

Systemes génératifs et résolution de problèmes mathématiques scolaires : une chronique

Le problèmes des chameaux et des dromadaires

Alain Favre & Luc-Olivier Pochon

Introduction

Les outils informatiques sont de plus en plus utilisés en mathématiques. Parmi les plus classiques et sophistiqués se trouvent les systèmes d'aide à la démonstration. De façon paradoxale mais compréhensible, les problèmes mathématiques scolaires ne sont pas embarqués dans ce train de résolutions automatisées. En effet, le support de l'activité mathématique au niveau élémentaire utilise le langage courant qui est peu formalisé. L'arrivée de systèmes génératifs modifie la situation. Si d'aucuns voient dans les problèmes mathématiques scolaires (nous utiliserons l'abréviation anglaise MWP pour Math Word Problem) un bon domaine pour tester ces systèmes, ceux-ci peuvent en retour donner des indications sur les difficultés liées à des formulations des énoncés des problèmes.

Dans ce cadre, le but de notre travail est d'examiner et d'analyser les réponses données à des MWP par des systèmes génératifs généraux ou plus spécialisés à la manière dont on examine les procédures et les solutions données par des élèves ou des étudiants. Par ailleurs, si les grands principes de fonctionnement de ces systèmes nous sont connus, plusieurs processus restent à éclaircir. Par conséquent, un autre objectif est d'étayer quelques hypothèses à propos du fonctionnement de ses systèmes qui mettent en oeuvre diverses stratégies centrées sur (ou autour) des architectures de réseaux neuronaux.

Ceci est une chronique ; nous respecterons le genre en commençant par le commencement. Depuis plusieurs années nous nous occupons de la mise à disposition de ressources informatiques dans l'enseignement des mathématiques élémentaires. Ce furent les projets Prof'Expert puis de l'Ermitage pour lesquelles des générateurs de problèmes ont été créés de même que des analyseurs de réponses. La banque de problèmes¹ élaborée pour le Rallye mathématique transalpin (RMT) nous a amenés à réfléchir à des questions de classification, de recherche voire de génération de problèmes. Pour cela nous avons mis en oeuvre des analyseurs (tokeniseur-parser-lemmatisation ...) et avons longuement testé les modèles LSA-LDA (Favre, Pochon, 2021). Nous nous sommes intéressés aux « taggers » et à l'apprentissage de données comprenant des termes mathématiques. Soudain, le sol s'est dérobé sous nos pieds. L'arrivée à grand fracas des systèmes génératifs, la profusion des articles, des avancées techniques et des discours médiatiques parfois confus nous a surpris. Après quelques tentatives d'adapter ces outils à nos différentes entreprises, nous nous sommes perdus dans la foudritude des approches. Pour appréhender cette nouvelle réalité, il fallait faire le tri. De fil en aiguille nous avons choisi de nous concentrer sur la résolution de quelques problèmes (MWP) ayant chacun une particularité pédagogique. Le premier de la liste dont la structure repose sur la résolution d'un système linéaire à deux inconnues² fait l'objet de ce document.

MWP

Dans l'enseignement des sciences, un problème de mots est un exercice mathématique (comme dans un manuel, une feuille de travail ou un examen) dans lequel des informations générales importantes

1 <http://arnt.projet-ermitage.org/>

2 Dit en termes savants, d'autres stratégies peuvent être mises en oeuvre par des écoliers n'ayant pas d'algèbre parmi les outils à leur disposition.

sur le problème sont présentées dans un langage ordinaire plutôt qu'en notations mathématiques. Comme la plupart des problèmes de mots présentent un petit récit, ils sont parfois appelés problèmes d'histoire. Le récit peut être plus ou moins long. Par ailleurs, la quantité de langage technique utilisé peut varier. En général on distingue le récit proprement dit (les données) et la consigne qui propose une ou plusieurs questions dont on demande la réponse.

Une façon d'entrer dans l'analyse d'un problème est de considérer les trois phases intervenant dans la procédure de résolution : la mathématisation (ou algébrisation) qui extrait des données les quantités et les relations en jeu, des séquences d'actions déductives (ou algorithme ou micro-tâches), et finalement la formulation de la solution. Il peut y avoir des va-et-vient entre les séquences. Le processus de déduction peut amener à revoir les données et les consignes. La formulation de la solution peut être en partie imbriquée dans la phase précédente. L'expression de la solution peut se réduire à une seule formule ou un nombre ou alors une explication voire une démonstration.

Les travaux en intelligence naturelle

Une branche de la recherche en didactique des mathématiques s'est attachée à analyser la résolution de problèmes mathématiques au niveau de la scolarité obligatoire essentiellement (élèves de 6 à 15 ans). Cette analyse est principalement focalisée sur l'analyse d'erreur. Les résultats de ces travaux mènent généralement à l'adaptation de l'enseignement aux représentations et capacités des enfants concernés. Nous renonçons à donner des références bibliographiques à ce sujet en nous contentant de citer le psychologue Jean Piaget parmi les précurseurs qui se sont penchés sur l'évolution des capacités logico-mathématiques des enfants³. Parmi les gens de terrain, Stella Baruck dont l'ouvrage paru en 1985, réédité en 2017, « L'âge du capitaine⁴, de l'erreur en mathématiques » résume une partie de ces travaux (Baruck, 1985). Mais d'autres recherches constellent les différentes revues consacrées à l'enseignement et à la didactique des mathématiques. Le RMT, parmi d'autres organisations de ce type, publie également des analyses de problèmes dans la Gazette de Transalpie⁵.

Réciproquement, les erreurs de enfants permettent aussi de déceler des formulations fautives ou inadaptées, ou de façon plus constructive d'identifier des aspects structurels qui permettent de classer les sources d'incompréhension ou d'erreurs. On trouve des indications à ce propos principalement dans d'anciens travaux, forcément puisque les MWP classiques ont disparus des programmes (un travail typique est celui de Josette Adda, 1975)⁶. Les travaux plus récents concernent plus spécifiquement le langage mathématique (Hache, 2013 ; Daniel Lacombe, 2011).

L'énoncé et sa solution

Le problème qui va servir à cette analyse est donnée dans l'encadré 1. Cette version est extraite du 5e rallye⁷ (1997). Le même problème a été reposé en 2016 lors du 24e rallye⁸ avec des quantités différentes : 23 bosses et 68 pattes, la consigne est alors plus précise : **Expliquez comment vous avez trouvé votre réponse.**

3 Ses travaux avaient également une visée épistémologique plus générale.

4 Référence à une histoire où après avoir donné des informations chiffrées sur un bateau, l'on demande l'âge du capitaine.

5 <https://armtint.eu/fr/la-gazette-de-transalpie-2/>

6 Jean Cardinet et Jean-Blaise Grize ont organisé dans les années 1980 un séminaire tenu à l'IRD sur le sujet dont il n'existe que des notes manuscrites. Par ailleurs, l'oeuvre de Grize est constellée d'exemple de difficultés logiques liées au langage.

7 http://www.projet-ermitage.org/ARMT/navi_fic2.php?code=sd10-fr&flag=1&langue=fr&enonce=05rmti_fr-9&w=0

8 http://www.projet-ermitage.org/ARMT/navi_fic2.php?code=op69-fr&flag=1&langue=fr&enonce=24rmti_fr-8&w=0

Chameaux et dromadaires

Cléopâtre a dessiné des chameaux et des dromadaires, cela fait 19 bosses et 52 pattes. Elle sait que les chameaux ont deux bosses et les dromadaires n'en ont qu'une.

Puis elle a encore dessiné un homme sur le dos de chaque chameau.

Combien a-t-elle dessiné d'hommes en tout ?

Expliquez votre réponse.

Encadré 1 : Le problème des chameaux et des dromadaires

Il est utile de préciser que dans la base de problème, l'énoncé complet figure sous forme d'image. Le texte disponible est donné sous une forme courte : « *Trouver, parmi des chameaux et des dromadaires, le nombre de chameaux connaissant le nombre de pattes, 52, et le nombre de bosses, 19* ». Le texte de l'énoncé complet se trouve toutefois dans d'autres documents (actes de colloques).

Solution IN (intelligence naturelle)

1. Mathématisation

x: nombre de chameaux

y: nombre de dromadaires

Relation liée au nombre de pattes : $4 \times (x + y) = 52$ $x + y = \frac{52}{4} = 13$

Relation liée au nombre de bosses : $2 \times x + y = 19$

2. Résolution

A résoudre le système de deux équations à deux inconnues :

$$\begin{cases} x + y = 13 \\ 2 \times x + y = 19 \end{cases}$$

... qui mène à la solution $x = 6$, $y = 7$.

3. Solution du problème (réponse à la question de la consigne)

Cléopâtre a dessiné 6 hommes.

Complément

Dans ce cas, l'explication est implicitement contenue dans la méthode de résolution. Mais comme on peut le vérifier dans les analyses a priori proposées dans la banque de problème, diverses autres stratégies peuvent être envisagées. En effet, certaines catégories d'élèves n'ont pas l'outil algébrique à leur portée. Par exemple : Procéder par essais-erreurs à partir de 9 chameaux (environ la moitié des animaux), ou encore après avoir déduit le nombre d'animaux à partir du nombre de pattes envisager le cas hypothétiques de dromadaires à deux bosses également.

A noter encore qu'il s'agit finalement de prendre de la distance par rapport au pur traitement arithmético-algébrique : la consigne concerne l'action de Cléopâtre et non la solution d'un système d'équations.

Résolution par les systèmes génératifs

Dans nos essais, effectués sauf mention du contraire en août et septembre 2024, nous allons distinguer les systèmes généralistes et les systèmes « laboratoires ».

Les systèmes généralistes sont ceux qui présentent une interface web grand public. Nous nous contenterons de les utiliser en formulant des requêtes (prompts) sans exploiter les éventuelles extensions que permettent des abonnements.

Nous désignons par systèmes « laboratoires » ceux qui permettent facilement d'ajuster quelques paramètres (notamment température et modèle) ou ceux que l'on peut diriger à partir d'une interface de programmation, notamment un « notebook » ou un environnement de développement Python.

Cas des systèmes généralistes

Dans cette partie, le problème sera posé à des systèmes devenus standard, où les possibilités de paramétrage sont très réduites ou nulles dans les versions « grand public ». Les modèles de langage sont souvent propres au système. L'architecture, hormis les phases techniques habituelles : tokenisation, embedding, ..., n'est pas apparente. A priori, nous ne savons pas si des modules spécialisés (agents de raisonnement par exemple) sont mis en œuvre.

ChatGpt⁹

La solution proposée est reproduite dans les encadrés¹⁰ 2a et 2b. Rien à redire, l'enjeu est *compris*. La difficulté à revenir à Cléopâtre (la tendance lorsque le problème est proposé à des enfants est de donner la réponse sans évoquer les hommes dessinés par Cléopâtre) est *surmontée*. Au vu de la réponse, l'ajout d'une partie déclarative est superflue (« zero-shot prompting » suffit). La question qui reste ouverte est de savoir quelles sont les sources utilisées. Certains de nos essais antérieurs ont-ils été exploités pour améliorer le résultat ?

On note encore que des codes de mise en forme subsiste (supprimée dans notre report). Par exemple : $\begin{cases} 2c + d = 19 \\ c + d = 13 \end{cases}$

Pour résoudre ce problème, il faut d'abord déterminer le nombre de chameaux et de dromadaires que Cléopâtre a dessinés. Nous savons que :

- Un chameau a 2 bosses.
- Un dromadaire a 1 bosse.
- Le nombre total de bosses est de 19.
- Le nombre total de pattes est de 52.

Désignons par c le nombre de chameaux et par d le nombre de dromadaires.

Équations à résoudre :

1. **Pour les bosses :** $2c + d = 19$
2. **Pour les pattes :** Chaque animal (chameau ou dromadaire) a 4 pattes, donc :
 $4c + 4d = 52$

Encadré 2a : Début de la réponse de ChatGPT (août 2024)

⁹ <https://chatgpt.com/auth/login>

¹⁰ Les encadrés reprennent textuellement les réponses générées par les systèmes. Toutefois, les coquilles sont signalées par un italique.

Simplification de l'équation des pattes :

Divisons l'équation par 4 : $c+d=13$

Systeme d'équations :

Nous avons maintenant un système de deux équations :

$$2c+d=19 ; c+d=13$$

Soustraction des deux équations :

Soustrayons la deuxième équation de la première pour éliminer d :

$$(2c+d)-(c+d)=19-13$$

Cela donne : $c=6$

Calcul de d :

Substituons $c=6$ dans l'équation $c+d=13$:

$$6+d=13$$

$$d=7$$

Vérification :

- Nombre de bosses : $2c+d=2(6)+7=12+7=19$ (OK)
- Nombre de pattes : $4c+4d=4(6)+4(7)=24+28=52$ (OK)

Conclusion :

Il y a donc 6 chameaux et 7 dromadaires. Puisque Cléopâtre a dessiné un homme sur le dos de chaque chameau, elle a dessiné **6 hommes en tout**.

Encadré 2b : La solution de chatGPT (août 2024)

Dans une tentative précédente (avril 2024) avec une version antérieure du système, un énoncé simplifié était proposé : « *Trouver, parmi des chameaux et des dromadaires, le nombre de chameaux connaissant le nombre de pattes, 52, et le nombre de bosses, 19* ».

Dans ce cas, les équations sont posées et résolues correctement, mais à partir de données erronées correspondant à ce premier commentaire :

Soit x le nombre de chameaux et y le nombre de dromadaires. On sait que chaque chameau a 4 pattes et 1 bosse, et chaque dromadaire a 2 pattes et 1 bosse.

Le système à résoudre est alors $x+y = 19$; $4x + 2y = 52$ (x étant le nombre de chameaux et y celui des dromadaires). La résolution s'effectue par la méthode de substitution $y = 19 - x$. Le système poursuit :

En remplaçant y dans l'équation 1, on obtient:

$$4x + 2(19 - x) = 52 ; 4x + 38 - 2x = 52 ; 2x + 38 = 52 ; 2x = 14 ; x = 7$$

Maintenant que nous avons trouvé le nombre de chameaux ($x = 7$), nous pouvons trouver le

nombre de dromadaires en utilisant l'équation 2:

$$y = 19 - 7y = 12$$

Donc, il y a 7 chameaux et 12 dromadaires.

En comparant les deux essais, outre les questions d'anatomie des animaux, on note les diverses façons de noter les « inconnues » : x, y dans la première version, c et d dans la version plus récente.

Gemini¹¹

La première solution proposée est reproduite dans l'encadré 3. Dans ce cas, le nombre de pattes n'est pas pris en compte. Le système se *concentre* sur les bosses.

Absolument ! Analysons ce problème étape par étape pour trouver le nombre d'hommes dessinés par Cléopâtre.

Comptons les animaux à bosses :

- **Nombre total de bosses :** 19
- **Chameaux (2 bosses chacun) :** Si chaque chameau a 2 bosses, alors il y a $19 \text{ bosses} \div 2 \text{ bosses/chameau} = 9,5$ chameaux. Mais on ne peut pas avoir une moitié de chameau, donc il y a **9 chameaux** au maximum.
- **Dromadaires (1 bosse chacun) :** Les 19 bosses initiales moins les 18 bosses des 9 chameaux nous donne 1 bosse restante. Donc, il y a **1 dromadaire**.

Récapitulons les animaux :

- **Chameaux :** 9
- **Dromadaires :** 1

Passons aux hommes :

- **Règle :** Un homme sur le dos de chaque chameau.
- **Calcul :** Comme il y a 9 chameaux, et qu'il y a un homme par chameau, Cléopâtre a dessiné **9 hommes**.

Réponse finale :

Cléopâtre a dessiné **9 hommes** en tout.

Pourquoi on ne compte pas les dromadaires pour les hommes ?

Dans l'énoncé, il est spécifié qu'elle a dessiné un homme sur le dos de **chaque chameau**. Comme le dromadaire n'est pas un chameau, on ne compte pas d'homme sur son dos.

En résumé, en se concentrant sur l'information clé que Cléopâtre a dessiné un homme par chameau, nous avons pu déterminer le nombre total d'hommes qu'elle a représentés dans son dessin.

Encadré 3 : Première solution proposée par Gemini (août 2024)

On retient pour une analyse ultérieure que la stratégie de considérer 9 comme première approximation (environ la moitié de 19) est une piste suggérée dans l'analyse de la tâche proposée sur la fiche correspondante du RMT. L'obstacle de revenir à l'action de Cléopâtre est *surmonté*.

On notera aussi le luxe de précision : « Comme le dromadaire n'est pas un chameau, on ne compte pas d'homme sur son dos. »

¹¹ <https://gemini.google.com/app>

Une deuxième tentative a été effectuée. La tâche est précédée d'une partie déclarative : « Tu es un étudiant en mathématique et un problème est à résoudre ». L'encadré 4 donne la réponse à cette déclaration.

La réponse (encadré 5) est provoquée par l'énoncé du problème précédé de « aide moi à poser les équations ». Comme dans le cas précédent, seul le nombre de bosses est pris en compte. Cependant une relation est fournie et la recherche de la solution par essais et erreurs (suggérée dans la banque de problèmes du RMT) est adoptée. Pourquoi le système *estime-t-il* que la solution est celle qu'il propose, mystère. L'affirmation induit évidemment que la donnée du nombre de pattes est superflue.

On notera la remarque concernant la parité.

Préciser le nombre de bosses que possèdent respectivement les dromadaires et les chameaux n'est pas utile. Dans un autre essai, sans cette information, Gemini précise (après son Absolutement:) que l'étape 1 est de « comprendre les animaux » :

- **Chameaux:** Ils ont généralement deux bosses.
- **Dromadaires:** Ils n'ont qu'une seule bosse.

En cours de route, Gemini pose la question : « Pourquoi on ne peut pas calculer le nombre exact de chameaux et de dromadaires ? »

Question à laquelle il apporte la réponse : « Le problème ne nous donne pas assez d'informations pour faire cette distinction. On ne sait pas combien de bosses appartiennent aux chameaux et combien aux dromadaires. »

Absolument ! En tant que modèle de langage, je peux t'aider à résoudre ton problème de mathématiques.

Pour que je puisse t'assister au mieux, pourrais-tu me fournir les informations suivantes :

- **L'énoncé exact du problème:** S'il y a des figures ou des schémas, essaie de les décrire le plus précisément possible.
- **Les notions mathématiques en jeu:** S'agit-il d'algèbre, de géométrie, de calcul, de statistiques, ou d'un autre domaine ?
- **Ce que tu as déjà essayé:** Quelles sont les étapes que tu as déjà franchies ? Quelles difficultés rencontres-tu ?

Voici quelques pistes pour t'aider à formuler ta question :

- "Je dois résoudre une équation du second degré, mais je bloque à la factorisation."
- "J'ai un problème de géométrie avec un triangle rectangle, et je ne sais pas comment calculer l'hypoténuse."
- "Je dois trouver la moyenne et l'écart-type d'une série de données, mais je ne suis pas sûr de la formule à utiliser."

N'hésite pas à me poser toutes les questions que tu veux ! Je peux t'expliquer les concepts, te donner des exemples, ou même te proposer une solution détaillée.

Ensemble, nous allons trouver la réponse à ton problème !

Alors, prêt(e) à commencer ?

Encadré 4 : Recommandations de Gemini (août 2024)

Absolument ! Ce problème est une énigme amusante qui combine logique et mathématiques.

Analysons d'abord l'information concernant les chameaux et les dromadaires :

- **Les bosses:** Chaque chameau a 2 bosses, chaque dromadaire 1 bosse. Au total, il y a 19 bosses.
- **Les pattes:** Chaque animal a 4 pattes.
- **L'équation des bosses:** Si on note "C" le nombre de chameaux et "D" le nombre de dromadaires, on peut écrire : $2C + D = 19$

Résolvons cette équation:

Il n'y a pas de solution évidente à première vue. Cependant, nous pouvons remarquer que le nombre de bosses (19) est impair. Or, si on multiplie un nombre quelconque par 2 (nombre de bosses d'un chameau), on obtient toujours un nombre pair.

Cela signifie qu'il **doit y avoir au moins un dromadaire** pour que la somme des bosses soit impaire.

Essayons différentes combinaisons:

- Si on avait 9 chameaux, cela ferait 18 bosses, il faudrait donc 1 dromadaire pour arriver à 19.
- Si on avait 8 chameaux, cela ferait 16 bosses, il faudrait donc 3 dromadaires pour arriver à 19.
- Et ainsi de suite...

En testant différentes combinaisons, on trouve que la seule solution possible est :

- 7 chameaux ($7 * 2 = 14$ bosses)
- 5 dromadaires ($5 * 1 = 5$ bosses)

Revenons à la question initiale :

- Cléopâtre a dessiné un homme sur le dos de **chaque chameau**.
- Elle a dessiné 7 chameaux.
- **Donc, elle a dessiné 7 hommes en tout.**

En conclusion:

Grâce à un raisonnement logique et en résolvant une petite équation, nous avons pu déterminer que Cléopâtre a dessiné 7 hommes.

Le nombre de pattes des animaux et l'homme sur le dos des chameaux n'étaient pas des informations nécessaires pour résoudre cette énigme particulière.

N'hésite pas si tu as d'autres questions ou si tu souhaites résoudre un autre problème !

Encadré 5 : Deuxième solution proposée par Gemini (août 2024)

Mistral¹²

L'encadré 6 donne la réponse de Mistral. Elle est comparable à celle de ChatGPT. Une variante numérique du problème (160 pattes 70 bosses) est résolu de la même manière (résultat 30 chameaux et 10 dromadaires).

12 <https://auth.mistral.ai/>

Pour résoudre ce problème, nous devons déterminer le nombre de chameaux et de dromadaires que Cléopâtre a dessinés. Nous savons que les chameaux ont deux bosses et les dromadaires en ont une. Nous avons également des informations sur le nombre total de bosses et de pattes.

1. Définir les variables :

- Soit c le nombre de chameaux.
- Soit d le nombre de dromadaires.

2. Établir les équations :

- Chaque chameau a 2 bosses et chaque dromadaire a 1 bosse. Donc, l'équation pour les bosses est : $2c+d=19$
- Chaque chameau a 4 pattes et chaque dromadaire a 4 pattes. Donc, l'équation pour les pattes est : $4c+4d=52$

3. Simplifier l'équation des pattes :

- Divisons l'équation des pattes par 4 : $c+d=13$

4. Résoudre le système d'équations :

- Nous avons maintenant deux équations : $2c+d=19$
- Soustrayons la deuxième équation de la première : $(2c+d)-(c+d)=19-13$
 $2c+d-c-d=6$ $c=6$

5. Trouver le nombre de dromadaires :

- Substituons $c=6$ dans l'équation $c+d=13$: $6+d=13$ $d=7$

6. Conclusion :

- Cléopâtre a dessiné 6 chameaux et 7 dromadaires.
- Puisqu'elle a dessiné un homme sur le dos de chaque chameau, le nombre d'hommes dessinés est égal au nombre de chameaux.

Donc, Cléopâtre a dessiné 6 hommes en tout.

Encadré 6 : Solution proposée par Mistral (août 2024)

YIAHO¹³

L'encadré 7 donne la déclaration destinée au système.

Dans la réponse (encadré 8) les relations utilisées sont labellisées.

YIAHO IQ 200 : Bonjour, je suis l'IA de Yiaho, comment puis-je vous aider ?

Guest: Voilà un problème mathématique, peux-tu m'aider à le résoudre?

YIAHO IQ 200 : Bien sûr, je suis là pour ça ! Quel est le problème mathématique que vous aimeriez résoudre ?

Encadré 7 : Introduction avec YIAHO (août 2024)

¹³ <https://www.yiaho.com/>

YIAHO IQ 200 : Pour résoudre ce problème, nous allons utiliser un système d'équations.

1. Définissons nos variables :

- Soit (c) le nombre de chameaux.
- Soit (d) le nombre de dromadaires.

2. Établissons les équations à partir des informations données :

- Chaque chameau a 2 bosses, et chaque dromadaire a 1 bosse. Le total des bosses est donné comme 19 : [$2c + d = 19$] (1)]
- Chaque chameau a 4 pattes, et chaque dromadaire a 4 pattes. Le total des pattes est donné comme 52 : [$4c + 4d = 52$] (2)]
- On peut simplifier l'équation (2) en divisant par 4 : [$c + d = 13$] (3)]

3. Résolvons le système d'équations :

- À partir de l'équation (3), nous pouvons exprimer (d) en fonction de (c) : [$d = 13 - c$] (4)]
- Remplaçons (d) dans l'équation (1) : [$2c + (13 - c) = 19$] [$2c + 13 - c = 19$] [$c + 13 = 19$] [$c = 19 - 13 = 6$]

4. Trouvons maintenant (d) :

- Remplaçons (c) dans l'équation (4) : [$d = 13 - 6 = 7$]

5. Conclusion :

- Cléopâtre a dessiné 6 chameaux et 7 dromadaires.
- Puisqu'elle a dessiné un homme sur le dos de chaque chameau, le nombre total d'hommes dessinés est de 6.

Ainsi, Cléopâtre a dessiné **6 hommes en tout**.

Encadré 8 : Solution proposée par YIAHO (août 2024)

Solution par des systèmes « laboratoires »

Le modèle NuminaMath-7B-TIR¹⁴

Ce modèle a été utilisé par les gagnants de la compétition « AIMO¹⁵ 2022-23 ». Le résultat donné dans l'encadré 9 est obtenu avec la température à 0.1 (c'est-à-dire le système a peu de choix).

Tout d'abord remarquons que le module Python Sympy est explicitement utilisé pour la résolution des équations. Ce qui ne semble pas être le cas dans des cas examinés précédemment ? A noter que cette formulation de la solution sous la forme d'un programme Python se trouve fréquemment. L'usage de sympy simplifie cette séquence¹⁶.

La réponse est juste, compte tenu du fait que les chameaux ont été gratifiés de 2 pattes (une source

14 <https://huggingface.co/AI-MO/NuminaMath-7B-TIR>

15 Artificial Intelligence Mathematical Olympiad. L'url proposée (<https://huggingface.co/spaces/AI-MO/math-olympiad-solver>) semble inactive. Pour l'utilisation du modèle on peut se référer à ce nouvel espace de huggingface: <https://huggingface.co/spaces/Pra-tham/deepseekmath>

16 Voir à ce propos Joy He-Yueya, Gabriel Poesia, Rose E. Wang, Noah D. Goodman (2023) ou l'autre Poesia pour une mise en perspectives de deux manières d'intégrer des séquences Python.

de confusion possible sera évoquée ultérieurement).

On notera encore que la réponse est en anglais, parsemée des mots français provenant de l'énoncé. Il est probable que plusieurs systèmes commencent à traduire les requêtes en anglais (nouvelle pseudo-énigme).

La relance avec la température portée à 0.9 remet quatre pattes aux chameaux et conduit à une solution exacte.

To solve this problem, we need to set up a system of equations based on the given information:

1. Let x be the number of chameaux.
2. Let y be the number of dromadaires.

From the problem, we have the following information:

- Each chameau has 2 bosses.
- Each dromadaire has 1 *boss*.
- The total number of bosses is 19.
- Each chameau has 2 pattes.
- Each dromadaire has 4 pattes.
- The total number of pattes is 52.

We can set up the following system of equations:

1. $2x+y=19$ (total number of bosses)
2. $2x+4y=52$ (total number of pattes)

We need to solve this system of equations to find the values of x and y . Then, we can determine the number of hommes (men) drawn on the backs of chameaux.

Let's solve this step-by-step using Python.

```
from sympy import symbols, Eq, solve

# Define the variables
x, y = symbols('x y')

# Define the equations
eq1 = Eq(2*x + y, 19)
eq2 = Eq(2*x + 4*y, 52)

# Solve the system of equations
solution = solve((eq1, eq2), (x, y))
print(solution)

{x: 4, y: 11}
```

The solution to the system of equations is: $x=4$ $y=11$

This means there are 4 chameaux and 11 dromadaires.

Now, since each chameau has an accompanying man on its back, the number of men drawn is equal to the number of chameaux.

Thus, Cléopâtre has drawn: 4 hommes in total.

Encadré 9 : Solution basée sur le modèle NuminaMath-7B-TIR et sympy

Une nouvelle tentative (température 0.1) se révèle plus catastrophique. Le nombre d'hommes est

introduit (« h: the number of hommes drawn on the backs of chameaux »). Le nombre de pattes du dromadaire est à nouveau 2. Les équations sont alors (avec c et d resp. nombre de dromadaires et de chameaux) :

1. $2c+d=19$ (total number of bosses)
2. $4c+2d+h=52$ (total number of pattes, where $h=c$)

Ainsi le nombre de jambes est introduit dans la situation, ce que l'unique mot anglais 'leg' pour patte et jambe a peut-être favorisé (et est peut-être à l'origine de l'erreur du premier essai).

Continuons à analyser cette réponse :

« Substituting $h=c$ into the second equation, now we have a system of linear equations: »

1. $2c+d=19$
2. $5c+2d=52$

L'usage de sympy conduit au résultat $c=14$ $d=-9$ et au commentaire :

« Since the number of dromadaires (d) cannot be negative, it seems there might be an inconsistency in our initial setup or interpretation. Let's re-evaluate the problem. »

La re-évaluation se base sur les mêmes équations mais fait appel à une résolution par substitution ($d = 19-2c$) : « We'll solve the equations *manually* to ensure accuracy and then verify using Python. » Elle mène au même résultat et au même constat. Vient alors une re-évaluation du problème.

Mêmes données, les mêmes solutions sont trouvées (après un petit accro avec sympy : *oubli* de `from sympy import symbols, Eq, solve`) ce qui conduit à un commentaire :

« We'll use the `solve` function from SymPy correctly and ensure we handle the equations properly ». Et à un nouveau passage par sympy et une nouvelle re-évaluation

« Given the negative value for d, let's recheck the problem constraints and assumptions ». Il s'agit alors d'une Detailed Re-évaluation. Le calcul est à nouveau exécuté « manuellement ».

La re-évaluation mène à une séquence « Corrected Setup » (mais rien n'est changé) : « Let's solve this system again using Python, ensuring we handle non-negative values ». Et abandon.

L'intérêt de cette séquence est de montrer l'évaluation du résultat et les possibles reprises, même si dans ce cas le même processus est répété plusieurs fois.

Perplexity

Le système Perplexity est proposé en mode généraliste¹⁷ mais aussi en mode « laboratoire¹⁸ ». En mode généraliste, il donne une solution dans une forme comparable à ChatGPT (en juin 2024 déjà). Nous commenterons les solutions obtenues avec le deuxième mode avec divers modèles et différentes variantes de l'énoncé.

Solution 1 : L'encadré 10 montre que les équations sont correctement posées. Toutefois la solution proposée correspond à une situation avec 19 bosses et 48 pattes.

Une nouvelle tentative avec l'énoncé simplifié : « *Trouver, parmi des chameaux et des dromadaires, le nombre de chameaux connaissant le nombre de pattes, 52, et le nombre de bosses, 19* » conduit à une réponse correcte (encadré 11). On notera que la manipulation algébrique a été remplacée par une « manipulation » sur les animaux. C'est d'ailleurs une variante proposée dans l'analyse de la tâche de la fiche correspondante de la banque de problèmes du RMT.

¹⁷ <https://www.perplexity.ai/>

¹⁸ <https://labs.perplexity.ai/>

L'argument concernant les bosses est repris textuellement de la fiche de la banque de problèmes du RMT¹⁹ (voir plus loin ce qui conforte cette hypothèse)

Cléopâtre a dessiné 19 bosses et 52 pattes. Les chameaux ont deux bosses et les dromadaires n'ont qu'une. Pour trouver le nombre de chameaux et de dromadaires, nous pouvons utiliser les équations suivantes :

- Nombre de bosses = Nombre de chameaux * 2 + Nombre de dromadaires * 1
- Nombre de pattes = Nombre de chameaux * 4 + Nombre de dromadaires * 4

En remplaçant les valeurs données, nous obtenons :

- $19 = \text{Nombre de chameaux} * 2 + \text{Nombre de dromadaires} * 1$
- $52 = \text{Nombre de chameaux} * 4 + \text{Nombre de dromadaires} * 4$

En résolvant ce système d'équations, nous trouvons :

- Nombre de chameaux = 7
- Nombre de dromadaires = 5

Maintenant, Cléopâtre a dessiné un homme sur le dos de chaque chameau. Donc, elle a dessiné 7 hommes en tout.

Encadré 10 : Solution proposée par Perplexity/lab avec le modèle llama-3-sonar-32k-online, données originales (24 juin 2024)

Pour résoudre ce problème, nous devons utiliser les informations suivantes :

- Chaque chameau a deux bosses et quatre pattes.
- Chaque dromadaire a une bosse et quatre pattes.

Nous savons que le nombre total de pattes est de 52, ce qui signifie que le nombre total d'animaux est de $52 \div 4 = 13$.

Nous savons également que le nombre total de bosses est de 19. Si tous les animaux étaient des dromadaires, il y aurait 13 bosses. Pour atteindre 19 bosses, nous devons remplacer 6 dromadaires par 6 chameaux, car chaque chameau a deux bosses supplémentaires.

Donc, il y a 6 chameaux et 7 dromadaires.

Réponse : 6 chameaux.

Encadré 11 : Solution proposée par Perplexity/lab avec le modèle llama-3-sonar-32k-online, données simplifiées (24 juin 2024)

LM-Studio avec le modèle deepseek-ai-deepseek-math-7b-rl-GGUF²⁰

L'environnement LM-Studio²¹ met à disposition quelques modèles de langage et permet de les mettre en oeuvre. L'encadré 12 montre la requête effectuée sur le modèle deepseek-ai-deepseek-math-7b-rl-GGUF (la version la moins gourmande en mémoire). L'encadré 13 montre la réponse du système en format LaTeX (indice qui est retenir) et l'encadré 14 la réponse que nous avons mise en forme.

19 http://www.projet-ermitage.org/ARMT/navi_fic2.php?code=sd10-fr&flag=1&langue=it&w=0

20 deepseek-math-7b-rl.Q4_K_M.gguf

21 <https://lmstudio.ai/>

```

from openai import OpenAI

# Point to the local server
client = OpenAI(base_url="http://localhost:1234/v1",
api_key="lm-studio")

completion = client.chat.completions.create(
    model="tastypear/deepseek-ai-deepseek-math-7b-rl-GGUF",
    messages=[
        {"role": "system", "content": "Résoudre et expliquer."},
        {"role": "user", "content": "Cléopâtre a dessiné des
chameaux et des dromadaires, cela fait 19 bosses et 52 pattes.
Elle sait que les chameaux ont deux bosses et les dromadaires
n'en ont qu'une.Puis elle a encore dessiné un homme sur le dos
de chaque chameau. Combien a-t-elle dessiné d'hommes en
tout ? ."}
    ],
    temperature=0.7,
)

print(completion.choices[0].message)

```

Encadré 12 : Un module de Python avec une requête sur le modèle deepseek-ai-deepseek-math-7b-rl-GGUF via le serveur LM-Studio

```

= RESTART: /Users/irpochon/Documents/Congres-projets/Python/LM-
Studio/essai-chameaux.py
ChatCompletionMessage(content="On note x le nombre de chameaux et y le
nombre de dromadaires. On a donc les équations suivantes :\n2x + y =
19 \\quad (E1) car chaque animal à deux bosses, sauf un seul qui n'en
ont qu'une ; 4x+2y=52\\quad(E2)car chacun des quatre pattes du chameau
ou dromadaire compte pour une patte.\nOn résoud le système (E1, E2),
on multiplie E1 par 2 : 4x+2y=38\\quad(2E1) et soustrait cette
équation à l'(E2) ce qui donne:0 + (52-38)=14 \\Longrightarrow 2y =
14. Donc y=\\frac{14}{2}=7 dromadaires.\nOn remplace dans (E1), on a
2x+7=19\\Longrightarrow 2x=(19-7) = 12, donne alors : x =\\frac {12}
{2} =6. Elle a 6 chameaux et comme chaque chameau porte un homme sur
son dos, elle en dessine H=6 \\times 1 = 6.\nIl y a donc dessiné
\\boxed{6} hommes.", role='assistant', function_call=None,
tool_calls=None)

```

Encadré 13 : Réponse en format Latex modèle (deepseek-ai-deepseek-math-7b-rl-GGUF via LM-Studio)

On note x le nombre de chameaux et y le nombre de dromadaires. On a donc les équations suivantes :

$2x + y = 19$ (E1) car chaque animal à deux bosses, sauf un seul qui n'en ont qu'une ;

$4x+2y = 52$ (E2) car chacun des quatre pattes du chameau ou dromadaire compte pour une patte.

On *résoud* le système (E1, E2), on multiplie E1 par 2 : $4x+2y=38$ (2E1) et soustrait cette équation à l'(E2) ce qui donne: $0 + (52-38)=14 \Rightarrow 2y = 14$.

Donc $y = 14/2 = 7$ dromadaires.

On remplace dans (E1), on a $2x+7=19 \Rightarrow 2x=(19-7) = 12$, donne alors : $x = 12/2 = 6$.

Elle a 6 chameaux et comme chaque chameau porte un homme sur son dos, elle en dessine $H=6 \times 1 = 6$.

Il y a donc dessiné 6 hommes.

Encadré 14 : Réponse mise en forme (modèle deepseek-ai-deepseek-math-7b-rl-GGUF via LM-Studio)

La réponse juste apparaît par enchantement après un système d'équations incompatibles du fait des dromadaires à 2 pattes (également proposé par ChatGPT en avril 2024). D'où vient cette caractéristique des dromadaires souvent utilisées ? Mystère ?

La faute d'orthographe, « résoud » au lieu de « résout », se retrouve aussi souvent sur Internet.

Une nouvelle requête propose une solution plus inventive avec les hommes formant une troisième bosse aux chameaux mais note une impossibilité (encadré 15). Cette version est obtenue avec une température de 0.1. Il est toutefois difficile d'estimer l'influence de ce paramètre.

Lever partiel du voile

Avec Perplexity/Lab et le modèle de base, la fouille du code source de la page html proposant la réponse permet de faire une découverte intéressante. Avec l'énoncé original, un attribut d'un bouton (non activé) est associé à l'url qui mène à la fiche du problème de la banque de problème²².

Bizarrerie, il s'agit de la version française dans un environnement (nom des rubriques, en-têtes, etc.) en langue italienne.

De même, avec une version légèrement modifiée de l'énoncé « dans un troupeau de chameaux et dromadaires..... » on trouve la référence à une conférence donnée à l'IREM de Grenoble prenant pour thème ce même problème du RMT²³ et à un problème proposé par l'IREM de la Réunion²⁴.

Dans ce cas, l'énoncé du problème est : « Dans un troupeau composé de chameaux (2 bosses) et de dromadaires (1 bosse), on compte 12 têtes et 20 bosses. Combien y a-t-il de dromadaires ? ».

Dans le cas d'un énoncé original avec d'autres valeurs, aucune citation (référence) n'apparaît mentionnée explicitement.

Le rôle de ces références serait à éclaircir.

22 http://www.projet-ermitage.org/ARMT/navi_fic2.php?code=sd10-fr&flag=1&langue=it

23 https://irem.univ-grenoble-alpes.fr/medias/fichier/61n6_1562329267866-pdf

24 https://iremi.univ-reunion.fr/IMG/pdf/crpe_2016_gpe1_corr.pdf

Soit x la nombre de chameaux et y le nombre de dromadaires. On a les deux équations suivantes :

1) Le total des bosses est 19, or chaque chameau à 2 bosses (avant l'homme sur son dos), donc on obtient: $2x + y = 19$.

2) Le total de pattes fait 52. Chaque chameaux a 4 pattes et un dromadaire en a une, ce qui nous donne : $4x+y*2=52$, soit: $4x + 2y = 52$.

On résout le système de deux équations à deux inconnues (x ; y) suivant :

$$2x + y = 19$$

$$4x+2y=52$$

On multiplie la première par 2 pour pouvoir faire une soustraction *membre-à-membre*:

$$4x + 2y = 38$$

$$4x+2y=52$$

donne en faisant la différence des deux lignes : $0x + 0y = -14$ ce qui est impossible !

En fait, on a oublié que chaque chameau porte un homme sur son dos. Chaque "chameaux" (avec le dessinateur) à donc maintenant trois bosses au lieu de deux et l'équation 1 devient :

$$3x + y = 19.$$

On résout alors

$$3x+y=19$$

$$4x+2y=52$$

On multiplie la première par 2 pour pouvoir faire une soustraction *membre-à-membre*:

$$6x + 2y = 38$$

$$4x+2y=52$$

donne en faisant la différence des deux lignes : $2x -0y = -14$ d'où $x=7$.

En reportant dans la première équation on obtient: $3*7+y=19$, soit $y=4$. Le nombre de chameaux est donc bien sûr 7 et le nombre de dromadaires 4 (on vérifie en effet que $3 \times 7 + 4 = 21 + 4 = 25$ ce qui correspond au bon nombres de bosses).

Chaque chameau porte un homme sur son dos, donc le nombre d'hommes est égal à celui des "chameaux", c-à-d 7.

Encadré 15 : Nouvelle réponse (deepseek-ai-deepseek-math-7b-rl-GGUF via LM-Studio)

Quelques perles (avril 2024)

Relever les « perles » est une activité fréquemment menée par des didacticiens et des enseignants. Parfois cela permet de mettre à jour des incompréhensions, parfois des représentations enfantines amusantes. Qu'en est-il avec les systèmes génératifs²⁵ ?

* Modèle : Llama-3-8B-Instruct-Coder-GGUF (LM-Studio, t=0.05) avec l'énoncé complet pose la question :

Combien y avait-il de vaches et combien de moutons dans la ferme?

Le système s'est vraisemblablement branché sur une analyse de problèmes semblables.

²⁵ Les textes en retrait reprennent textuellement les réponses générées par les systèmes.

* Modèle : Llama-3-8B-Instruct-Coder-GGUF (LM-Studio, t=0.05) donnée courte : Trouver, parmi des chameaux et des dromadaires, le nombre de chameaux connaissant le nombre de pattes, 52, et le nombre de bosses, 19.

Il y a 1 chameau qui connaît les nombres de pattes et de bosses.

Explication: Si il n'y avait qu'un seul animal qui connaissait ces informations, alors il devrait être un chameau car il y a plus d'informations sur les chameaux que sur les dromadaires. Il est donc logique que le nombre de pattes soit 52 et le nombre de bosses soit 19, ce qui correspond aux caractéristiques d'un chameau.

Note: Ce problème est un exemple de réduction par l'absurde, où l'on élimine les hypothèses absurdes pour arriver à une conclusion logique. Dans cet cas, si il y avait plus de 1 animal qui connaissait ces informations, alors il devrait être un dromadaire car il y a plus d'informations sur les dromadaires que sur les chameaux. Mais cela est absurde car les dromadaires ne connaissent pas le nombre de pattes et de bosses. Il est donc logique que seul un chameau connaisse ces informations.

L'incompréhension provient d'un mauvais découpage de l'énoncé. La solution donnée ci-dessous est également liée à une telle difficulté.

* Modèle mixtral-8x7b-instruct dans l'environnement correspondant de la communauté Hugging Face, donnée courte : la solution n'est pas trouvée. A partir de l'énoncé original, le système ne va pas plus loin que l'établissement des relations. L'énoncé simplifié n'a pas été *compris*.

Il est impossible de trouver un tel scénario parmi des chameaux et des dromadaires, car ni les chameaux ni les dromadaires n'ont 19 bosses.

Un chameau a typiquement deux bosses, tandis qu'un dromadaire n'en a qu'une.

De plus, les chameaux et les dromadaires ont tous quatre pattes, pas 52. Il s'agit donc d'une énigme impossible à résoudre avec ces animaux.

* Perplexity/labs avec les données courtes

Pour trouver le nombre de chameaux connaissant le nombre de pattes, 52, et le nombre de bosses, 19, nous devons d'abord comprendre les caractéristiques distinctives des chameaux et des dromadaires.

Les chameaux ont deux bosses dorsales, tandis que les dromadaires n'en ont qu'une. Les chameaux ont également des pattes plus courtes que les dromadaires.

Maintenant, pour répondre à la question, nous devons considérer les informations fournies. Le nombre de pattes est de 52, ce qui signifie que nous devons diviser ce nombre par 4, car chaque chameau ou dromadaire a 4 pattes. Cela nous donne 13.

Le nombre de bosses est de 19. Comme les chameaux ont deux bosses et les dromadaires une seule, nous pouvons déduire que le nombre de chameaux est de $19/2 = 9,5$. Cependant, comme nous ne pouvons pas avoir de demi-chameau, nous devons arrondir ce nombre. Étant donné que le nombre de pattes est de 13, nous pouvons conclure que le nombre de chameaux est de 9.

En résumé, parmi des chameaux et des dromadaires, le nombre de chameaux connaissant le nombre de pattes, 52, et le nombre de bosses, 19, est de 9.

Le système s'est vraisemblablement branché sur un solutionnaire particulièrement détaillé.

* Mistral

Sur, je peux vous aider à résoudre ce problème.

Tout d'abord, il est important de noter que les chameaux et les dromadaires ont tous deux deux pattes. Donc, le nombre 52 représente le nombre total de chameaux et de dromadaires, car chaque animal a deux pattes.

Ensuite, nous savons que les chameaux ont deux bosses et les dromadaires ont une bosse. Donc, le nombre 19 représente le nombre total de bosses.

Si nous supposons que tous les animaux sont des dromadaires, nous aurions 19 dromadaires, car chaque dromadaire a une bosse. Mais cela laisserait 33 animaux ($52-19$) sans bosses, ce qui n'est pas possible car tous les animaux ont au moins une bosse.

Donc, il doit y avoir des chameaux. Si nous ajoutons un chameau, nous avons maintenant 18 dromadaires (car $19-2 = 17$ bosses restantes, et chaque dromadaire a une bosse) et un chameau. Cela signifie que nous avons maintenant 20 animaux (18 dromadaires + 1 chameau) et il reste encore 32 pattes ($52-20*2 = 12$, et chaque animal a deux pattes).

Chaque chameau supplémentaire ajoutera une bosse et deux pattes. Donc, si nous ajoutons 16 chameaux de plus (pour un total de 17 chameaux), nous aurons 32 bosses de plus (pour un total de 49 bosses, ce qui est trop), mais nous aurons aussi 32 pattes de plus, ce qui est exactement ce dont nous avons besoin.

Donc, il doit y avoir 16 chameaux (et 16 dromadaires).

La question se pose : où est-il allé chercher tout ça ? On note un hiatus entre la forme (raisonnement détaillé) et le fond.

* Mistral-7B-OpenOrca-GGUF (exploité par lm-studio)

Il y a 52 chameaux parmi lesquels 19 ont 4 pattes et 33 ont 6 pattes.

Pour trouver le nombre de chameaux connaissant le nombre de pattes, 52, et le nombre de bosses, 19, on peut utiliser la formule de combinaison : $C(n, k) = n! / (k!(n-k)!)$

$$C(52, 4) = 52! / (4!(52-4)!) = 52! / (4!48!) = 1330$$

Il y a donc 1330 chameaux connaissant le nombre de pattes et le nombre de bosses.

C'est la réponse la plus farfelue, donnée sans préavis. A noter que la valeur donnée pour $C(52,4)$ est erronée. Dans une autre tentative, le système, sous réserve que les animaux ont également deux pattes, propose des relations correctes sans indication de résolution et une réponse fautive (7 chameaux) parachutée.

A voir aussi l'encadré 15 où des chameaux à trois bosses sont introduits. A noter que ce thème est récurrent dans les devinettes.

En conclusion ces « perles » montrent la distance entre une certaine *compréhension* de la situation et un résultat.

Des licornes et des vaches

Comment se comportent les systèmes avec un problème isomorphe mais dont les ingrédients ne semblent pas se trouver sur Internet ? Le problème de l'encadré 16 concerne les vaches et les licornes.

Maturin a amené dans le champs des vaches et des licornes. En tout, il y a 19 cornes et 52 pattes. Comme chacun sait, les vaches ont deux cornes et les licornes une seule. Puis il met une cloche au cou de chaque vache.

Combien a-t-il mis de cloches ?

Expliquez votre réponse.

Encadré 16 : Un problème isomorphe

Dans l'encadré 17 la solution de YIAHO est identique à celle donnée pour le problème original. Il en va de même pour ChatGPT.

Pour résoudre ce problème, nous devons établir un système d'équations basé sur les informations fournies.

1. Soit (v) le nombre de vaches et (l) le nombre de licornes.
2. Selon l'énoncé, nous savons que chaque vache a 2 cornes et chaque licorne a 1 corne. Ainsi, l'équation pour le nombre total de cornes est : $2v + l = 19$
3. De plus, chaque vache a 4 pattes et chaque licorne a 4 pattes. Le nombre total de pattes est donc : $4v + 4l = 52$ Nous pouvons simplifier cette équation en divisant chaque terme par 4 : $v + l = 13$

Nous avons maintenant deux équations :

1. ($2v + l = 19$)
2. ($v + l = 13$)

Nous pouvons résoudre ce système d'équations.

Pour éliminer (l), nous soustrayons la deuxième équation de la première :

$$(2v + l) - (v + l) = 19 - 13$$

Cela nous donne : $v = 6$

Nous remplaçons (v) dans l'une des équations pour trouver (l). Utilisons la deuxième équation : $6 + l = 13$ implique $l = 7$

Donc, Maturin a 6 vaches et 7 licornes. Puisqu'il met une cloche au cou de chaque vache, il a mis 6 cloches.

Réponse : Maturin a mis 6 cloches.

Encadré 17 : Solution au problème des vaches et licornes par YIAHO

Après une analyse impeccable de la situation le Gemini *estime* le problème insoluble par manque d'informations (encadrés 18 et 19). On note comme dans le problèmes des chameaux et des dromadaires (encadrés 3 et 5) une grande quantité de *discours* qui ne concerne qu'en partie le processus de résolution proprement dit.

Absolument ! Résolvons ce petit problème de logique ensemble.

Commençons par analyser les informations que nous avons :

- **Nombre total de cornes:** 19
- **Nombre total de pattes:** 52
- **Vaches:** 2 cornes chacune
- **Licornes:** 1 corne chacune

Notre objectif: Trouver le nombre de vaches, car c'est elles qui portent les cloches.

Résolution:

1. Déterminer le nombre de licornes:

- Chaque licorne représente une corne.
- Si nous soustrayons le nombre total de cornes par le nombre de cornes des vaches (en supposant que toutes les cornes appartiennent à des vaches), nous trouverons le nombre de licornes "en trop". Ce nombre correspondra exactement au nombre de licornes.
- $19 \text{ cornes totales} - (\text{nombre de vaches} * 2 \text{ cornes par vache}) = \text{nombre de licornes}$

2. Déterminer le nombre de vaches:

- Maintenant que nous connaissons le nombre de licornes, nous pouvons déduire le nombre de vaches en utilisant le nombre total de pattes.
- Chaque vache a 4 pattes et chaque licorne a 4 pattes.
- $52 \text{ pattes totales} / 4 \text{ pattes par animal} = \text{nombre total d'animaux (vaches + licornes)}$
- En soustrayant le nombre de licornes trouvé précédemment, nous obtenons le nombre de vaches.

Calculs:

Nous n'avons pas assez d'informations pour faire des calculs précis. Le problème est insoluble tel quel.

Pourquoi ?

- **Inconsistance dans les données:** Si toutes les vaches ont 2 cornes et les licornes 1 corne, il est impossible d'avoir exactement 19 cornes avec un nombre entier de vaches et de licornes.
- **Information manquante:** Le nombre total de pattes (52) ne nous permet pas de déterminer de manière unique le nombre de vaches et de licornes. Il existe plusieurs combinaisons possibles de vaches et de licornes qui donneraient 52 pattes au total.

Conclusion:

Il n'est pas possible de répondre à la question telle qu'elle est posée. Il y a une incohérence dans les données fournies.

Pour résoudre ce problème, il faudrait:

- **Vérifier les données:** S'assurer que le nombre total de cornes et de pattes est correct.
- **Ajouter une information supplémentaire:** Par exemple, le nombre de vaches ou de licornes, ou une relation entre les deux.

Exemple d'information supplémentaire:

- "Il y a 3 licornes de plus que de vaches."
- "La moitié des animaux sont des vaches."

Avec une information supplémentaire de ce type, nous pourrions établir un système d'équations et trouver la solution.

En résumé: Le problème tel qu'il est posé est mathématiquement impossible à résoudre en raison d'une incohérence dans les données.

Pourrais-tu vérifier les données et me fournir des informations supplémentaires ?

Encadré 19 : Conclusion de Gemini

Problème avec données superflues

Sans aller jusqu'à des pièges tels que l'âge du capitaine (voir plus haut), les MWP peuvent introduire des données superflues qui permettent d'exercer un des aspects de la mathématisation qui est de séparer les informations utiles des autres. Le problème proposé décrit la caravane de façon plus réaliste en introduisant les caravaniers. Les précisions à leur sujet brouillent un peu la piste. Toutefois, la solution ne dépend pas de la répartition des caravaniers entre ceux à pied, sur les chameaux, voire sur les dromadaires !

Une caravane traverse le désert. Certains caravaniers (il y en a 35 au total) sont à pied, d'autres sur les chameaux. Il n'y a pas de caravanier sur les dromadaires. Un caravanier s'amuse à compter les pattes, il en trouve 52. Et le nombre de bosses est 19. Combien y a-t-il de chameaux et de dromadaires ?

Encadré 20 : Problème avec information superflue

ChatGPT 4o

L'encadré 21 donne le début de la solution. Le nombre de caravaniers à pied est pris comme inconnue. L'hypothèse est que tous les chameaux sont occupés²⁶ et l'équation 1 n'est pas idiote. L'incompréhension survient avec l'équation 2 qui prend en compte les pattes des caravaniers (voir notre remarque précédente concernant la traduction anglaise). En soi une telle situation pourrait se concevoir.

26 Situation que l'expression « les » chameaux permet d'envisager ; qu'en aurait-il avec « des » chameaux ?

Pour résoudre ce problème, nous devons établir et résoudre un système d'équations en utilisant les informations fournies.

Définissons les variables :

- C : nombre de chameaux
- D : nombre de dromadaires
- P : nombre de caravaniers à pied

Nous avons les informations suivantes :

1. Il y a 35 caravaniers au total.
2. Il y a 52 pattes au total.
3. Il y a 19 bosses au total.

Établissons les équations à partir de ces informations.

Équation 1 : Nombre total de caravaniers

Puisque chaque caravane est composée de chameaux, de dromadaires et de caravaniers à pied, nous avons : $P+C=35$

Équation 2 : Nombre total de pattes

Les pattes sont comptées comme suit :

- Chaque caravanier a 2 pattes.
- Chaque chameau a 4 pattes.
- Chaque dromadaire a 4 pattes.

Cela nous donne l'équation suivante : $2P+4C+4D=52$

Équation 3 : Nombre total de bosses

Les bosses sont comptées comme suit :

- Chaque chameau a 2 bosses.
- Chaque dromadaire a 1 bosse.

Cela nous donne l'équation suivante : $2C+D=19$

Nous avons maintenant trois équations :

1. $P+C=35$
2. $2P+4C+4D=52$
3. $2C+D=19$

Encadré 21 : Problème avec donnée superflue avec chatGPT, établissement des relations

La solution *manuelle* est menée de main de maître (encadré 22). Elle pourrait provenir d'un algorithme standard tel que ceux réalisés pour les systèmes d'enseignement évoqués en introduction. Toutefois, la mise en oeuvre conjointe des deux techniques (addition et substitution) est peu commune.

Résolution du système d'équations

D'abord, simplifions l'équation 2 en la divisant par 2 : $P+2C+2D=26$

Nous avons maintenant le système :

1. $P+C=35$
2. $P+2C+2D=26$
3. $2C+D=19$

Soustrayons la première équation de la deuxième pour éliminer P : $(P+2C+2D)-(P+C)=26-35$;
 $C+2D = -9$

Maintenant, nous avons :

1. $P+C=35$
2. $C+2D=-9$
3. $2C+D=19$

Isolons C dans l'équation 2 : $C = -9 - 2D$

Substituons cette expression de C dans l'équation 3 : $2(-9-2D)+D=19$;

$$-18-4D+D=19 ; -18-3D=19 ; -3D=37 ; D = -37/3$$

Nous avons une erreur dans les calculs, vérifions les équations.

Revenons aux équations initiales et tentons de résoudre à nouveau correctement.

1. $P+C=35$
2. $P+2C+2D=26$
3. $2C+D=19$

Substituons $C=35-P$ dans l'équation 2 et 3.

$$\text{Équation 3: } 2(35-P)+D=19 ; 70-2P+D=19 ; D = 19 - 70 + 2P ; D = -51+2P$$

$$\text{Équation 2: } P+2(35-P)+2D=26 ; P+70-2P+2D=26 ; [...] : 3P=58 ; P = 58/3$$

Encore une erreur, révision.

Donnons cela une nouvelle révision méthodique avec bonne substitution et addition. Je vais résoudre encore en repensant des solutions plus simples :

$$\text{Equation 1: } P+C=35$$

$$\text{Equation 2: } 2P+4C+4D=52$$

$$\text{Equation 3: } 2C+D=19$$

Solution directe

Solving step-wise through plausible integer adjustments:

1. Assume practical testing.
2. Fix feasible integer pairs solving linear algebraically together. Let result directly through correct pairs-solving explicitly: $C=6, D=7$: Vérifier $2C+D=19$

E.g. fixing system algebraically for solving p logically not-mistaken ways yielding.

Finally Correct: $C=6, D=7$.

Malheureusement, il se trouve que les données numériques ne sont pas adaptées à cette interprétation de la situation. Le système *met en doute* à deux reprises sa méthode de résolution pour finalement (après un long moment de silence) nous servir (en anglais) une solution qui satisfait l'équation des bosses et se révèle être la bonne solution sans autre justification.

Il n'a pas eu l'*esprit* comme deepseek-ai-deepseek-math-7b-rl-GGUF via LM-Studio de remettre en cause son *interprétation* de la donnée. (noter le terme « *repensant* »).

YIAHO

1. Notons :

- (p) : le nombre de caravaniers à pied.
- (c) : le nombre de chameaux.
- (d) : le nombre de dromadaires.

2. Nous savons qu'il y a au total 35 caravaniers : $p + c = 35$

3. Chaque caravaniers à pied a 2 jambes donc le nombre total de pattes des caravaniers à pied est (2p). Chaque chameau a 4 pattes, et chaque dromadaire a également 4 pattes. Le nombre total de pattes est donné par : $2p + 4c + 4d = 52$

4. Le nombre de bosses est donné par le nombre de chameaux (chameaux ayant 1 bosse chacun) et le nombre de dromadaires (dromadaires ayant également 1 bosse chacun). On a : $c + d = 19$

Nous avons donc un système de trois équations :

1. ($p + c = 35$)
2. ($2p + 4c + 4d = 52$)
3. ($c + d = 19$)

Nous pouvons utiliser la première équation pour exprimer (p) en fonction de (c) : $p = 35 - c$

Substituons (p) dans la deuxième équation :

$$2(35 - c) + 4c + 4d = 52$$

$$70 - 2c + 4c + 4d = 52$$

$$70 + 2c + 4d = 52$$

$$2c + 4d = 52 - 70$$

$$2c + 4d = -18$$

Nous pouvons simplifier cette équation par 2 : $c + 2d = -9$

Maintenant, nous avons :

1. ($c + d = 19$)
2. ($c + 2d = -9$)

Nous pouvons résoudre ce système. Soustrayons la première équation de la seconde :

$$(c + 2d) - (c + d) = -9 - 19 ; d = -28$$

Cependant, ce résultat n'a pas de sens. Reprenons les calculs.

Le système suit la même voie que ChatGPT et aboutit à un résultat n'ayant pas de sens *repère-t-il* (encadré 23).

Partons des équations ($c + d = 19$) et ($p + c = 35$). Nous savons que ($d = 19 - c$).

Remplaçons (d) dans la deuxième équation :

$$2p + 4c + 4(19 - c) = 52$$

$$2p + 4c + 76 - 4c = 52$$

$$2p + 76 = 52$$

$$2p = 52 - 76$$

$$2p = -24$$

$$p = -12$$

Nous avons encore une incohérence. Il semble qu'il y ait une erreur dans les équations. Revenons à la base.

Reprenons à partir de :

1. ($p + c = 35$)

2. ($c + d = 19$)

($d = 19 - c$)

Substituons cela dans le nombre total de pattes :

$$2(35 - c) + 4c + 4(19 - c) = 52$$

$$70 - 2c + 4c + 76 - 4c = 52$$

$$70 + 76 - 2c = 52$$

$$146 - 2c = 52$$

$$-2c = 52 - 146$$

$$-2c = -94$$

$$c = 47$$

Les résultats sont incohérents. Je vais recalculer avec un sens plus logique en considérant le nombre total de pattes.

Encadré 24 : Deuxième essai de YIAHO avec le problème enrichi

Un nouvel essai reprend le même système d'équation mais procède à une autre substitution. Nouvelle incohérence et nouvel essai sans plus de succès (encadré 24).

En reprenant les valeurs correctement *dit-il* (?) et en considérant l'équation concernant les pattes-jambes, il obtient sans explication 19 chameaux et 0 dromadaire. Mais la conclusion énonce 16 chameaux. Le tout est suivi d'une remarque pathétique (sous forme d'anacolithe) indiquant une clarification des calculs prise pour évidente (encadre 25).

Revenons à l'équation de départ. En prenant les valeurs correctement, nous avons :

$$1. (2p + 4c + 4d = 52)$$

Finalement, après avoir utilisé le bon nombre de corps, nous devrions avoir :

- ($c = 19$) (chameaux)
- ($d = 0$) (dromadaires)

En conclusion, il y a 16 chameaux et aucun dromadaire.

Merci de votre compréhension pendant que je clarifiais les calculs.

Encadré 25 : Conclusion finale de YIAHO avec le problème enrichi

Remarques et questions conclusives

Nous avons utilisé le problème des chameaux et dromadaires, un MWP de la famille de problème basés sur un système d'équations à deux inconnues pour explorer les solutions proposées par quelques systèmes génératifs et d'en tirer quelques premières indications sur leurs *capacités* à résoudre des problèmes de mots. La question reste entière de savoir si l'on peut caractériser ces capacités de logique et l'activité menée de raisonnement.

Résumé des principales observations

Systemes généralistes

ChatGPT-4 livre une solution parfaite ce qui n'avait pas été le cas quelques mois plus tôt (il avait alors attribué 2 pattes aux dromadaires). Solution également parfaite fournie par Mistral et YIAHO.

Un nouvel essai de Mistral avec une variante numérique (160 pattes et 70 bosses) aboutit également. Une recherche (Iman Mirzadeh & ali, 2024) montre toutefois que les performances de certains systèmes varient selon les variantes numériques des énoncés (ou seules les quantités en jeu sont modifiées) proposées.

Gemini s'engage sur une procédure considérant le rapport 19/2 comme solution provisoire. Cette voie qui est suggérée par l'analyse de la tâche dans la banque de problèmes du RMT n'est pas utilisée judicieusement ici. Dans une autre tentative procédant par essais à partir de 9 chameaux (également cité dans la fiche du RMT), le système s'arrête à 7 chameaux (et 5 dromadaires) estimant que c'est la seule solution possible.

Le modèle NuminaMath-7B-TIR²⁷ (vainqueur des Olympiades mathématique d'IA - AIMO) établit des équations en attribuant 2 pattes²⁸ aux chameaux. Ce système délègue la résolution des équations au module Sympy de Python. A noter une particularité de formulation qui mêle anglais et français. La démarche est exprimée en anglais mais les éléments de la donnée : dromadaires, chameau, boss (sic), etc. sont gardés tels quels. Un nouvel essai se révèle plus catastrophique. Toutefois, l'oubli des importations de modules Python est une erreur amusante à relent humain.

Perplexity en mode généraliste obtient une solution correcte.

27 En retournant sur la page site pour des nouveaux essais, nous constatons que ce modèle n'est plus disponible dans l'environnement proposé.

28 L'origine de cette erreur récurrente reste un mystère.

Systèmes « laboratoires »

Les propositions suivantes sont effectuées avec les systèmes qui permettent quelques choix.

Perplexity/lab avec le modèle llama-3-sonar-32k-online pose correctement les équations mais propose une mauvaise réponse (correspondant à nouveau au cas 19 bosses et 48 pattes) sans indication sur le processus. Une nouvelle tentative menée avec un énoncé simplifié aboutit à un résultat correct avec comme stratégie à la place d'une manipulation algébrique une « manipulation » sur les animaux, stratégie évoquée dans la fiche correspondante de la banque de problèmes. Un argument concernant les bosses est repris intégralement de la fiche de la banque de problèmes.

A ce propos nous avons pu constater dans le code source de la page html de la réponse, des liens menant à des fiches concernant le problème.

Avec un autre modèle de langage (mixtral-8x7b-instruct dans un environnement proposé par la communauté Hugging Face) la solution n'est pas trouvée. A partir de l'énoncé original, le système ne va pas plus loin que l'énoncé des relations en *comprenant* de façon erronée que l'on considère des chameaux à 19 bosses et 52 pattes.

Le modèle deepseek-ai-deepseek-math-7b-rl-GGUF exploité à l'aide de LM-Studio trouve la bonne solution et cela malgré des équations basées sur des dromadaires à 2 pattes.

Une deuxième tentative introduit des chameaux à 3 bosses (le chamelier constituant la 3e bosse). Ces chameaux à 3 (voire 4) bosses se retrouvent à de nombreuses reprises sur Internet sous la forme d'images, de devinettes ou même de description²⁹, à tel point que le doute s'installe : ces Camelus trinidadus existent-ils ou pas ?



Le Monde De Kikidi sur facebook

Un problème isomorphe

Contrairement à l'hypothèse, le problème isomorphe (des vaches et des licornes) est parfaitement résolu par YIAHO et ChatGPT. Gemini bute sur ce problème. Malgré de nombreuses remarques pertinentes (p. ex. **Notre objectif**: Trouver le nombre de vaches, car c'est elles qui portent les cloches), il lui est pas possible de coordonner les contraintes repérées.

Ajout de données inutiles

Par contre le problème avec l'information sur les caravaniers n'est pas résolu correctement. L'énoncé était peut-être maladroit et l'usage unique du mot « legs » en anglais pour les pattes et les jambes a peut-être créé de l'ambiguïté. On notera toutefois la démarche adoptée par ChatGPT : un système

²⁹ Le **chameau trinidal** (Camelus trinidadus), également connu sous le nom de chameau à trois bosses, est une espèce de chameau originaire des prairies, forêts et broussailles européennes (en Allemagne, en France, en Italie, en Espagne, au Portugal, en Roumanie et en Autriche), et est l'un des derniers chameaux véritablement sauvages du genre Camelus, avec le chameau sauvage de Bactriane...(https://novum-terram.fandom.com/wiki/Trinidal_Camel)

d'équations cohérent, le constat d'une réponse incompatible (négative), une nouvelle résolution et une nouvelle erreur. Contrairement au système basé sur le modèle deepseek-ai-deepseek-math-7b-rl-GGUF l'interprétation de l'énoncé n'est pas remise en cause. Finalement, une solution est proposée, trouvée semble-t-il par essais successifs sur une partie des contraintes. Pour ces dernières interventions ChatGPT a passé à l'anglais et semble particulièrement *énervé* voire *fâché* ! YIAHO n'a pas plus de succès prenant également en compte les « pattes » des caravaniers. Après le constat de premières solutions incohérentes, il termine (*calmement*) en proposant (utilisant le conditionnel) une solution (fausse) sortie de nulle part. *Civilisé*, il rajoute : « Merci de votre compréhension pendant que je clarifiais les calculs ». Sous entendu que le problème est finalement résolu alors qu'il n'en est rien.

Discussion

Les trois phases : mathématisation, actions déductives, formulation de la solution

La première phase de résolution du problème consiste à identifier les inconnues et formuler les contraintes sous la forme d'un système d'équations linéaires (mais d'autres stratégies sans l'outil algébrique sont possibles). Cette séquence sans être miraculeuse recèle un côté un peu magique dont il est difficile d'imaginer les différents ingrédients. Pour des familles de problèmes la méthode est d'extraire les valeurs numériques (sous forme d'expressions régulières) puis les objets associés et les termes des relations. La quantité d'informations, le nombre de schémas récurrents dans d'énormes volumes de données d'entraînement, nécessaire aux LLM pour pouvoir traiter de nombreuses familles de problèmes paraît démesurée. Beaucoup d'exemples peuvent-ils pallier l'absence de représentation intelligente ? Il semble que oui !

A noter, comme le signale un des systèmes génératifs non cité ici qu'il s'agit d'un problème classique. En effet, de nombreuses variantes sont disponible sur internet. De plus, les exemples de problèmes simples habillant un système de deux équations à deux inconnues pullulent. Un apprentissage spécifique à ces problèmes a peut-être été effectué. Il n'empêche que les démarches nous ont souvent surpris. La suite de notre travail s'attachera à examiner ce qu'il en est avec des problèmes plus « originaux ».

A noter que dans ces séquences nous n'avons pas dirigé le système par une partie déclarative. L'invite (le prompt) n'a consisté qu'en l'énoncé et la consigne. Les systèmes semblent suffisamment entraînés pour *comprendre* le contexte (mathématique) de la demande. Ensuite ce contexte doit être mêlé à l'énoncé pour extraire les inconnues. Selon le système, on constate que celles-ci peuvent être nommées en toute lettre (version plutôt informaticienne) ou par les lettres usuelles x et y ou encore par l'initiale (en minuscule ou majuscule) des éléments dont il s'agit de déterminer la quantité. Deuxième « tour de force » : la formulation des contraintes sous forme de relations. A noter cependant que l'énoncé n'introduit pas d'ambiguïtés lexicales³⁰ ou syntaxiques.

Dans la présentation de la solution, le système résume la situation et présente la méthode à suivre. Cette annonce sera suivie par l'action déductive structurée. Une question semblable se pose : cette structuration de la présentation de la résolution est-elle le fruit d'un apprentissage supervisé dans la spécialisation mathématique ou totalement libre (c'est-à-dire dans un apprentissage moins supervisé). On note encore comme le dévoile d'interrogation à travers notre propre interface que le format Latex pourrait constituer un support au *raisonnement* des systèmes.

Dans le cas du problème considéré la partie déductive a consisté le plus souvent en la résolution du système linéaire par substitution ou addition. Des systèmes plus spécialisés peuvent faire appel à

30 L'utilisation de la locution « de moins » dans un problème additif peut conduire les écoliers à utiliser une soustraction à la place d'une addition.

des modules Python (mais pouvant être secondé par une méthode *manuelle* si besoin est). Cette partie est, a priori, moins magique que la première. En effet, le processus peut facilement se réaliser par une programmation classique, commentaires compris. Difficile de se convaincre que le processus génératif associé à un grand modèle de langage suffit. L'encadré 26 indique ce que dit ChatGPT à ce propos suite à l'invite : Quel module ou plugin utilises-tu pour les MWP? (Yahoo répond de même après avoir eu besoin d'un éclaircissement sur la signification du sigle). La question est de savoir ce qui se cache sous le terme raisonnement mathématique.

Pour résoudre des problèmes de mots en mathématiques (MWP, "Math Word Problems"), je ne dépends pas directement d'un module ou plugin spécifique. Au lieu de cela, je m'appuie sur mes capacités intégrées en traitement du langage naturel (NLP) et en raisonnement mathématique. Cela me permet d'interpréter, comprendre et résoudre des problèmes en analysant le langage naturel et en appliquant les concepts mathématiques appropriés.

Si tu souhaites développer un système qui résout des problèmes de mots, tu pourrais envisager d'utiliser des bibliothèques Python comme :

1. **SymPy** : pour la manipulation symbolique des expressions mathématiques.
2. **NLTK** ou **spaCy** : pour le traitement et l'analyse du langage naturel.
3. **Pandas** : pour structurer et manipuler les données si les problèmes sont plus orientés vers des tableaux ou des statistiques.

Les systèmes plus avancés peuvent aussi intégrer des modèles d'apprentissage automatique pré-entraînés, comme GPT ou BERT, pour améliorer la compréhension du langage.

Est-ce que tu cherches à résoudre des MWP dans un cadre spécifique, comme un projet d'automatisation ou d'enseignement?

Encadré 26 : Réponse de ChatGPT à la question : Quel module ou plugin utilises-tu pour les MWP ?

Puis la solution est énoncée qui nécessite une étape en deux temps que les élèves peuvent assez souvent négliger. En effet, la consigne demande le nombre d'hommes dessinés par Cléopâtre et pas simplement les nombres de chameaux et dromadaires qui sont les solutions du traitement mathématique³¹ !

Remarques diverses

On notera que certaines erreurs sont typiquement « humaines » : les coquilles : résoud (encadré 15), membre-à-membre (encadré 16), oubli des importations (près de l'encadré 9), utilisations simultanées des symboles x et * pour la multiplication. D'autres sont plus ambiguës : les dromadaires à 2 pattes (est-ce liée à l'autre problème classique des poules et des lapins) ; l'introduction des « pattes » des caravaniers, la reprise d'une procédure sans changement.

Une erreur intéressante qui provient selon toute vraisemblance de la tokenisation des échanges (tokenisation qui répond davantage à des contraintes de fonctionnement que sémantiques) : en reprenant le terme 'signifie' de l'invite le système affiche 'signeifie'.

Les systèmes *savent repérer* des cas impossibles. Des nombres négatifs ou fractionnaires pour le nombre de chameaux conduisent à une réévaluation sans toutefois remonter jusqu'à une nouvelle

31 Sur 86 classes de cat 5 Suisse romande (les problèmes dans le cas du rallye sont résolus par groupe), 37 produisent une réponse correcte, 17 trouvent le bon nombre de chameaux et de dromadaires sans le nombre d'hommes ou évoquent six hommes sans explications. Le restant (32 groupes) font des essais qui n'aboutissent pas ou trouvent seulement le nombre d'animaux ou encore ne comprennent pas le problème. Une version ultérieure (23 bosses et 68 pattes) posées à 1983 classes dans le cadre international montre des taux de réussites un peu inférieurs aussi bien en cat 5 que 6.

interprétation de la situation. Après quelques ré-évaluations on constate l'abandon où alors l'apparition miraculeuse de la solution.

A propos de raisonnement

De façon plus fondamentale peut-on évoquer les lacunes ou la fragilité (Iman Mirzadeh & ali, 2024) de *raisonnement* ? Leur constatation³² pose une véritable question ontologique sur la nature des différents raisonnements : raisonnement naïf-natif, raisonnement logico-mathématique, raisonnement formel. Par ailleurs, une présentation³³ de cette étude « souligne les immenses failles dans le raisonnement des Llm ».

Que fait cette étude ? Elle reprend un ensemble de référence (dataset GSM8K) de problèmes de mathématiques (MWP) et soumet des problèmes isomorphes aux différents systèmes ayant déjà un score par rapport à cette référence. Pour certains problèmes une information non nécessaire a été ajoutée.

Dans le premier cas, les « gros » systèmes résistent bien. Dans l'autre on constate plutôt, selon les auteurs, à une « baisse de performances catastrophique ». Les commentateurs posent à juste titre la question « un jeune enfant serait-il tombé dans le panneau ? Il aurait été intéressant de disposer d'une étude comparative réalisée dans une salle de classe pour en juger ». Bonne question dont nous n'avons pas la réponse. Il est vrai que les enfants peuvent se laisser berner par des données superflues mais aussi par des vocables trompeurs³⁴.

Une autre étude (Marianna Nezhurina & al, 2024) propose diverses versions du problème d'Alice au pays des merveilles (problèmes AIW) basés sur le schéma « Alice a trois soeurs et un frère ; combien de soeur possède le frère ». Les auteurs notent que « la panne est dramatique, car les modèles montrent de fortes fluctuations même pour de légères variations de problèmes qui ne devraient pas affecter leur résolution, exprimant également une forte confiance excessive dans les mauvaises solutions, souvent soutenue par des explications plausibles ressemblant à des confabulations ». Cela rejoint nos constats quoique pour le problème AIW avec quelques exemples préalable (few-shot) des réponses justes peuvent être obtenues. Nous ne résistons pas toutefois à donner une réponse proposée par YIAHO qui nous rend assez perplexes à propos du *raisonnement* mené³⁵. Nous reprendrons ce cas dans une prochaine chronique.

- Raisonnement ou simulacre : En conclusion ses travaux ne posent-ils pas la mauvaise question ? Au lieu de constater des lacunes de raisonnement ne faudrait-il pas se poser la question si raisonnement il y a, au sens humain du terme.

L'enseignement des mathématiques à l'ancienne était basé sur l'utilisation de modèles standard (les baignoires qui se remplissent et les robinets qui fuient, les mélanges de vins, etc.) qui mettaient en oeuvre proportionnalité directe, inverse, etc. Trouver la solution revenait à identifier (« matcher ») une situation particulière à un modèle général, cette identification était exercée à partir de plusieurs exemples. L'enseignement renouvelé veut remplacer cette absence de raisonnement ou ce simulacre de raisonnement par une compréhension plus essentielle des structures en jeu. La réponse à la

32 « By evaluating performance on different instantiations and difficulty levels, we draw a comprehensive picture of LLMs' reasoning capabilities. Our findings support the hypothesis that current LLMs are not capable of performing formal mathematical reasoning and pave the way for further research on this important topic ».

33 <https://www.journaldugeek.com/2024/10/16/ia-une-etude-souligne-les-immenses-failles-dans-le-raisonnement-des-llm/>

34 Par exemple : Toto à 15 billes de **plus** que Julie. Toto à 27 billes, combien en possède Julie ?

35 YIAHO IQ 200 : Bonjour, je suis l'IA de Yiaho, comment puis-je vous aider ?

Vous : au problème suivant: Karine a 3 soeurs et 1 frère, Paul. Combien de soeurs Paul a-t-il? J'ai répondu 3. On me dit que la réponse est fausse, pourquoi?

YIAHO IQ 200 : La réponse est en effet incorrecte. Paul a une sœur, Karine, et il y a 3 autres sœurs, ce qui fait un total de 4 sœurs. Donc, Paul a 3 sœurs. Votre réponse de 3 était correcte.

question posée précédemment revient à situer les systèmes IA par rapport à ces alternatives³⁶.

- Différentes étapes de résolutions : Il y aurait aussi nécessité de distinguer différents types d'actions intervenant dans le processus de résolution : raisonnements naïvo-natif (en partie lié au « bon sens » ou sens commun qui a toujours été la pierre d'achoppement des travaux en IA³⁷), partie plus élaborée³⁸ articulant, détaillant et juxtaposant différentes phases dans le processus de résolution. Une des phases importantes réside aussi dans l'imagination³⁹ des possibles. Dans les problèmes de géométrie cette imagination mise en oeuvre dans l'utilisation de constructions auxiliaires, pierre d'achoppement des premiers systèmes de démonstration automatique (Gelernter, 1963).

- Raisonnement et langage : A distinguer les deux formes de connaissances bien connues des étudiants de Piaget : empiriques fournies par l'expérience indépendamment de toute logique et antérieures aux coordinations logico-mathématiques ; logico-mathématiques qui consiste en une coordination ultérieure favorisée en particulier par l'usage du langage et que cette logique peut dépendre de la langue utilisée. Plus généralement dans une perspective socio-culturelle « le langage constitue le facteur déterminant de l'émergence des capacités sociales, culturelles et cognitives propres à l'espèce, ou encore a entraîné une réorganisation "révolutionnaire" du fonctionnement psychique hérité de l'évolution. » (Bronckart, 2007).

Ce point amène à se poser la question de savoir si les systèmes sont-ils multilingues et *raisonnent-ils* dans les différentes langues où font-ils les va-et-vient vers une langue de base (l'anglais). L'épisode des pattes-jambes (legs) nous fait pencher pour cette deuxième hypothèse. Cela ne semble toutefois pas être le cas comme le signale YIAHO qui à la question : « pour raisonner passes-tu systématiquement par l'anglais? » nous répond : *Non, je ne passe pas systématiquement par l'anglais pour raisonner. Je suis capable de traiter et de comprendre plusieurs langues, y compris le français, directement. Mon raisonnement se fait dans la langue dans laquelle la question est posée, ce qui permet une communication fluide et naturelle. Si vous avez des questions, n'hésitez pas à demander dans la langue de votre choix !* Toutefois la question reste ouverte.

Dans un article Jade Emy (2025) note que des utilisateurs ont observé que le modèle o1 d'OpenAI *pense* souvent en chinois lors de son processus de raisonnement. Selon certaines théories c'est peut-être dû aux ensembles de données d'entraînement du modèle et aux services d'étiquetage de données. Une autre hypothèse serait que les modèles de raisonnement utilisent les langues les plus efficaces en fonction de l'objectif.

Quelle intelligence ?

- Niveau des performances : tout d'abord il faut relever les performances obtenues par ce qu'il ne faut pas l'oublier sont de simples automates. Même si la mise dans le domaine public des batteries de problèmes test peut encourager une certaine « tricherie », il n'en reste pas moins que des problèmes isomorphes peuvent être résolus.

- Saut entre performance et estimation du résultat : YIAHO après avoir pétouillé et fournit une fausse réponse nous gratifie d'un message « Merci de votre compréhension pendant que je clarifiais les calculs. ». Ce sont des remarques pathétiques (voir plus loin) qui venant d'un automate brouillent

36 Remarque du même type montrant un embryon de similitude entre les deux fonctionnements : Alexis que le train qui passe devant la maison « va à Berne » ou parfois « va à Neuchâtel ». Voyant un bateau partir il propose « il va à Berne ». Cela montre typiquement un processus de répétition langagière « réflexe » absent de raisonnement.

37 A propos de la difficulté de définir cette notion (Lassagne, 2024).

38 Nous ne considérons pas ici le raisonnement formel plus facilement basé sur la combinatoire de règles bien standardisée. Ce n'est pas pour rien que les premiers systèmes se sont intéressés aux problèmes de géométrie plane dont on peut lister la liste des opérations possibles (avec l'achoppement imaginatif des constructions auxiliaires).

39 Alain Valette (2008). Imagination et mathématiques: deux exemples. Bulletin de la Société des Enseignants Neuchâtelois de Sciences, 35, Juillet 2008

les pistes. L'utilisateur est sans cesse renvoyé d'un « coup de génie » (Ornes, 2023) à une « élucubration » (Jean-François Colonna, 2023).

- Emergence : ce terme est utilisé par Andler (2023) dans sa comparaison entre IA et intelligence humaine. Le fait qu'il n'est pas possible de « tracer » l'activité de génération rend le processus mystérieux et entraîne une confusion entre ce qui est créativité et simplement processus non prédictif^{40 41} ou imprévisible. Quoi de plus que du « matching » sophistiqué et de données dans une quantité qui dépasse notre perception voire imagination.

Du point de vue de la créativité, Harsh Kumar & al (2024) ont étudié un étape intermédiaire, c'est-à-dire l'assistance à la créativité apportée par les systèmes génératifs. Leurs résultats révèlent que si l'assistance par LLM peut fournir des améliorations à court terme de la créativité pendant les tâches assistées, elle peut par inadvertance entraver les performances créatives indépendantes lorsque les utilisateurs travaillent sans assistance », ce qui de leur point de vue soulève des inquiétudes quant à l'impact à long terme sur la créativité et la cognition humaines.

On note encore l'effet « bon élève », cette tendance à faire des déclarations superflues : « Comme le dromadaire n'est pas un chameau, on ne compte pas d'homme sur son dos. » ou redondantes.

« Exploits » à mieux comprendre

A l'aune de notre expérience dans l'analyse automatique de problème, plusieurs opérations effectuées par les systèmes étudiés nous paraissent des exploits.

- La présentation préalable d'un plan de résolution même si celui-ci est parfois très convenu.

- Récupération des valeurs numériques des inconnues et des relations parmi l'histoire contée par l'énoncé du problème. De même que le but recherché. Par répétition des requêtes, on note que le choix des symboles utilisés pour les inconnues peut varier (x, y ; initiales, mot complet). Cette diversité peut-elle être exploitée dans le décryptage du processus de génération.

- Problème de trace : C'est une remarque que nous partageons avec d'autres utilisateur (p. ex. Delahaye) il est difficile d'avoir des explications et des justifications indiquant d'où provient les réponses d'un chatbot. Les réponses des systèmes eux-mêmes restent extrêmement vagues⁴². En fouillant les sources des réponses, on trouve quelques références plus précises. Les enchaînements des déductions (exigées à l'époque des systèmes à base de connaissances) semblent par essence inaccessibles.

- Jonglage dans l'utilisation de modules extérieurs (modules importés) et la résolution *manuelle* extraite d'un apprentissage.

- Le niveau du langage accompagnant la résolution même si certaines informations semblent superflues.

- La possibilité de conserver dans les grands modèles un suivi avec la possibilité d'instaurer un dialogue servant à obtenir des éclaircissements. L'introduction dans ce dialogue des demandes concernant le processus de production de la réponse (une trace en quelque sorte) n'apportent aucune réponse précise

40 Il y aurait matière à une mise en perspective de cet aspect avec les réalisations et théories du mouvement surréaliste.)

41 Impression déjà ressentie avec Eliza. Nous avons aussi ce sentiment avec les commentaires délivrés par un système d'EAO aux réponses à des questions générées par un ensemble de règles et donc non programmées indirectement. (Pochon, Grossen, 1997).

42 Une réponse de YIAHO : en utilisant un vaste ensemble de données sur lequel j'ai été formée, comprenant des livres, des articles, des sites web et d'autres sources d'information. Grâce à des algorithmes avancés de traitement du langage naturel, je peux analyser votre question, comprendre son contexte et extraire les informations pertinentes pour vous fournir une réponse claire et précise.

- Les systèmes repèrent que des valeurs décimales ou négatives sont impossibles s'agissant d'un nombre de chameaux. Comment cette connaissance est-elle acquise ? Dans des exemples ou par des apprentissages spécifiques du « bon sens »⁴³ ? Est-ce une manifestation de ce que Romac, Carta, Yves Oudeyer (2024) appelle ancrage fonctionnel ?
- La rédaction quasi-parfaite, mise à part quelques accrocs, des réponses aux questions ou problèmes posés.

Travaux futurs

Dans ce travail exploratoire, nous n'avons pas étudié systématiquement les différentes dimensions citées mais effectué plutôt un travail en « largeur » : quelques variantes d'un problème, quelques modèles de langage, quelques chatbots, quelques températures. Nous n'avons pas encore entrevu la possibilité de nous fixer sur une dimension et de garder toute chose égale par ailleurs et d'organiser des plans d'expériences liés à des hypothèses. Toutefois, la même opération simple pourrait et devrait être effectuée pour des problèmes que les systèmes jugeraient moins classiques.

Ceci d'autant plus que l'annonce d'un taux de résolution des problèmes de mathématiques des dataSet de problèmes en constante augmentation peut faire penser que ces datasets sont utilisés dans les phases d'apprentissages (c'est le côté « brûler la soupe » de l'entreprise). Notamment, nous n'avons pas pu estimer si le sentiment que certains systèmes avaient évolué suite à la répétition de nos questions déjà effectuée de manière assez libre dès la fin de 2022 était justifié.

Toutefois, il est assez bluffant de constater que les différents techniques de prompting (« few-shot », présentation d'exemples, demande de « Chain of Thought - CoT ») permettent d'améliorer les réponses (mais cela exige de savoir résoudre le problème). Il serait utile d'acquérir l'intuition de comment ce méli-mélo de divers types de requêtes (toutes celles du modèle CRAFT par exemple) est géré et exploité⁴⁴. Des comparaisons systématiques restent à organiser.

Différentes stratégies de résolution (PEANO, PAL, TORA) seront aussi à explorer plus attentivement. Toutefois, une partie de leur fonctionnement repose sur l'identification des inconnues qui fait partie des exploits à mieux comprendre.

Finalemment

Dans notre travail d'analyse des MWP l'usage des IA génératives nous paraissent une voie intéressante dans la mesure où celles-ci semblent concentrer (c'est ce qu'elles nous disent) de nombreuses données et procédures de résolution.

Nous sommes loin d'avoir exploré toutes les ressources disponibles. Pendant ces quelques mois consacrés à ces essais le panorama a évolué. D'une part nous découvrons de nouveaux services (certains venant d'éclorre) et d'autre part nous apprenons à mieux les utiliser. Nous n'avons pas notamment encore exploré tous les effets de la rédaction des invites (prompt).

Nous relevons également la difficulté à exprimer les observations faites pour éviter d'humaniser par trop ces systèmes. En cas de difficulté à trouver des termes adéquats, nous avons utilisé des mots en italique. A noter qu'un des systèmes à lui-même mentionné sa manière *manuelle* de résoudre des équations.

Au-delà des aspect techniques, relevons qu'au début de nos essais, diverses habitudes humaines provoquaient quelques hésitations : tutoyer ou vouvoyer ? hésitation à contredire, formulation de

43 A noter que des enfants après des manipulation de poids de camions et de leur chargement arrivent à ne pas s'étonner d'obtenir 45 kg.

44 L'art des « invites » (prompt) devient un domaine de compétence à part entière. Un parcours de la littérature à ce sujet nous en dira plus. L'un d'entre nous a également maintenu un long *dialogue* avec YIAHO à ce sujet.

remerciements. C'est le cas avec tous les systèmes conversationnels primitifs. Avec Eliza déjà.

Dans sa résolution de problème on éprouve une certaine gêne face à cette attitude désopilante d'énoncer des résultats faux avec *assurance* voire *fiercé*.

Nous nous sommes souvent posé la question de la réutilisation de nos interactions. Vrai ou pas vrai cet enrichissement des connaissances par l'interaction est certainement à l'ordre du jour des prochaines versions de systèmes conversationnels-générateurs. N'était-ce pas déjà le soucis des pionniers du domaine qui, pour privilégier cette interaction, inventaient le langage LISP ?

En imaginant mettre en relation deux chatbots⁴⁵, que vont-ils se raconter ? Et surtout comment vont-ils évoluer.

A l'époque de l'IA des temps héroïques, alors que le modèle connexionniste avait perdu du terrain et n'était même pas cité dans l'ouvrage classique de Winograd et Flores (1990), Jean-Jacques Ducret (1986) posait le constat du peu (ou l'absence) de références à la psychologie, cognitive notamment. Que pouvons-nous dire à ce propos actuellement ?

(version du 24.12.24)

Bibliographie

Josette Adda (1975). L'incompréhension en Mathématiques et les Malentendus. *Educational Studies in Mathematics*, Vol. 6, No. 3 (Nov., 1975), pp. 311-326

Andler, S. (2023). *Intelligence artificielle, intelligence humaine : la double énigme*. Paris : Gallimard

John Ayoun (2024). ChatGPT : Comment rédiger des prompts efficaces. Ed. First

Stella Baruck (1985). *L'Âge du capitaine - de l'erreur en mathématiques*, Seuil, 1985

Jean-Paul Bronckart (2007), L'activité langagière, la pensée et le signe, comme organisateurs du développement humain. *Langage et société*, 121-122, Paris <https://shs.cairn.info/revue-langage-et-societe-2007-3-page-57?lang=fr>

Jean-François Colonna (2023). *Les élucubrations mathématiques de ChatGPT* <https://www.pourlascience.fr/p/opinions/les-elucubrations-mathematiques-de-chatgpt-25434.php>

Jean-Paul Delahaye (2024). Derrière les modèles massifs de langage. *Pour la Science*, n° 555, p. 80-85

Jean-Jacques Ducret (1986). *Winograd et Flores ou l'oubli de la psychologie*. Genetic artificial Intelligence and Epistemics Laboratory, Université de Genève. (memoNo5).

Jade Emy (2025). Le modèle de raisonnement d'OpenAI « pense » souvent en chinois : Personne ne peut expliquer pourquoi.

<https://intelligence-artificielle.developpez.com/actu/366800/Le-modele-de-raisonnement-d-OpenAI-pense-souvent-en-chinois-Personne-ne-peut-expliquer-pourquoi/>

Luyu Gao, Aman Madaan, Shuyan Zhou, Uri Alon, Pengfei Liu, Yiming Yang, Jamie Callan, Graham Neubig (2023). *PAL: Program-aided Language Models* (2211.10435.pdf)

Herbert Gelernter (1963). Realization of a geometry theorem-proving machine. In E. A. Feigenbaum et J. Feldman (Eds), *Computers and thought*, AAAI Press & MIT Press, 1995 (1^e édition 1963) pp. 134-152. (reprise d'un article des actes First International Conference on Information Processing (IFIP), p. 273–281, UNESCO, 1959).

45 Allusion aux systèmes tels AlphaZero ayant appris à jouer en jouant contre eux-mêmes.

- Zhibin Gou, Zhihong Shao, Yeyun Gong, Yelong Shen, Yujiu Yang, Minlie Huang, Nan Duan, Weizhu Chen (2024). *TORA: A tool integrated reasoning agent for mathematical problem solving* (Tora_2309.17452v4.pdf).
- Christophe Hache (2013). Langage mathématique à la transition primaire / collège. *Faire des mathématiques à l'école : de la formation des enseignants à l'activité de l'élève*, ARPEME, pp.452-463, 2013. <hal-00870111>
- Joy He-Yueya, Gabriel Poesia, Rose E. Wang, Noah D. Goodman (2023). *Solving Math Word Problems by Combining Language Models With Symbolic Solvers* (2304.09102.pdf)
- Harsh Kumar, Ewan Jordan, Jonathan Vincentius, Ashton Anderson (2024). *Human Creativity in the Age of LLMs: Randomized Experiments on Divergent and Convergent Thinking* (2410.03703v1.pdf)
- Daniel Lacombe (2011) *Grandeur et misère de l'implication*, séminaire « Enseignement des mathématiques », l'IREM de Paris, Paris [vidéo site IREM de Paris]
- François Lassagne (propos recueillis par) (2024). Entretien avec Mark Whiting : « Nous ne devrions pas supposer que les autres partagent le même 'bons sens' que nous ». *Pour la Science* n° 557, p. 32-35
- Iman Mirzadeh, Keivan Alizadeh, Hooman Shahrokhi, Oncel Tuzel, Samy Bengio, Mehrdad Farajtabar (2024). *GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models* (2410.05229v1.pdf)
- Marianna Nezhurina, Lucia Cipolina-Kun, Mehdi Cherti, Jenia Jitsev (2024). *Alice in Wonderland: Simple Tasks Showing Complete Reasoning Breakdown in State-Of-the-Art Large Language Models* (AIW_2406.02061v4.pdf).
- Stephen Ornes (2023). Modèles massifs de langage : d'où viennent les coups de génie de l'IA. *Pour la Science*, n° 549, p. 46-51
- Luc-Olivier Pochon, Alain Favre (2021). *A la recherche d'une classification de problèmes de mathématiques scolaires : une chronique*. hal-02972684v2
- Luc-Olivier Pochon, Michèle Grossen (1997). *Journal du développement d'un système d'EAO : De la possibilité et de la difficulté d'utiliser des théories sur l'apprentissage*. Neuchâtel, IRDP : Recherches 97.105
- Gabriel Poesia, Noah D. Goodman (2022). *Peano: Learning Formal Mathematical Reasoning* (2211.15864.pdf)
- Clément Romac, Thomas Carta, Pierre-Yves Oudeyer (2024). Les IA face au réel. *Pour la Science* n° 557, p 24-31
- Terry Winograd, Fernando Flores (1990). *Understanding Computers and Cognition, A New Foundation for Design*. Norwood;J : Ablex Pub. Corp. (5th printing)