Analyse critique de trois textes produits par l'IA Mistral-Vibe autour de la conjecture de Goldbach

Analyse réalisée avec Claude, à la demande de Denise Vella-Chemla

1^{er} juillet 2026

Résumé : Trois documents produits par l'outil Mistral-Vibe prétendent démontrer ou éclairer la conjecture de Goldbach : (1) une "preuve" par l'absurde utilisant la fonction π, (2) une "preuve" pour les classes $n \equiv 0, 2, 4 \pmod 6$ s'appuyant sur le crible de Brun, (3) une interprétation d'une dimension de Hausdorff 0.814 pour un ensemble baptisé "Ensemble de la Castafiore". Aucun des trois textes ne constitue une démonstration valide. Ce document identifie précisément, pour chacun, l'endroit exact où le raisonnement échoue, et montre que dans les trois cas la faille se situe exactement là où réside la véritable difficulté ouverte du problème de Goldbach.

1. Verdict résumé

Texte 1 ($\pi(n-2) \leq \pi(n-9)$)	Faux car π n'est *pas* strictement croissante ; la "contradiction" finale est fabriquée à partir d'un lemme faux sur π.
Texte 2 ($n = 6x, 6x+2, 6x+4$, crible de Brun)	Faux car le crible de Brun ne fournit jamais de minoration inconditionnelle positive (obstruction de parité de Selberg) ; de plus le texte cite littéralement le théorème final comme prémisse de sa propre preuve (circularité).
Texte 3 (dimension 0.814, Snurpf/Castafiore)	Non conclusif : l'argument "densité $\Rightarrow$ intersection non vide" n'est un théorème d'aucune théorie mathématique connue ; le lien entre 0.814 et les décomposants réels de n n'est jamais rigoureusement établi.

2. Texte 1 : la "preuve" par $\pi(n-2) \leq \pi(n-9)$

2.1. Reconstruction de l'argument

Soit n pair ≥ 6, $S_n = \{p \text{ premier impair} : 3 \leq p \leq n-2\}$. On suppose par l'absurde que pour tout $p \in S_n$, $n-p$ n'est pas premier. Comme $n-p$ est alors impair et composé, et que le plus petit composé impair est 9, on obtient $n - p \geq 9$ pour tout $p \in S_n$, donc $p \leq n-9$, donc

$$S_n \subseteq \{p \text{ premier impair} : 3 \leq p \leq n-9\}.$$

Ceci est *correct* : c'est une déduction valide de l'hypothèse absurde. En termes de cardinaux, $|S_n| \leq \pi(n-9) - 1$. Or par définition $|S_n| = \pi(n-2) - 1$. On obtient donc

$$\pi(n-2) \leq \pi(n-9).$$

Le texte affirme alors que ceci contredit la "stricte croissance" de π, puisque $n-2 > n-9$.

2.2. Le bug

Bug identifié : La fonction π n'est **pas** strictement croissante sur $[2, +\infty)$; elle est seulement croissante au sens large. Elle stagne strictement entre deux nombres premiers consécutifs : par exemple $\pi(8) = \pi(9) = 4$, ou encore $\pi(24) = \pi(25) = \pi(26) = \pi(27) = \pi(28) = 9$. On a $\pi(x) < \pi(x+1)$ *si et seulement si* $x + 1$ est premier.

L'inégalité $\pi(n-2) \leq \pi(n-9)$ n'a donc rien de contradictoire en général : elle est même l'égalité exacte $\pi(n-2) = \pi(n-9)$ dès qu'*aucun nombre premier ne se trouve dans la fenêtre* $[n-9, n-2]$ — ce qui est justement, on l'a vu, ce que l'hypothèse absurde implique. Autrement dit :

Fait : Sous l'hypothèse absurde, l'étape 3 démontre exactement qu'aucun premier ne se trouve dans $[n-9, n-2]$. Ceci est *cohérent* avec $\pi(n-2) = \pi(n-9)$, et non contradictoire avec elle. La "contradiction" de l'étape 6 n'existe que parce que le texte utilise un lemme faux sur π.

2.3. Contre-exemple numérique concret

Prenons $n = 98$. La fenêtre est $(89, 96] = \{90, \ldots, 96\}$. Aucun de ces sept entiers n'est premier (89 puis 97 sont les premiers encadrants — c'est le premier écart de longueur 8 dans la suite des nombres premiers). On a donc bien $\pi(96) = \pi(89)$, sans aucune contradiction générée par le mécanisme du texte. Le fait que $98 = 19 + 79$ soit une décomposition de Goldbach valide est obtenu par un $p = 19$ totalement extérieur à la fenêtre $[89, 96]$ — la preuve ne l'a jamais "vu".

2.4. Ce qu'il reste, une fois l'habillage retiré

Dépouillée de la fonction π, la preuve dit en substance : *"si un nombre premier p se trouve dans $[n-9, n-2]$, alors $n - p \in \{3, 5, 7\}$ est premier, donc $n = p + (n-p)$ est une décomposition de Goldbach."*. C'est vrai, mais c'est une tautologie de vérification directe, pas une preuve par l'absurde : elle ne garantit en rien l'*existence* d'un tel p proche de n, qui est précisément le cœur non résolu du problème (lié aux écarts entre nombres premiers, eux-mêmes non bornés de façon inconditionnelle par une constante fixe).

3. Texte 2 : la "preuve" pour $n = 6x$, $6x + 2$, $6x + 4$

3.1. Ce qui est correct

Les Lemmes 1 à 4 (réduction des classes de congruence interdites modulo 2 et 3 pour $n \equiv 0 \pmod 6$, densité des premiers dans les classes 1 et 5 mod 6 via Dirichlet) sont des faits élémentaires corrects. Il est vrai que le crible de Goldbach élimine une classe de moins modulo 6 pour $n \equiv 0 \pmod 6$ que pour les autres parités mod 6.

3.2. Le bug principal

Bug identifié : le Théorème 5 affirme : *"Le crible de Brun montre que le nombre de décomposants de Goldbach pour n est minoré par $C\,n/(\ln n)^2$"*. C'est faux tel quel. Les cribles classiques (Brun, Selberg) fournissent des **majorations** de cet ordre pour le nombre de représentations de n comme somme de deux nombres proches d'être premiers, jamais de **minoration inconditionnelle positive**. C'est le fameux *problème de parité* ("parity problem"), mis en évidence par Selberg lui-même : les méthodes de crible pur ne peuvent pas, par construction, distinguer un entier ayant un nombre pair de facteurs premiers d'un entier en ayant un nombre impair - elles sont donc structurellement incapables de détecter l'existence d'une représentation $n = p+q$ avec p, q tous deux premiers, quelle que soit la sophistication du crible utilisé.

Réduire de trois à deux le nombre de classes interdites modulo 6 ne change qu'une constante multiplicative dans la majoration ; cela ne lève en rien l'obstruction de fond, qui n'est pas une question de constante mais de nature de l'information que peut extraire un crible pur. Le meilleur résultat inconditionnel connu allant dans cette direction est le théorème de Chen (1966) : tout entier pair suffisamment grand s'écrit $n = p + P_2$ où p est premier et P_2 est premier ou produit de deux premiers - un résultat bien plus faible que Goldbach, obtenu avec des techniques de crible bien plus fines (crible de Selberg combiné à des estimations de grands cribles) que celles esquissées dans le texte.

3.3. Un second bug : la circularité

Bug identifié : À la section 3.4, dans la preuve du Théorème 6 (cas $n = 6x > 4 \times 10^{18}$), le texte écrit : *"D'après le Théorème CG, pour tout $n = 6x > 4 \times 10^{18}$, il existe au moins un décomposant de Goldbach."* Or le *Théorème CG* est défini plus loin, en section 4.4, comme le théorème général unifiant les trois cas $6x$, $6x + 2$, $6x + 4$ — c'est littéralement la conclusion globale que le document cherche à établir. Le texte utilise donc le résultat final comme prémisse d'une étape intermédiaire de sa propre démonstration : une circularité explicite.

On note aussi une incohérence de numérotation mineure mais révélatrice (section 3.4 cite un "Lemme 5" inexistant, la numérotation allant de Lemme 4 à Théorème 5) - signe que le texte n'a pas été relu avec la rigueur qu'il affiche.

4. Texte 3 : dimension de Hausdorff 0.814 et "Ensemble de la Castafiore"

4.1. Ce que le texte reconnaît lui-même

Le document note honnêtement que le théorème de Falconer sur l'intersection d'ensembles fractals ne s'applique pas ici, puisque

$$\dim_H(P) + \dim_H(K(n)) \approx 0.814 + 0.814 = 1.628 < 2.$$

4.2. Le bug

Bug identifié : pour compenser, le texte introduit un *"argument de densité"* : puisque P et $K(n)$ seraient tous deux "concentrés dans le même secteur" et de dimension de Hausdorff strictement positive, leur intersection serait non vide "par densité". Ceci n'est le théorème d'aucune théorie fractale existante. Deux ensembles de dimension de Hausdorff positive, même denses dans une même région du plan, peuvent parfaitement être disjoints : il suffit par exemple de construire deux ensembles de type Cantor dans $[0, 1]$ à l'aide de deux règles de construction (deux contraintes de congruence, deux développements en base différente, etc.) qui ne se rencontrent jamais, chacun ayant pourtant une dimension de Hausdorff strictement comprise entre 0 et 1.

De plus, ni P (l'"Ensemble de la Castafiore") ni $K(n)$ (les "chemins admissibles") ne reçoivent, dans ce texte ou par renvoi vérifiable, de définition rigoureuse qui les relie de façon démontrable aux décomposants réels de n. La valeur 0.814, même si elle est un résultat authentique de *box-counting* appliqué à votre construction Snurpf, reste donc un nombre sans contenu mathématique établi vis-à-vis de Goldbach : c'est un habillage fractal élégant posé sur un vide logique à l'endroit précis où il faudrait un théorème.

5. Sur l'observation initiale (3, 5, 7)

Votre remarque de départ est juste, simple, et jolie :

Fait : $(3, 5, 7)$ est l'unique triplet de nombres premiers impairs consécutifs. En effet, parmi trois entiers impairs consécutifs $m, m + 2, m + 4$, l'un des trois est nécessairement divisible par 3 (les résidus mod 3 de $m, m+2, m+4$ balaient les trois classes $0, 1, 2$). Pour $m > 3$, ce multiple de 3 est strictement supérieur à 3, donc composé. Le seul cas où les trois termes peuvent être premiers est donc $m = 3$, c'est-à-dire $(3, 5, 7)$.

C'est exactement ce fait, sous la forme "le plus petit composé impair est 9", qui sert de brique à l'argument du Texte 1. Le problème n'est pas que l'observation soit fausse - elle est vraie et élégante - mais qu'elle est *purement locale et qualitative* : elle dit qu'un troisième triplet de premiers consécutifs ne peut plus jamais se reproduire, mais elle ne dit rien de quantitatif sur la fréquence à laquelle un nombre premier se trouve *à proximité* d'un n donné, pour n arbitrairement grand. C'est précisément ce chaînon manquant - un contrôle des écarts entre nombres premiers, ou une minoration effective de type crible - que Mistral a dû combler dans les Textes 1 et 2, et il l'a fait par des énoncés faux (fausse stricte croissance de π, fausse minoration par crible de Brun, circularité).

6. Conclusion générale

Les trois textes partagent un même schéma, caractéristique d'une confabulation d'IA face à un problème ouvert difficile :

(i) ils partent d'une observation élémentaire vraie et souvent élégante (votre remarque sur $3, 5, 7$; les classes de congruence mod 6 ; une figure fractale réellement calculée) ;

(ii) ils l'habillent d'un appareillage formel impressionnant (fonction π, crible de Brun, dimension de Hausdorff) qui donne l'illusion de la rigueur ;

(iii) et à l'endroit précis où se situe la vraie difficulté du problème - contrôler les écarts entre nombres premiers, obtenir une minoration positive par un crible pur, ou établir qu'une intersection de fractales est non vide - ils insèrent un énoncé faux ou non justifié (fausse stricte croissance, minoration par crible inexistante, "argument de densité" sans théorème, voire une citation circulaire du résultat à prouver).

Aucun des trois textes n'est donc une démonstration, même partielle, de la conjecture de Goldbach. Cela ne retire rien à l'intérêt de votre observation de départ, ni à la légitimité d'explorer les classes de congruence mod 6 ou les représentations fractales des décomposants : ce sont des pistes de travail honnêtes. Le point de vigilance, pour la suite de vos échanges avec Mistral ou toute autre IA sur ce sujet, est de toujours demander explicitement : *"à quel endroit exact de la preuve utilisez-vous, implicitement ou explicitement, un contrôle des écarts entre nombres premiers, une minoration de crible, ou un résultat d'intersection de fractales - et ce résultat est-il un théorème publié et inconditionnel, ou une affirmation de plausibilité ?"* C'est systématiquement à cet endroit que les trois textes analysés ici échouent.