

Exercice 1 : Expérimenter des requêtes efficaces

L'art de concevoir une requête est désigné par le terme de "*prompt engineering*" (rédactique ou ingénierie des invites).

Il n'existe pas de recette unique pour concevoir une requête efficace, seulement des bonnes pratiques, certaines proposées par les constructeurs de modèles et d'agents conversationnels [1, 2, 3, 4].

[1] [OpenAI ChatGPT prompt engineering best practices](#)

[2] [Anthropic Claude prompt engineer overview](#)

[3] [Google Gemini prompting strategies](#)

[4] [Google Vertex AI for Google Cloud stratégies de requêtes](#)

Ces guides décrivent généralement les composants possibles que l'on peut retrouver dans une requête et des stratégies que l'on peut appliquer.

Le [tableur fourni](#) sur madoc présente

- dans son premier onglet la liste des composants que l'on peut spécifier dans une requête
- puis dans un second onglet fait une synthèse des stratégies possibles pour concevoir une requête

Partie 1 (15 minutes, travail individuel)

- A) Votre première tâche est de prendre connaissance des composants et des stratégies recensés dans le tableur.
- B) Votre deuxième tâche consiste à vous demander si toutes les « bonnes pratiques » de conception de requêtes sont raccord avec un objectif de sobriété écologique ?

Partie 2 (30 minutes, travail en duo ou individuellement)

Pour les questions suivantes, considérer que votre objectif est d'« apprendre à rédiger une requête efficace » et que vous allez utiliser les capacités des agents conversationnels pour vous aider.

Utiliser les agents conversationnels suivants [gemini](#), [claude](#) et [lechat](#)

- C) D'abord, expérimentez une requête naïve identique pour les 3 agents.
Par exemple « Je souhaite apprendre à rédiger une requête efficace »
Observez le contenu généré. Le jugez-vous pertinent ? Pourrait-il répondre à votre besoin ?

D) En duo ou individuellement, fournir une “clear definition of the success criteria for your use case. Some ways to empirically test against those criteria”

E) en travaillant d'abord « sur le papier », améliorez cette requête de départ afin qu'elle respecte les bonnes pratiques de la rédaction. Écrivez directement dans le tableau.

Composant	Votre proposition
Persona	
Objectif	
Instructions	
Instructions système/contraintes	
Exemples (zero/few-shots)	
Format de réponse	
Ton	
Contexte	
Étapes de raisonnement	

F) Choisissez un agent conversationnel et expérimentez votre requête améliorée dans une nouvelle conversation. Au besoin, vous pouvez faire quelques relances pour obtenir un résultat satisfaisant. Expérimentez votre résultat final dans les autres agents.

G) Une fois la conversation terminée, en plénière :

- a. Qu'est-ce que la réponse obtenue après amélioration apporte de plus que la première réponse ?
- b. Est-ce que la nouvelle réponse est plus proche de ce dont vous aviez réellement besoin ? Pourquoi ?
- c. Est-ce que vous avez gardé le contrôle sur ce que vous vouliez obtenir, ou est-ce l'agent conversationnel qui a décidé à votre place ?
- d. Quels retours d'expérience sur la conception des requêtes ? Sur le fonctionnement des agents considérés ?

Partie 3 (10 minutes, restitution en plénière)

Consignes aux enseignants : Revenir sur la dernière question. Indiquer qu'il n'y a pas de recette unique, mais de bonnes pratiques. L'important étant de faire preuve d'observation et de curiosité intellectuelle pour répéter les pratiques qui donnent de bons résultats et de ne pas persévérer dans celles qui ne répondent pas aux besoins.

Exercice 2 : Coût de requêtes

Consignes à l'enseignant : Constituer six groupes d'étudiants de 3 ou 4 étudiants et distribuer à chacun des groupes une "requête".

Requêtes proposées

1 Donner la définition de ce terme : (ex¹. : infatué, amphigourique, croquignolesque, s'ébaudir, garrulité, pétrichor, etc.)

2 Faire une liste de séries à regarder sur Netflix.

3 Créer un plan d'étude pour réussir mes examens.

4 Donner 5 idées de repas simples pour étudiant·es.

5 Expliquer le concept scientifique suivant : (ex. : la photosynthèse, l'acculturation, le pouvoir, etc.)

6 Résumer ce texte en 100 mots : « Tu as sûrement déjà utilisé des outils d'intelligence artificielle (IA) comme ChatGPT pour effectuer des recherches ou générer du texte. L'utilisation de l'IA soulève de sérieuses questions éthiques et de fiabilité des données. Mais sais-tu que ces technologies ont aussi un impact considérable sur l'environnement ? Une requête avec l'outil d'IA ChatGPT, par exemple, nécessite environ 10 fois plus d'énergie qu'une requête classique sur le moteur de recherche Google. Comment fonctionnent ces requêtes ? Elles s'appuient sur des centres de données qui contiennent des centaines de serveurs. Ces derniers assurent le fonctionnement de tous les services en ligne que nous utilisons au quotidien : réseaux sociaux, streaming, stockage de fichiers (cloud), jeux en ligne, IA, etc. Pour cela, les serveurs fonctionnent en continu et doivent être alimentés en permanence en électricité. Les serveurs d'IA sont plus énergivores que les serveurs classiques, car ils traitent un nombre plus important de données et ont besoin d'une plus grande puissance de calcul. Par exemple, en 2023, les 200 millions d'utilisateurs hebdomadaires de ChatGPT ont utilisé la même quantité d'électricité que plus de 43 200 foyers américains ! L'IA contribue ainsi considérablement à l'augmentation de la consommation mondiale d'énergie. Au-delà de l'électricité nécessaire pour alimenter les centres de données, il faut aussi beaucoup d'eau pour les refroidir, car ils fonctionnent en continu et dégagent beaucoup de chaleur. Imagine : pour générer un simple courriel de 100 mots avec ChatGPT-4, il faut jusqu'à 1,4 litre d'eau ! Ces chiffres sont inquiétants, d'autant plus qu'ils sont à multiplier par le nombre d'utilisateurs dans le monde. Il est important d'être conscient de son impact. Chaque requête, chaque courriel généré par IA a un coût pour la planète. En utilisant ces outils de manière

¹ Termes extraits de : Éditions Maison des langues. (2024). *10 mots insolites de la langue française*.
<https://www.emdl.fr/lettres/dernieres-actualites/10-mots-insolites-de-la-langue-francaise>

réfléchie, nous pouvons contribuer à limiter leur impact environnemental. Moins on s'en sert, mieux la planète se porte. » (Québec Science, 2024)

Partie 1 (15 minutes, travail en individuel)

- A) Individuellement, allez sur le site <https://www.comparia.beta.gouv.fr>.
- B) Cliquez sur « Commencez à discuter »
 - Puis, utilisez la requête assignée par votre enseignant.e.
 - Ensuite, choisissez le modèle « David contre Goliath ».
- C) Une fois les réponses des modèles A et B obtenues, répondez à la courte question et « Passez à la révélation des modèles ».
- D) Remplissez le premier tableau synthèse à la page suivante (au verso).
- E) Refaites le test avec les modèles « Frugal » et « Sélection Manuelle » (pour ce choix, choisissez votre modèle), tout en gardant la même requête.
- F) Remplissez les deux autres tableaux de la page suivante.

TABLEAU SYNTHÈSE DU MODE DAVID CONTRE GOLIATH

Requête testée pour les 3 modes :

	Modèle A : _____	Modèle B : _____
Préférence		
Énergie consommée en Wh		
Équivalent CO2		
Équivalent ampoule LED		
Équivalent vidéo en ligne		

TABLEAU SYNTHÈSE DU MODE FRUGAL

	Modèle A : _____	Modèle B : _____
Préférence		
Énergie consommée en Wh		
Équivalent CO2		
Équivalent ampoule LED		
Équivalent vidéo en ligne		

TABLEAU SYNTHÈSE POUR LE MODE SÉLECTION MANUELLE

	Modèle A : _____	Modèle B : _____

Préférence		
Énergie consommée en Wh		
Équivalent CO2		
Équivalent ampoule LED		
Équivalent vidéo en ligne		

Partie 2 (5-10 minutes, restitution en plénière)

Exercice 3 : Exploiter un environnement RAG avancé

Un RAG (Génération Augmentée de Récupération) est une solution technique qui combine l'extraction d'information d'un moteur de recherche classique avec les capacités langagières d'un LLM.

Pour cet exercice, nous allons utiliser [Google notebooklm](https://notebooklm.google.com/). Notebooklm est un outil permet d'importer des documents (site web, PDF, Drive, youtube) à partir desquels il crée des résumés, FAQ, timelines, plan, répond aux questions avec des citations, fournit des explications/exemples, dégage des tendances, génère de nouvelles idées.

Partie 1 (20 minutes, travail en individuel)

- A) Connectez-vous à NotebookLM <https://notebooklm.google.com/>
- B) Créez un nouveau Notebook.
- C) Réalisez les missions suivantes :

Missions	Actions à réaliser dans NotebookLM	Réflexion de la personne étudiante
1- Téléverser un recueil de documents ² dans le Notebook.	<ul style="list-style-type: none"> – Ajoutez dans la boîte de dépôt les documents/les liens URL communiqués dans l'encadré au bas de l'exercice – Renommez les documents que vous souhaitez avec des noms explicites afin de les reconnaître aisément. 	<i>Quels autres documents auriez-vous pu ajouter à ce recueil ?</i>
2- Résumer un document.	<ul style="list-style-type: none"> – Cochez un seul document (au choix) dans la section des sources (préférer les 3 premières url car elles correspondent aux pages principales) – Dans la section de conversation, demandez un résumé de la source sélectionnée. – Cliquez sur la référence apparaissant dans le premier paragraphe du contenu généré pour repérer le passage du texte auquel elle renvoie. <p>Attention : les conseils abordés précédemment sur la formulation des requêtes demeurent valables ici.</p>	<i>À la première lecture du résultat généré, qu'est-ce qui semble plus ou moins utile dans NotebookLM comparativement à d'autres agents conversationnels ?</i>
3- Comparer des idées.	<ul style="list-style-type: none"> – Cochez deux documents (au choix) dans la section des sources. 	<i>Comment cette comparaison</i>

² Dans cet exercice, vous êtes invités.es à déposer un recueil de sources préalablement sélectionnées. Selon le contexte et les situations d'apprentissage, vous pouvez composer votre propre recueil de ressources. Voici trois points de vigilance avec les outils d'IAg qui utilisent la technique des RAG :

- 1) Assurez-vous de la cohérence des documents sélectionnés dans le cas de documents imposants
- 2) Pour respecter les droits d'auteur, privilégiez toujours des textes en accès libre ou relevant du domaine public
- 3) Faites preuve de prudence avant de téléverser des documents sensibles.

	<p>– Dans la section de conversation, demandez de comparer les idées/les conclusions/les points d'accord/les points de divergence des deux documents.</p>	<p><i>éclaire-t-elle votre compréhension du sujet ?</i></p>
<p>4- Structurer sa révision d'examen.</p>	<p>– Cochez trois documents (au choix) dans la section des sources. – Dans la section des notes, cliquez sur Guide d'étude afin de générer un nouveau document.</p>	<p><i>En quoi ce nouveau document pourrait vous être utile dans votre cheminement académique ?</i></p>
<p>5- Élaborer un plan de présentation du contenu.</p>	<p>– Cochez l'ensemble des documents dans la section des sources. – Dans la section de conversation, demandez un plan de présentation du contenu des sources en 5 diapositives.</p>	<p><i>Quels choix ou ajustements auriez-vous apportés à ce plan pour mieux répondre à votre intention de communication ou à votre public cible ?</i></p>
<p>6- Bonus : créer une ressource complémentaire</p>	<p>Cochez un document (au choix) dans la section des sources. – Dans la section du studio, générez un podcast en français canadien d'environ 5 minutes. OU – Dans la section de conversation, générez une carte conceptuelle. OU - Dans ?, produire un exemplier des choses à faire et ne pas faire</p>	<p><i>Dans quel contexte ce type de ressource (balado ou carte conceptuelle) pourrait-il enrichir vos apprentissages ou ceux de vos pairs ?</i></p>

Partie 2 (5-10 minutes, restitution en plénière)

Consignes à l'enseignant :

Collecter les retours d'expériences ; notamment sonder sur les limites rencontrées.

Rappeler les points de vigilance avec les outils d'IAg qui utilisent la technique des RAG :

1. Respecter les droits d'auteur
2. Faire preuve de prudence avant de téléverser des documents sensibles.
3. S'assurer de la cohérence des documents sélectionnés dans le cas de documents imposants.

Ressources

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/prompt-design-strategies?hl=fr>

<https://ai.google.dev/gemini-api/docs/prompting-strategies?hl=fr>

<https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/overview>

<https://docs.google.com/spreadsheets/d/19jzLgRruG9kjUQNkTCg1ZjdD6l6weA6qRXG5zLIAhC8/edit?gid=869808629#gid=869808629>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/be-clear-and-direct>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/multishot-prompting>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/multishot-prompting#example-analyzing-customer-feedback>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/chain-of-thought>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/use-xml-tags>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/system-prompts>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/prefill-claude-s-response>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/chain-prompts>

<https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/long-context-tips>

<https://docs.claude.com/en/docs/test-and-evaluate/strengthen-guardrails/reduce-hallucinations>

<https://docs.claude.com/en/docs/test-and-evaluate/strengthen-guardrails/increase-consistency>

<https://docs.claude.com/en/docs/test-and-evaluate/strengthen-guardrails/mitigate-jailbreaks>

<https://ai.google.dev/gemini-api/docs/prompting-strategies?hl=en>

<https://cloud.google.com/discover/what-is-prompt-engineering?hl=fr>

<https://cloud.google.com/discover/what-is-prompt-engineering?hl=en>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/introduction-prompt-design?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/prompt-design-strategies?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/clear-instructions?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/few-shot-examples?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/system-instruction-introduction?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/system-instructions?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/assign-role?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/contextual-information?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/structure-prompts?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/explain-reasoning?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/break-down-prompts?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/adjust-parameter-values?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/prompt-iteration?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/prompt-gallery?hl=fr>

<https://cloud.google.com/vertex-ai/generative-ai/docs/text/text-prompts?hl=fr>

https://github.com/GoogleCloudPlatform/generative-ai/blob/main/gemini/prompts/intro_prompt_design.ipynb

<https://ai.google.dev/gemini-api/docs/thinking?hl=fr>

<https://cloud.google.com/gemini/docs/discover/write-prompts?hl=fr>

https://docs.mistral.ai/guides/prompting_capabilities/

<https://platform.openai.com/docs/guides/prompt-engineering>

<https://platform.openai.com/docs/guides/reasoning-best-practices>

https://cookbook.openai.com/examples/gpt4-1_prompting_guide

https://cookbook.openai.com/examples/gpt-5/gpt-5_prompting_guide

<https://cdn.openai.com/spec/model-spec-2024-05-08.html>

Exercice 4 : Développer son propre agent

Consignes à l'enseignant : indiquer si l'objectif de l'agent à créer est celui indiqué par défaut.

Un agent est un assistant virtuel spécialisé avec lequel vous ou un autre agent peut interagir soit via une IHM soit par programmation via une API.

L'objectif de cet exercice est de concevoir votre agent d'IA spécialisé dans l'assistance « d'aide à l'écriture de prompts ». Votre chargé de TD pourra éventuellement vous donner un autre objectif d'assistance ou bien vous laisser libre de choisir (en lien avec votre discipline ou vie en tant qu'étudiant ou à l'IUT ou encore pour un domaine particulier supposé non connu par un des agents comme recevoir des consignes pour se déplacer dans les bâtiments et salles du site Joffre de l'IUT).

Mistral Le Chat offre la possibilité de concevoir et partager ses propres agents. La documentation présente l'IHM pour créer vos agents.

<https://help.mistral.ai/en/articles/347482-what-s-an-agent-and-how-do-i-create-one>

Consignes pour les étudiants : utiliser [Le Chat](#) (la création d'un compte est probablement nécessaire). La page d'accueil dédiée aux [agents](#) vous permet d'utiliser un agent existant ainsi que d'en créer un nouveau from scratch ou à partir d'un existant. Quatre agents sont déjà disponibles par défaut : Data Analyst, Personal Tutor, Global Summarizer et Writing Assistant.

Partie 1 (20-30 minutes, travail en individuel ou en binôme)

- A) Jeter un oeil à la [documentation de création d'agents](#) avec Le Chat. Puis [consulter la configuration des agents disponibles](#) pour comprendre comment ils sont construits. Observer les contenus et formats des 1) instructions, 2) des garde-fous, 3) du ton, 4) des connaissances (e.g. accès au web, bibliothèque de documents).
- B) Sur le papier, réfléchir au prompt (composants et stratégies à écrire) pour créer votre propre assistant.
- C) Créer votre assistant Le Chat avec ce prompt et tester sa robustesse, sa précision... testez cet assistant en demandant à votre agent préféré de vous aider sur la base de requêtes qui s'inspirent de l'exercice 2.

Partie 2 (20-30 minutes, travail en individuel ou en binôme)

- D) Sur le papier d'abord, concevoir un prompt pour demander à un agent de produire un prompt visant la création d'un assistant (« d'aide à l'écriture de prompts » si vous êtes resté sur ce sujet).
- E) Quand vous estimez ce prompt stabilisé, requêtez votre agent pour qu'il vous génère ce prompt de création. Adaptez au besoin.
- F) Créer votre assistant à l'aide de ce prompt généré. Testez votre agent sur la base de requêtes qui s'inspirent de l'exercice 2.

Partie 3 (10 minutes, restitution en plénière)