manille :
 - L'importance économique et stratégique de cette route maritime entre Acapulco (Mexique) et Manille.
 - L'échange de marchandises entre l'Asie et l'Amérique par cette route.
3. Relations diplomatiques avec le Japon :
 - Les premiers contacts et missions de Jésuites au Japon.
 - Les persécutions des chrétiens et l'expulsion des missionnaires.
4. Relations avec la Chine :
 - Les missions franciscaines et jésuites en Chine.
 - Le commerce entre la Chine et l'Espagne via les Philippines.
5. La concurrence avec d'autres puissances européennes en Asie :
 - Les tensions et conflits avec des puissances comme le Portugal, la Hollande et l'Angleterre dans la région.
6. La fin de la domination espagnole en Asie :
 - Les causes et les conséquences de la perte des Philippines après la guerre hispano-américaine de 1898.
7. L'influence culturelle et l'héritage espagnol en Asie :
 - L'impact de la langue, de la religion et des traditions espagnoles dans les pays asiatiques, notamment aux Philippines.

Ce ne sont que quelques sujets généraux que tu pourrais envisager. En fonction de ton intérêt et de la longueur du devoir dont tu as besoin, tu peux te concentrer sur un seul de ces thèmes ou en explorer plusieurs ensemble. Bonne chance pour ton travail !

Utilisateur. Je dois me concentrer sur le XIX^e siècle, donne-moi trois sujets pour le devoir centrés sur cette période

ChatGPT

Bien sûr ! En se concentrant sur le XIX^e siècle et les relations de l'Espagne avec l'Asie, voici trois thèmes spécifiques :

1. Le déclin des Galions de Manille au XIX^e siècle :
 - Les raisons économiques, politiques et technologiques qui ont conduit à l'arrêt de cette route maritime en 1815.
 - Les conséquences pour le commerce entre l'Asie et l'Amérique et comment cela a affecté les colonies espagnoles sur les deux continents.
2. Révoltes et mouvements indépendantistes aux Philippines :
 - La montée des mouvements nationalistes et réformistes comme José Rizal et la Ligue Philippine.
 - La Révolution Philippine de 1896 et sa relation avec la Guerre hispano-américaine.
 - Le rôle des leaders révolutionnaires tels qu'Emilio Aguinaldo et Andrés Bonifacio.
3. La Guerre hispano-américaine et la perte des Philippines (1898) :
 - Les causes et le contexte du conflit avec les États-Unis.
 - Les principales batailles et événements aux Philippines pendant la guerre.
 - Les conséquences du Traité de Paris, qui a abouti à la cession des Philippines aux États-Unis et marqué la fin de l'empire colonial espagnol en Asie.

Ces sujets te permettront d'explorer en profondeur les relations de l'Espagne avec l'Asie au cours du XIX^e siècle, en te concentrant sur les événements et les changements les plus significatifs de cette période. J'espère qu'ils te seront utiles pour ton travail !

Utilisateur. Y avait-il d'autres territoires espagnols en Asie que les Philippines au XIX^e siècle ?

ChatGPT

Au cours du XIX^e siècle, les Philippines étaient le principal et le plus important territoire espagnol en Asie. Cependant, outre les Philippines, l'Espagne avait de petites possessions et des établissements ailleurs en Asie, même si leur contrôle et leur importance n'étaient pas comparables à ceux des Philippines. Parmi ces territoires et établissements, on retrouve notamment :

1. Les Îles Mariannes : Bien que ces îles aient été découvertes par Magellan au XVI^e siècle, l'Espagne a consolidé son contrôle sur elles au XVII^e siècle. Les Mariannes, dont Guam est l'île principale, étaient administrées comme une partie du territoire espagnol aux Philippines. Bien que Guam ait été cédée aux États-Unis après la Guerre hispano-américaine en 1898, les autres îles Mariannes ont été vendues à l'Allemagne en 1899.
2. Les Îles Carolines : L'Espagne a officiellement revendiqué ces îles au XIX^e siècle, bien qu'elle ait déjà établi des missions au XVII^e siècle. À la fin du XIX^e siècle, des tensions ont émergées avec l'Allemagne et le Royaume-Uni dû aux ambitions coloniales dans la région. Enfin, en 1899, l'Espagne a vendu les îles Carolines, ainsi que les îles Palau, à l'Allemagne.

3. Les Îles Palau : Ces îles, comme les Carolines, faisaient partie des possessions espagnoles dans le Pacifique et ont été vendues à l'Allemagne en 1899.
4. Établissements commerciaux en Chine : Bien que l'Espagne n'ait pas de colonies officielles en Chine, elle a établi une présence commerciale dans des ports tels que Canton. De plus, les missionnaires espagnols, principalement les jésuites, étaient déjà bien présents en Chine depuis plusieurs siècles.

En bref, si les Philippines étaient la possession la plus importante de l'Espagne en Asie au XIX^e siècle, il existait d'autres territoires et établissements dans le Pacifique et en Asie continentale. Cependant, à la fin du XIX^e siècle, l'Espagne avait perdu ou vendu la plupart de ces possessions.

Utilisateur. Propose-moi un sujet de devoir sur la présence commerciale de l'Espagne à Canton

ChatGPT

Bien sûr ! La présence commerciale de l'Espagne à Canton au cours du XIX^e siècle est un sujet intéressant et moins exploré que d'autres aspects de l'histoire coloniale espagnole. Voici un exemple de sujet détaillé pour ton devoir :

« Les échanges commerciaux entre l'Espagne et la Chine au XIX^e siècle : Le rôle de Canton en tant que porte d'entrée du marché chinois »

Dans le cadre de ce thème, tu pourrais aborder les points suivants :

1. Contexte historique :
 - Bref aperçu des relations commerciales entre l'Espagne et la Chine avant le XIX^e siècle.
 - L'importance de Canton en tant que principal port commercial de la Chine avec l'Occident.
2. Développement de la présence commerciale espagnole :
 - Les motivations économiques et politiques de l'Espagne pour établir des relations commerciales avec la Chine.
 - La création d'usines et d'établissements commerciaux espagnols à Canton.
3. Produits et biens échangés :
 - Les principaux produits exportés d'Espagne vers la Chine et inversement.
 - L'importance du commerce de l'argent et comment Canton est devenu un point clé pour le flux de ce métal précieux.
4. Défis et concurrence :
 - Les tensions et les défis rencontrés par les commerçants espagnols à Canton, notamment les problèmes avec les autorités locales et la concurrence avec les autres puissances européennes.
 - Les conséquences des Guerres de l'opium et la façon dont elles ont affecté la dynamique commerciale à Canton.
5. Conséquences et héritage :
 - L'impact économique et culturel du commerce entre l'Espagne et la Chine au XIX^e siècle.
 - L'influence et l'héritage de la présence commerciale espagnole à Canton sur les relations hispano-chinoises ultérieures.

Ce sujet te permettra d'explorer un aspect spécifique et moins connu de l'histoire commerciale de l'Espagne en Asie, offrant une perspective unique sur les interactions mondiales au XIX^e siècle. J'espère qu'il te sera utile et je te souhaite beaucoup de succès dans tes recherches !

Exemple 3 : Traiter et comprendre les articles scientifiques

Lis l'article suivant [ADRESSE HTML]. Ne me donne pas d'informations pour l'instant

ChatGPT

A utilisé Link Reader

J'ai lu avec succès l'article intitulé « Spectral Clustering for Sensing Urban Land Use using Twitter Activity ». Je garderai les informations en tête dans le cas où tu aurais des questions ou besoin de détails sur l'article. Fais-moi savoir si je peux t'aider davantage !

Utilisateur (source anonyme) Quel est le titre ?

ChatGPT

L'article est intitulé « Spectral Clustering for Sensing Urban Land Use using Twitter Activity ».

Utilisateur. Explique-moi ce que dit la section 4.1

ChatGPT

Bien sûr, la section 4.1 de l'article est intitulée « Twitter Datasets » et voici un résumé de son contenu :

- Twitter Datasets : Cette section décrit comment les données de Twitter ont été collectées et utilisées pour l'étude.
- Géolocalisation de Tweets : Les utilisateurs de Twitter peuvent marquer les tweets avec leur position géospatiale actuelle. Plus précisément, ils peuvent définir leur emplacement géographique en spécifiant une ville ou une région ou en permettant à Twitter de suivre leurs coordonnées GPS de longitude et de latitude. Lorsqu'un nouveau tweet est écrit, Twitter enregistre les informations géographiques de l'utilisateur à ce moment-là, ainsi que plusieurs autres métadonnées. Pour modéliser un schéma d'utilisation par zone dans un environnement urbain, ils n'ont recueilli que les tweets dont la localisation est automatiquement enregistrée par Twitter via le GPS et non signalée par l'utilisateur.
- Collecte des données : Ils ont utilisé l'API de streaming de Twitter pour collecter des tweets géolocalisés en temps quasi réel. Cette API permet d'établir un flux très performant avec Twitter grâce auquel un grand nombre de positions publiques de tweets peut être collecté. Plus précisément, l'API de streaming de Twitter affiche toutes les positions publiques de tweets, soit actuellement environ un pour cent de l'ensemble des tweets de ce flux de données. Leur ensemble final de données Twitter se compose de 49 jours (sept semaines) de tweets géolocalisés provenant du monde entier, du 25 octobre au 12 décembre 2010.
- Zones géographiques sélectionnées : L'aire géographique de Londres est définie par la zone délimitée par ce qui aurait été le Ringway 1. Pour Madrid, ils considèrent approximativement la zone urbaine située à l'intérieur de l'autoroute M-30.

Ceci est un résumé de la section 4.1 de l'article. Si tu as besoin de plus de détails ou d'informations sur d'autres sections, fais-le moi savoir !

Utilisateur. Je ne comprends pas bien la méthode utilisée dans la section 4.2. Explique-la-moi dans un langage simple, s'il te plaît.

ChatGPT

A utilisé Link Reader

La section 4.2 de l'article est intitulée « Land Segmentation and Land Use Clustering ». Voici une explication simplifiée de cette section :

La section 4.2 se concentre sur la façon dont l'environnement urbain est segmenté en différents secteurs et sur la façon dont ces derniers sont regroupés en fonction de leur activité sur Twitter afin d'identifier les différentes utilisations de la zone.

1. Segmentation d'une Zone avec des Tweets Géolocalisés :

- La méthode utilise un type de réseau neuronal appelé cartes auto-organisatrices (SOM) pour diviser la zone urbaine en différents secteurs en fonction de l'activité de Twitter.
- Ces secteurs sont définis en fonction de la concentration de tweets dans les différentes zones. Par exemple, une zone avec beaucoup de tweets aura plus de segments (ou de neurones) qu'une zone avec moins de tweets.
- Pour déterminer la taille et la forme de ces secteurs, la méthode explore différentes tailles de cartes et sélectionne celle qui correspond le mieux à la distribution des tweets dans la zone urbaine.

2. Groupement d'Utilisation des Zones :

- Une fois les secteurs définis, chacun est caractérisé par son « vecteur d'activité de tweet », qui est essentiellement un résumé de quand et combien de tweet ont été publiés dans ce secteur.
- Ces vecteurs d'activité sont utilisés pour regrouper les secteurs similaires. L'idée est que les secteurs ayant des modèles d'activité Twitter similaires sont susceptibles d'avoir des utilisations similaires.
- Pour effectuer ce regroupement, la méthode utilise une technique appelée « partitionnement spectral ». Cette technique tente de regrouper les secteurs qui sont similaires les uns aux autres et différents des autres groupes.
- Le nombre de groupes est déterminé par une technique appelée « détection d'eigengap ».

3. Interprétation des Groupes :

- Une fois les groupes formés, ils peuvent être interprétés pour identifier les différentes utilisations des zones. Par exemple, un groupe qui a beaucoup d'activité Twitter la nuit pourrait représenter une zone de loisirs ou de vie nocturne.
- L'emplacement des groupes dans la zone urbaine permet également d'interpréter leur utilisation. Par exemple, un groupe situé près d'un parc pourrait représenter un zone de loisirs.

En résumé, la section 4.2 décrit comment l'activité de Twitter peut être utilisée pour segmenter une zone urbaine en différents secteurs et comment ces derniers peuvent être regroupés pour identifier les différentes utilisations des zones. L'idée est que la façon dont les gens tweetent dans différentes zones peut nous donner des indices sur la façon dont cette zone est utilisée dans la vie réelle.

Utilisateur. Décris-moi quelles sont les groupes identifiés dans la section 4.2

ChatGPT

A utilisé PDF.AI

La section 4.2 identifie plusieurs groupes en fonction de l'activité de Twitter dans différentes zones urbaines. Ces groupes représentent différents types d'utilisation pour chaque secteur. Voici les groupes identifiés :

1. Type de groupe 1 : Entreprises (Business)

Ce groupe montre une activité sur Twitter constante pendant la journée et diminuant la nuit. Géographiquement, il couvre des zones associées à des activités commerciales/de bureaux, telles que le quartier financier de Manhattan, la City et Canary Wharf à Londres, et les quartiers Castellana et AZCA à Madrid.

2. Type de groupe 2 : Loisirs/Weekend (Leisure/Weekend)

Ce groupe montre une grande différence entre l'activité sur Twitter le week-end et pendant de la semaine. Géographiquement, il couvre des régions telles que Central Park à Manhattan, Hyde Park à Londres et le parc du Retiro à Madrid. Il comprend également des zones touristiques.

3. Type de groupe 3 : Vie nocturne (Nightlife)

Ce groupe a une activité sur Twitter qui augmente considérablement la nuit, en particulier le week-end. Géographiquement, il couvre des zones associées à la vie nocturne, telles que le West End à Londres et Malasaña/Chueca à Madrid.

4. Type de groupe 4 : Résidentiel (Residential)

Ce groupe montre une activité sur Twitter constante pendant le week-end et atteignant son maximum les soirs de semaine. Géographiquement, il couvre des zones principalement résidentielles dans les trois villes.

5. Type de groupe 5 : Industriel (seulement identifié pour Londres)

Ce groupe a une activité sur Twitter qui est la plus élevée les matins de semaine et qui diminue le reste de la journée. Géographiquement, il couvre des zones de l'est et du sud de Londres, comme la zone autour de la gare de Battersea et le parc olympique.

Ces groupes ont été identifiés en analysant la « signature d'activité Twitter » de chaque groupe, qui est un résumé de quand et combien de tweets ont été publiés dans ce groupe, et en la comparant à la répartition géographique des groupes sur la carte de la ville.

Il est important de noter que ces groupes représentent des hypothèses d'utilisation des différents secteurs d'une ville en fonction de l'activité sur Twitter, et qu'ils sont interprétés en fonction de leur activité et de leur emplacement dans la zone urbaine.

Exemple 4 : Revue de l'état de l'art

Utilisateur. I am writing a review of the state of the art on the topic "machine learning applications in higher education". Please find other scientific papers that have already presented a review on this topic.

ChatGPT

Certainly! Reviewing the state of the art on "machine learning applications in higher education" is a comprehensive task, and numerous researchers have written about this. Here are some papers and articles from before my last update in 2022 that could be pertinent to your research:

1. Huda, M., Maseleno, A., Atmotiyoso, P., Siregar, M., Ahmad, R., Jasmi, K. A., & Muhamad, N. (2018). Big data emerging technology: Insights into innovative environment for online learning resources. *International Journal of Emerging Technologies in Learning*, 13(1), 23-36.
2. Dutt, R., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5, 15991-16005.
3. Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data—evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63-71.
4. Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira Eds., *Proceedings of 5th Annual Future Business Technology Conference*.
5. Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414-422.
6. Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17.
7. Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting students drop out: A case study. In *2nd International Conference on Educational Data Mining*.
8. Ali, A., & Smith, A. (2015). Review of machine learning approaches to predict learning outcomes. in *Proc. Science and Information Conference (SAI)*.
9. Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, 1-15.
10. Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. In *Proceedings of the 11th International Conference on Data Engineering*.

Remember that while these articles can provide an overview and insights on the topic, it is essential to access each paper and evaluate its relevance to your specific focus within the broader topic. Additionally, considering the rapid advancement of technology and methodologies, it's crucial to look for more recent articles and reviews (post-2022) to ensure your review is comprehensive and up-to-date.

Utilisateur. Just give me references from 2020 or newer

ChatGPT

Certainly, focusing on more recent work is a great idea to ensure that your review is up-to-date. Here are some articles and papers from 2020 onwards that delve into machine learning applications in higher education:

1. Alghamdi, A., Karpinski, A. C., Lepp, A., & Barkley, J. (2020). Data mining in higher education: University student dropout case study. *International Journal of Information and Learning Technology*.
2. Jiang, S., Williams, A. E., Schenke, K., Warschauer, M., & O'Dowd, D. K. (2020). Predicting MOOC performance with week 1 behavior. *Journal of Computer Assisted Learning*.
3. Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., & Ramírez-Quintana, M. J. (2020). From predicting educational achievement to identifying learning strategies: A meta-analysis. *User Modeling and User-Adapted Interaction*.
4. Gogoulou, A., Gouli, E., Boas, G. V., & Liakou, E. (2020). Exploiting machine learning for predicting student's performance. *Journal of Educational Computing Research*.
5. Vieira, C., Parsons, P., & Byrd, V. (2020). Data Mining Techniques on Students' Learning Behavior. In *Computer Supported Education*. Springer.
6. Shen, L., Wang, M., & Shen, R. (2020). Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment. *Educational Technology & Society*.
7. Khan, I. A., Alotaibi, F. M., & Ahmad, F. (2021). A predictive model using machine learning for improving higher education enrollment management. *Education and Information Technologies*.
8. Tan, C., & Shao, P. (2021). Using artificial intelligence to support higher education learning: Possibilities and challenges. *British Journal of Educational Technology*.
9. Almarabeh, H., & Mohammad, H. (2021). Machine Learning Techniques in E-Learning Systems: A Systematic Review. *Education and Information Technologies*.
10. Ramírez-Quintana, M. J., Martínez-Plumed, F., Ferri, C., & Hernández-Orallo, J. (2021). Machine Learning for Student Modelling and Adaptive Education: A Survey. *Artificial Intelligence Review*.

To ensure a thorough and comprehensive review, I recommend accessing these papers and analyzing their methodologies, findings, and relevance to your specific area of interest within the topic of machine learning in higher education. Additionally, don't forget to also consider more recent publications from 2022 onwards, which may not be covered here.

Utilisateur. Considering these references, write an introduction that uses them

ChatGPT

Certainly! Here's an introduction that references the provided articles:

Introduction. The application of machine learning in higher education has emerged as a transformative force, shaping educational paradigms, predicting student outcomes, and personalizing learning experiences. Recent years have witnessed an upsurge in research seeking to harness the potential of these intelligent algorithms for various aspects of higher education. For instance, Alghamdi et al. (2020) delved into the arena of data mining, presenting a pivotal study on the prediction of student dropout rates in universities. Similarly,

highlighting the potential of early behavior in Massive Open Online Courses (MOOCs), Jiang et al. (2020) posited that the initial week's behavior could be instrumental in predicting overall MOOC performance.

Yet, the implications of machine learning in education are not confined to just prediction. The work of Martínez-Plumed et al. (2020) ushered in a novel perspective by transitioning from mere predictive analyses to identifying effective learning strategies through meta-analyses. Complementing this, Gogoulou et al.'s (2020) study further substantiated the value of machine learning, specifically in augmenting students' performance in e-learning environments.

As e-learning platforms become increasingly sophisticated, the role of emotional data in shaping learning experiences cannot be understated. Shen et al. (2020) championed this cause, emphasizing the importance of affective e-learning and how "emotional" data can be harnessed to foster better learning in pervasive educational environments.

The administrative aspects of higher education have not remained untouched either. Khan et al. (2021) crafted a predictive model leveraging machine learning, aiming to enhance the efficiency of higher education enrollment management. This intertwining of administration and machine learning underscores the holistic impact of such technologies on the academic realm.

Furthermore, as Tan and Shao (2021) posit, while the potential of artificial intelligence, including machine learning, in supporting higher education learning is vast, it is not devoid of challenges. Their comprehensive overview provides a balanced viewpoint, elucidating both the possibilities and hurdles of integrating AI into higher education. Complementing this, both Almarabeh & Mohammad (2021) and Ramírez-Quintana et al. (2021) offer exhaustive reviews on machine learning techniques in e-learning systems and the broader spectrum of student modeling and adaptive education.

Exemple 5 : Construire un programme et une présentation

Utilisateur. Je prépare une matière pour la dernière année d'ingénierie informatique. Je veux préparer un programme pour la matière « Introduction au Deep Learning ». Le cours se déroule sur 14 semaines, à raison de 2 leçons par semaine de 90 minutes. Les étudiants ont déjà des connaissances en machine learning et en programmation Python. Le programme doit inclure des exemples pratiques sur Python et 2 projets à rendre, le premier en semaine 7 et le dernier en semaine 14.

ChatGPT

Quel projet passionnant ! Le Deep Learning est l'un des domaines les plus fascinants de l'intelligence artificielle. En me basant sur la structure que tu as donnée, je te propose le programme suivant pour ton cours « Introduction au Deep Learning » :

Semaine 1 : Introduction au Deep Learning

- Histoire et motivation du Deep Learning.
- Différences entre le Machine Learning et le Deep Learning.

Semaine 2 : Réseaux Neuronaux Artificiels

- Structure et fonctionnement.
- Application et exemples pratiques en utilisant TensorFlow/Keras.

Semaine 3 : Fonctions d'Activation et de Rétropropagation

- Fonctions d'activation : ReLU, Sigmoid, Tanh, etc.
- L'algorithme de rétropropagation.

Semaine 4 : Régularisation et Optimisation

- Méthodes de régularisation : Abandon, régularisation L1 et L2.
- Optimiseurs : SGD, Adam, RMSprop, etc.

Semaine 5 : Réseaux Neuronaux Convolutifs (CNN)

- Introduction et applications.
- Application des CNN avec des exemples pratiques.

Semaine 6 : Réseaux Neuronaux Récurrents (RNN)

- Concept et applications.
- LSTM et GRU : application et exemples.

Semaine 7 : Techniques avancées et projet 1

- Batch normalization, Transfer learning, Data Augmentation.
- Présentation et développement du premier projet.

Semaine 8 : Autoencodeurs

- Introduction et applications.
- Application et exemples pratiques.

Semaine 9 : Réseaux Antagonistes Génératifs (GAN)

- Concept et applications.
- Application de GAN basiques.

Semaine 10 : Attention et Mécanismes de l'Attention

- Introduction au concept d'attention dans les réseaux neuronaux.
- Transformers et BERT : aperçu général.

Semaine 11 : Deep Learning pour le Traitement Automatique des Langues (NLP)

- Word embeddings : Word2Vec, GloVe.
- Applications de modèles NLP en utilisant des RNN et des Transformers.

Semaine 12 : Deep Learning dans la Vision par Ordinateur

- Détection d'objets, Segmentation sémantique.
- Application d'exemples pratiques.

Semaine 13 : Apprentissage par Renforcement

- Concepts basiques et applications.
- Applications d'exemples simples avec DQN.

Semaine 14 : Révisions et Projet 2

- Révision générale et réponses aux doutes.
- Présentation et développement du deuxième projet.

Projets :

- Projet 1 (semaine 7) : Classification d'images à l'aide de CNN. Les élèves peuvent travailler avec des datasets comme CIFAR-10 ou Fashion MNIST.
- Projet 2 (semaine 14) : Choix entre :
 - Création d'un modèle NLP pour l'analyse des sentiments.
 - Développement d'un GAN pour la génération d'images.
 - Mise en œuvre d'un modèle d'apprentissage par renforcement dans un environnement basique, comme un jeu simple.

Ce programme couvre un large éventail de sujets sur le deep learning et devrait donner aux étudiants une solide introduction au domaine. J'espère qu'il te sera utile et je te souhaite beaucoup de succès dans tes cours !

Utilisateur. Les étudiants ont déjà des matières spécifiques en NLP et en vision par ordinateur. Supprime ces deux semaines et ajoute du contenu supplémentaire aux semaines 6 et 9

ChatGPT

D'accord ! Étant donné que les étudiants ont déjà de l'expérience en NLP et en vision par ordinateur, nous pouvons approfondir des sujets spécifiques liés aux Réseaux Neuronaux Récurrents (RNN) et aux Réseaux

Antagonistes Génératifs (GAN), et éventuellement introduire des sujets avancés qui seraient d'un grand intérêt. Voici le programme ajusté :

Semaine 1 : Introduction au Deep Learning

- Histoire et motivation du Deep Learning.
- Différences entre le Machine Learning et le Deep Learning.

Semaine 2 : Réseaux Neuronaux Artificiels

- Structure et fonctionnement.
- Application et exemples pratiques en utilisant TensorFlow/Keras.

Semaine 3 : Fonctions d'Activation et de Rétropropagation

- Fonctions d'activation : ReLU, Sigmoid, Tanh, etc.
- L'algorithme de rétropropagation.

Semaine 4 : Régularisation et Optimisation

- Méthodes de régularisation : Abandon, régularisation L1 et L2.
- Optimiseurs : SGD, Adam, RMSprop, etc.

Semaine 5 : Réseaux Neuronaux Convolutifs (CNN)

- Introduction et applications.
- Application des CNN avec des exemples pratiques.

Semaine 6 : Réseaux Neuronaux Récurrents (RNN) Avancés

- LSTM et GRU : application et exemples.
- Applications avancées des RNN : génération de textes, séries temporelles et musique.

Semaine 7 : Techniques avancées et projet 1

- Batch normalization, Transfer learning, Data Augmentation.
- Présentation et développement du premier projet.

Semaine 8 : Autoencodeurs

- Introduction et applications.
- Application et exemples pratiques.

Semaine 9 : Réseaux Antagonistes Génératifs (GAN) Avancés

- GAN conditionnels, les CycleGAN et les StarGAN.
- Exemples de GAN avancés et applications innovantes.

Semaine 10 : Attention et Mécanismes de l'Attention

- Introduction au concept d'attention dans les réseaux neuronaux.
- Transformers : structure et applications.

Semaine 11 : Deep Learning et Graphes

- Réseaux Neuronaux en Graphes (GNN).
- Applications et exemples pratiques.

Semaine 12 : Apprentissage par Renforcement

- Concepts basiques et applications.
- Applications d'exemples simples avec DQN.

Semaine 13 : Modèles Génératifs et Audio

- Introduction à la génération d'audios avec des modèles profonds.
- WaveNet et autres modèles pour la génération de paroles et de musiques.

Semaine 14 : Révisions et Projet 2

- Révision générale et réponses aux doutes.
- Présentation et développement du deuxième projet.

Projets :

- Projet 1 (semaine 7) : Classification d'images à l'aide de CNN. Les élèves peuvent travailler avec des datasets comme CIFAR-10 ou Fashion MNIST.
- Projet 2 (semaine 14) : Choix entre :
 - Génération de texte en utilisant des RNN avancés.
 - Développement d'un GAN avancé pour une application innovante.
 - Mise en œuvre d'un modèle d'apprentissage par renforcement dans un environnement basique, comme un jeu simple.

Avec ces ajustements, les RNN et les GAN sont approfondis, en plus d'introduire quelques sujets supplémentaires qui peuvent être très intéressants pour les étudiants ayant déjà des connaissances en deep learning. J'espère que cette proposition te sera utile !

Utilisateur. Quelle bibliographie essentielle recommanderais-tu ?

ChatGPT

Bien entendu, pour un cours « Introduction au Deep Learning », il existe plusieurs textes fondamentaux qui sont fortement recommandés et qui couvrent de manière exhaustive les concepts clés de ce domaine. Voici la bibliographie essentielle recommandée :

1. « Deep Learning » par Ian Goodfellow, Yoshua Bengio et Aaron Courville.

Il s'agit probablement du livre le plus complet et le plus essentiel sur le deep learning. Il traite des fondements mathématiques jusqu'aux sujets avancés comme les autoencodeurs, les GAN et l'apprentissage semi-supervisé.

2. « Python Deep Learning » par Ivan Vasilev, Daniel Slater et Gianmario Spacagna.

Ce livre est excellent pour ceux qui veulent apprendre à mettre en place des modèles de deep learning avec Python. Il explique TensorFlow, Keras et d'autres frameworks populaires.

3. « Réseaux Neuronaux et Deep Learning : A Textbook » par Charu Aggarwal.

Il offre une compréhension approfondie des réseaux neuronaux et du deep learning. Charu Aggarwal a une façon d'expliquer facile à suivre, qui convient aussi bien aux débutants qu'aux experts.

4. « Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow » par Aurélien Géron.

Bien qu'il couvre également le machine learning en général, il comporte des sections dédiées au deep learning. Il est très pratique et orienté sur la pratique, ce qui le rend parfait pour un cours qui cherche à combiner la théorie avec des exemples concrets sur Python.

5. « Deep Learning pour la Vision par Ordinateur » par Rajalingappaa Shanmugamani.

Comme tu vas aborder des sujets liés aux CNN et aux applications de vision par ordinateur, ce livre peut être utile comme ressource complémentaire pour ces sujets spécifiques.

6. « Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play » par David Foster.

Excellent pour approfondir des sujets comme les GAN, les autoencodeurs et d'autres modèles génératifs.

7. « Deep Reinforcement Learning Hands-On » par Maxim Lapan.

Pour la partie sur l'apprentissage par renforcement, ce livre est l'une des ressources les plus complètes et les plus pratiques qui existe.

N'oublie pas que le monde du deep learning est en constante évolution. Bien que ces livres fournissent une base solide, il est toujours bon de compléter avec des articles de recherche, des tutoriels et des blogs en ligne pour se tenir au courant des dernières techniques et avancées. J'espère que cette liste te sera utile et je te souhaite beaucoup de succès dans ton cours !

Utilisateur. Rappelle-moi la semaine 3

ChatGPT

Bien sûr, voici la semaine 3 que j'ai proposée dans le programme :

Semaine 3 : Fonctions d'Activation et de Rétropropagation

- Fonctions d'activation : ReLU, Sigmoid, Tanh, etc.
- L'algorithme de rétropropagation.

Cette semaine, les étudiants découvriront les différentes fonctions d'activation utilisées dans les réseaux neuronaux et la façon dont elles affectent la propagation de l'information dans le réseau. De plus, l'algorithme de rétropropagation, essentiel pour l'entraînement des réseaux neuronaux, serait introduit, permettant d'ajuster les poids en fonction des erreurs calculées.

Utilisateur. Prépare-moi une présentation pour le cours 1

ChatGPT

Malheureusement, je ne peux pas créer de présentations directement ici, mais je peux te donner un plan détaillé que tu pourras utiliser pour construire ta présentation sur les fonctions d'activation dans le contexte du deep learning :

Présentation : Fonctions d'Activation dans des Réseaux Neuronaux

- Diapositive 1 : Titre
 - Titre : Fonctions d'Activation dans des Réseaux Neuronaux
 - Sous-titre : Semaine 3 - Cours 1
 - Image : Graphique ou image en lien avec les réseaux neuronaux.
- Diapositive 2 : Introduction
 - Définition : Les fonctions d'activation déterminent la sortie d'un nœud ou d'un neurone dans un réseau neuronal.
 - Importance : Ils sont essentiels pour ajouter de la non-linéarité au modèle, ce qui permet au réseau d'apprendre des modèles complexes.
- Diapositive 3 : Types de fonctions d'activation
 - Liste générale des fonctions d'activation à aborder :
 - Sigmoidé
 - Tanh (Tangente hyperbolique)
 - ReLU (Rectified Linear Unit)
 - (D'autres, si tu décides d'en inclure davantage).
- Diapositive 4 : Fonction sigmoïde
 - Formule : $S(x) = \frac{1}{1+e^{-x}}$
 - Graphique : Courbe sigmoïde
 - Avantages et inconvénients :
 - Plage de sortie entre 0 et 1.
 - Problème de disparition du gradient.
- Diapositive 5 : Fonction Tanh
 - Formule : $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
 - Graphique : Courbe Tanh
 - Avantages et inconvénients :
 - Plage de sortie entre -1 et 1.
 - Bien qu'elle puisse toujours provoquer une disparition du gradient, elle est centrée sur le zéro, ce qui la rend préférable à la fonction sigmoïde dans de nombreux cas.
- Diapositive 6 : Fonction ReLU
 - Formule : $f(x) = \max(0, x)$
 - Graphique : Graphique représentant la partie positive de x et zéro pour des valeurs négatives.
 - Avantages et inconvénients :
 - Aide à limiter la disparition du gradient.
 - Elle peut provoquer des « neurones morts » où certains nœuds n'activent pas.
- Diapositive 7 : Autres Fonctions d'Activation
 - Brève mention des autres fonctions, si tu décides de les inclure, par exemple :
 - Leaky ReLU
 - Swish
 - Softmax (particulièrement utilisée en sortie de couche pour classification).

- Diapositive 8 : Comment Choisir une Fonction d'Activation
 - En règle générale : ReLU est une bonne fonction pour commencer.
 - Problèmes de Classification Binaire : Sigmoidé en sortie de couche.
 - Classification multiclasse : Softmax en sortie de couche.
 - À prendre en considération : la nature du problème, l'architecture du réseau et l'expérimentation.

Avec ce schéma, tu peux utiliser des outils comme PowerPoint, Google Slides ou tout autre outil de présentation de ton choix pour créer les diapositives. Ajoute des images, des graphiques et des animations comme tu le souhaites pour rendre la présentation plus interactive et attrayante. Bonne chance !

Utilisateur. Que dirais-tu pour la diapositive 4 ?

ChatGPT

Pour la diapositive 4 : Fonction sigmoïde, je te fournis un script détaillé que tu peux suivre lors de la présentation :

[Montrer le titre et la formule]

« Passons maintenant à l'une des fonctions d'activation les plus connues : la fonction sigmoïde. »

[Montrer le graphique de la courbe sigmoïde]

« Comme vous pouvez le voir, la fonction sigmoïde prend n'importe quel nombre réel et le "réduit" (par normalisation) dans une plage comprise entre 0 et 1. Cette fonction est utile parce qu'elle convertit les grands ou les petits nombres en valeurs gérables. Nous pouvons voir que sa courbe est en forme de S »

[Montrer les avantages et les inconvénients]

« Cependant, toutes les utilisations ne conviennent pas à la fonction sigmoïde. Bien qu'elle soit excellente pour certains types de problèmes, notamment pour la classification binaire sur la couche de sortie, elle présente des inconvénients. L'un des principaux problèmes est la disparition du gradient. Qu'est-ce que cela signifie ? Si vous regardez les zones où la courbe est presque plate, vous pouvez imaginer que le gradient est très faible. Cela peut rendre les mises à jour insignifiantes pendant l'entraînement, ce qui peut ralentir la convergence ou, dans certains cas, arrêter l'apprentissage »

« Il est important de garder en tête qu'il n'existe pas de fonction d'activation unique qui fonctionne pour tous les scénarios. Le choix dépendra du problème, de l'architecture du réseau et souvent de l'expérimentation »

Ce texte donne un aperçu général de la fonction sigmoïde, en mentionnant ses principales caractéristiques et ses problèmes. Pendant la présentation, il est essentiel d'interagir avec les élèves, peut-être en posant des questions pour s'assurer qu'ils ont bien compris ou pour voir s'ils ont déjà une expérience du sujet.