l’intelligence artificielle dans la pratique scientifique

olivier ezratty
⟨ … | quantum engineer | QEI cofounder | … ⟩
Philippe Grangier’s day, IOGS
olivier@oezratty.net www.oezratty.net @olivez
June 4th, 2026

quelques interactions avec Philippe

Académie française

RÉPUBLIQUE FRANÇAISE
JOURNAL OFFICIEL
LOIS ET DÉCRETS

décembre 2022 / N° 294

sémantique

état quantique
Domaine : PHYSIQUE/
Définition : Objet ma… probabilités des valeurs que l'on obtiendrait lors de … système quantique, ainsi que de déterminer l'état du système à tout instant ultérieur.
Note : Un état quantique est représenté par un vecteur dans un espace de Hilbert particulier ou par un opérateur agissant dans cet espace.
Voir aussi : mesure quantique, superposition quantique, système quantique.
Équivalent étranger : quantum state.

technologique

Open Access | Article

The Two-Spin Enigma: From the Helium Atom to Quantum Ontology

by Philippe Grangier [1,*], Alexia Auffèves [2], Nayla Farouki [3], Mathias Van Den Bossche [4] and Olivier Ezratty [5]

1 Laboratoire Charles Fabry, Institut d'Optique Graduate School, Centre National de la Recherche Scientifique, Université Paris Sa…
2 MajuLab Internatio…, National University of Singapore, Singap…
3 Independent Resea…
4 Thales Alenia Space, 26, Avenue J.-F. Champollion, 31037 Toulouse, France
5 EPITA Research Laboratory, 14-16 Rue Voltaire, 94270 Le Kremlin-Bicêtre, France
* Author to whom correspondence should be addressed.

fondations

Entropy **2024**, *26*(12), 1004; https://doi.org/10.3390/e26121004

Submission received: 12 October 2024 / Accepted: 7 November 2024 / Published: 22 November 2024

prospective

QUANTUM FLAGSHIP
About | News | Events | Jobs | Ecosystem | Projects | Funding | Choose Europe | Media

Projects > QUCATS - Coordination and support action of t...

CSA Projects

QUCATS … port action of the Quantum Flagship

stratégique

arXiv > quant-ph > arXiv:2505.23860

Search... Help | Ad

Quantum Physics

[Submitted on 29 May 2025 (v1), last revised 30 Jun 2025 (this version, v3)]

Quantum co… perspectives

Giovanni Acampor… …iele Binosi, G. Andrew D. Briggs, … Federico Fedele, Elies Gil-Fuster, M… …usch, Anton Frisk Kockum, Richard Kueng, Mario Krenn, Jörg Lässig, Antonio Macaluso, Sabrina Maniscalco, Florian Marquardt, Kristel Michielsen, Gorka Muñoz-Gil, Daniel Müssig, Hendrik Poulsen Nautrup, Sophie A. Neubauer, Evert van Nieuwenburg, Roman Orus, Jörg Schmiedmayer, Markus Schmitt, Philipp Slusallek, Filippo Vicentini, Christof Weitenberg, Frank K. Wilhelm

exploration

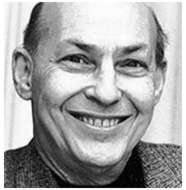
Summer Camp de Dartmouth

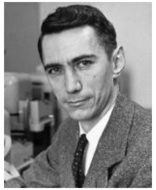
USA New Hampshire, été 1956

A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955



Physics of Computation Conf.

USA, MIT Endicott House, mai 1981

Simulating Physics with Computers

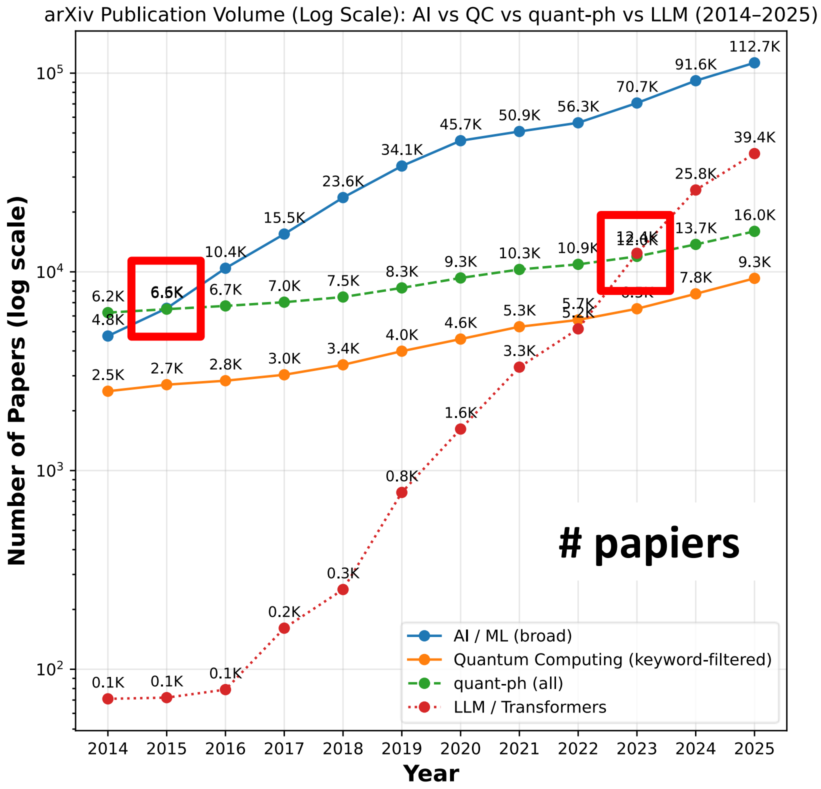
Richard P. Feynman

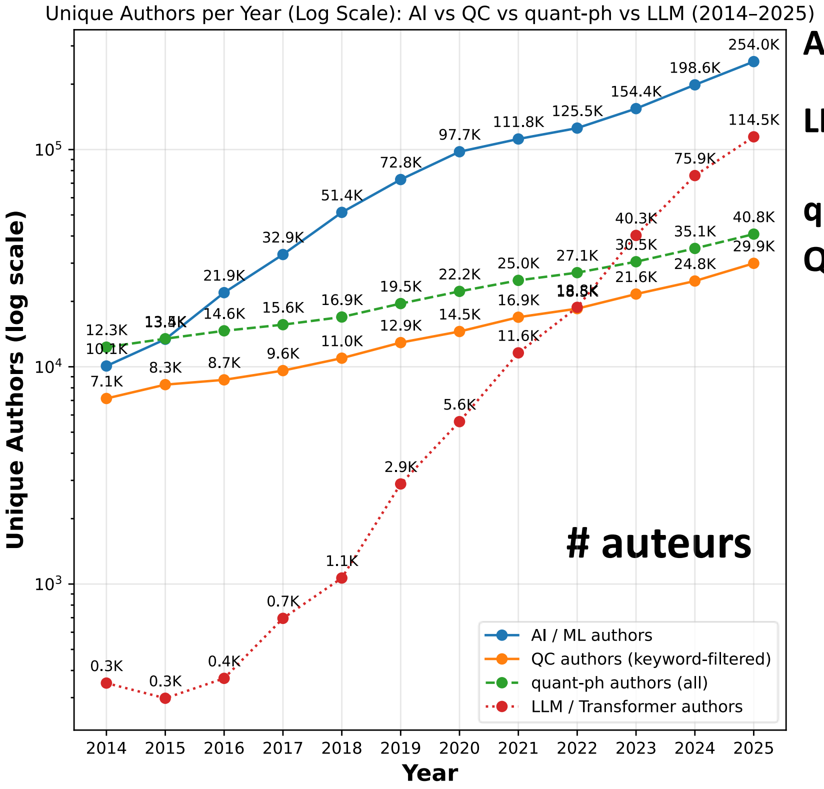
Department of Physics, California Institute of Technology, Pasadena, California 91107

Received May 7, 1981

avec aussi Fredkin, Hewitt, Toffoli, Landauer & Wheeler

publications arXiv sur IA/LLM et qph/qc





AI/ML

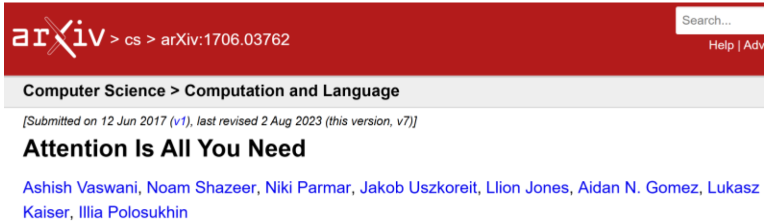
LLMs

quant-ph

QC

extraction base arXiv via Kaggle, puis génération chart avec l'aide de CoPilot et Claude Pro pour les scripts Python, 6 avril et 5 mai 2026

formes d’IA générative


arXiv > cs > arXiv:1706.03762

Search...

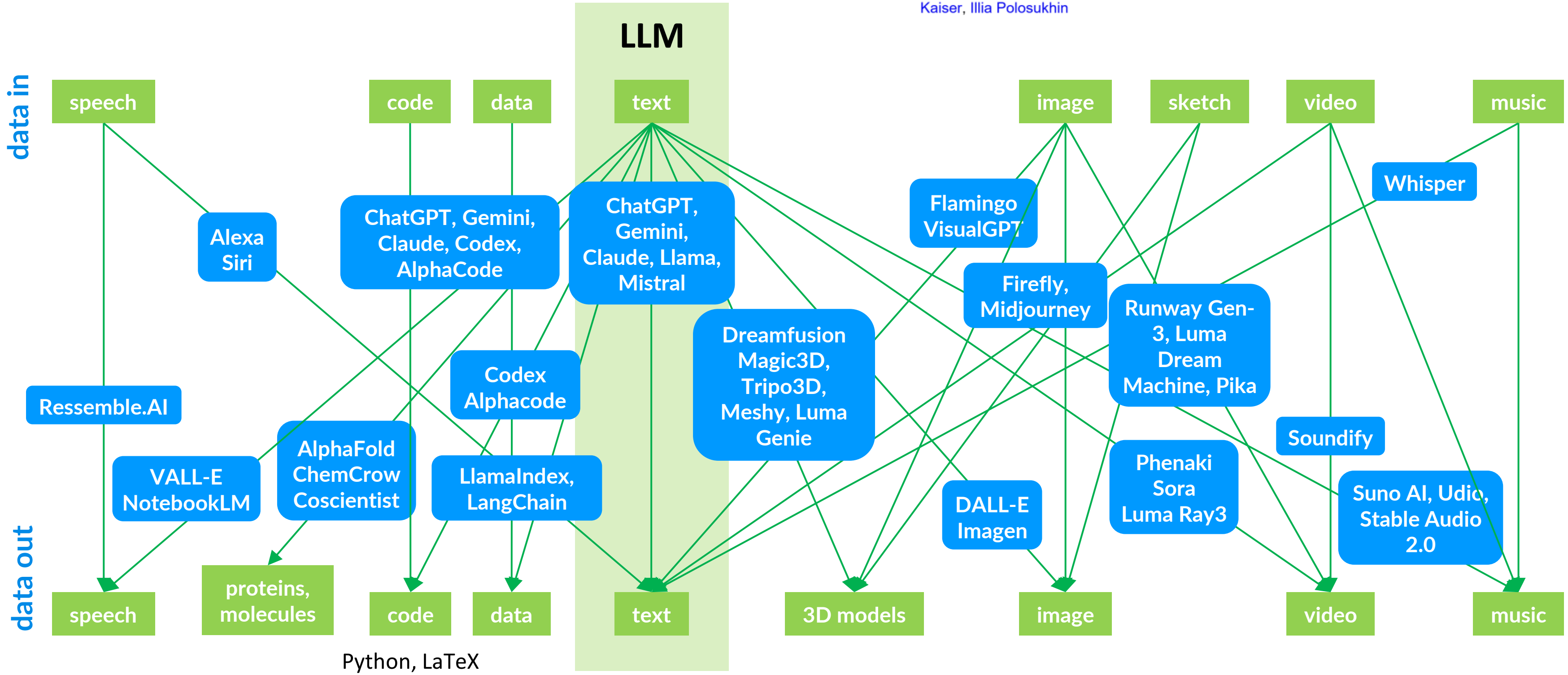
Help | Adv

Computer Science > Computation and Language

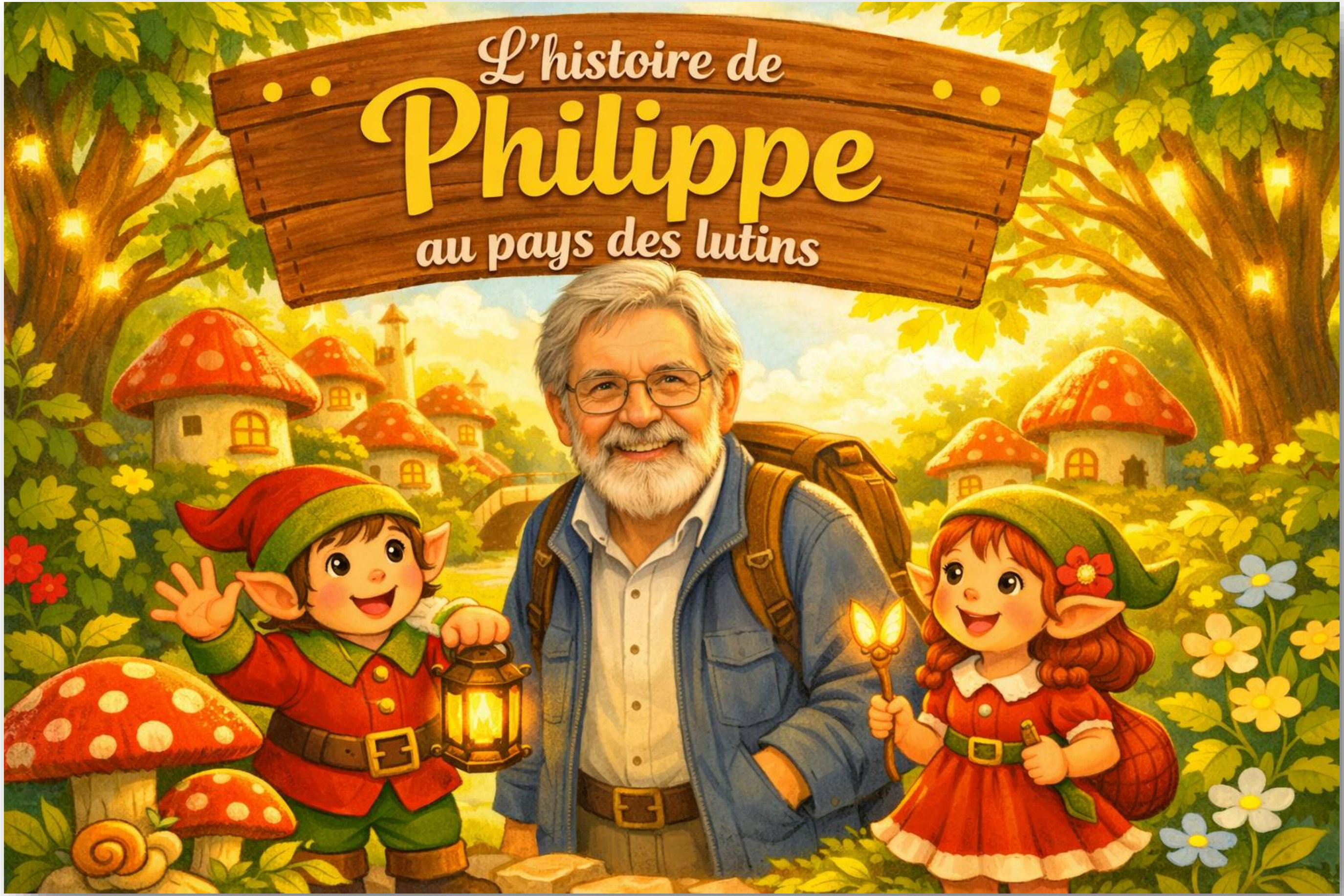
[Submitted on 12 Jun 2017 (v1), last revised 2 Aug 2023 (this version, v7)]

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin



Python, LaTeX
quantum circuits
web site, QEC, QFT


L'histoire de
Philippe
au pays des lutins

vécu pratique avec Philippe

OpenAI
ChatGPT

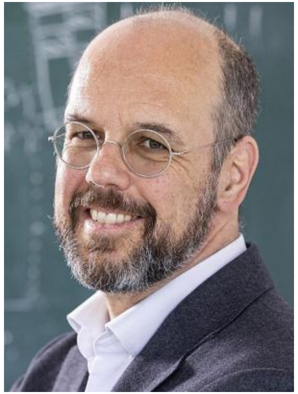
- rubidium 87
- **création de schémas**
- **plan quantique d'un pays européen**
- mémo pour l'UE
- quiz étudiants
- **Shor « à la Trump »**

2023	2024	2025	2026

ChatGPT 4.0 on drafting a national quantum initiative

December 30th, 2023

Tested by Olivier Ezratty.



let's say I am quantum scientist in a small European country and I want to convey to the government the message that it should fund a national quantum initiative to fund research and startup development in this burgeoning field, given it is already competitive and many countries have launched such initiatives. can you write a memo that would persuade my government to implement such a plan for $300M?

ChatGPT

Certainly! Below is a draft of a persuasive memo aimed at convincing your government to fund a national quantum initiative.

To: [Recipient's Name], [Recipient's Title] **From:** [Your Name], Quantum Scientist **Date:** [Date] **Subject:** Proposal for a National Quantum Initiative - A Strategic Investment in Our Nation's Future

Executive Summary:

This memo proposes the establishment of a National Quantum Initiative (NQI) with an investment of $300 million. This initiative is crucial for positioning [Country's Name] as a leader in the rapidly evolving field of quantum science and technology. With many countries already investing heavily in quantum research and development, it is imperative for our national competitiveness and security that we do not fall behind in this critical field.

Background:

Quantum technology, representing the next frontier of scientific advancement, promises revolutionary changes in computing, communication, and sensing. It holds the potential to solve complex problems beyond the reach of classical computers, ensuring advancements in material science, pharmaceuticals, and encryption. Globally, nations like the United States, China, and members of the European Union have already recognized its potential, investing billions into quantum research and startups.

The Need for a National Quantum Initiative:

1. **Competitiveness in Global Science and Technology**: As quantum technology rapidly evolves, not participating at this juncture means falling behind in a critical area of future technology and innovation.
2. **Economic Opportunities**: The global quantum computing market is projected to reach [Projected Value] by 2030. An early investment in quantum technologies will position [Country's Name] to capture a significant share of this market, fostering economic growth and job creation.
3. **National Security**: Quantum computing poses unique challenges and opportunities in the realm of cybersecurity. Developing domestic quantum capabilities is vital for safeguarding our national security interests.
4. **Support for Academia and Startups**: A national initiative will provide much-needed funding for academic research and support the growth of quantum startups, driving innovation and technological breakthroughs.

Proposed Allocation of $300 Million Investment:

1. **Research and Development**: $150 million for advancing quantum computing, communication, and sensing technologies.
2. **Infrastructure Development**: $50 million for establishing state-of-the-art quantum laboratories and research facilities.
3. **Education and Workforce Training**: $30 million for developing educational programs and training the next generation of quantum scientists and engineers.
4. **Startup and Industry Support**: $40 million for grants and incentives to encourage quantum technology startups and industry collaborations.
5. **International Collaboration and Standards**: $30 million for participating in international quantum research projects and contributing to global standards in quantum technology.

Conclusion:

Investing in a National Quantum Initiative is not just about keeping pace with global trends; it is about seizing an opportunity to lead in a

Could explain how the Shor integer algorithm works using a Shakespeare style?

Could you do the same using the style of Donald J Trump ?

ChatGPT

Certainly! Let's imagine how Donald J. Trump might describe Shor's algorithm, renowned for its ability to factorize large numbers, a key challenge in cryptography, in his characteristic style:

"Okay, folks, listen up, I've got something tremendous to tell you about. It's called Shor's algorithm – and believe me, it's huge. This thing, it's a quantum algorithm, not your ordinary, everyday algorithm, but a quantum one. It's the best, really fantastic, for breaking codes and solving big, big numbers. We're talking about numbers so big, you wouldn't believe it.

Now, let me tell you, nobody knows algorithms better than me. And this Shor, he's a smart guy, really smart. He figured out how to use these tiny, tiny things called qubits. Qubits, unlike regular bits, they can be in multiple states at once – it's like, they can be in New York and Florida at the same time. Can you believe it?

schéma expliquant l'algorithme de Peter Shor…

ou pas !

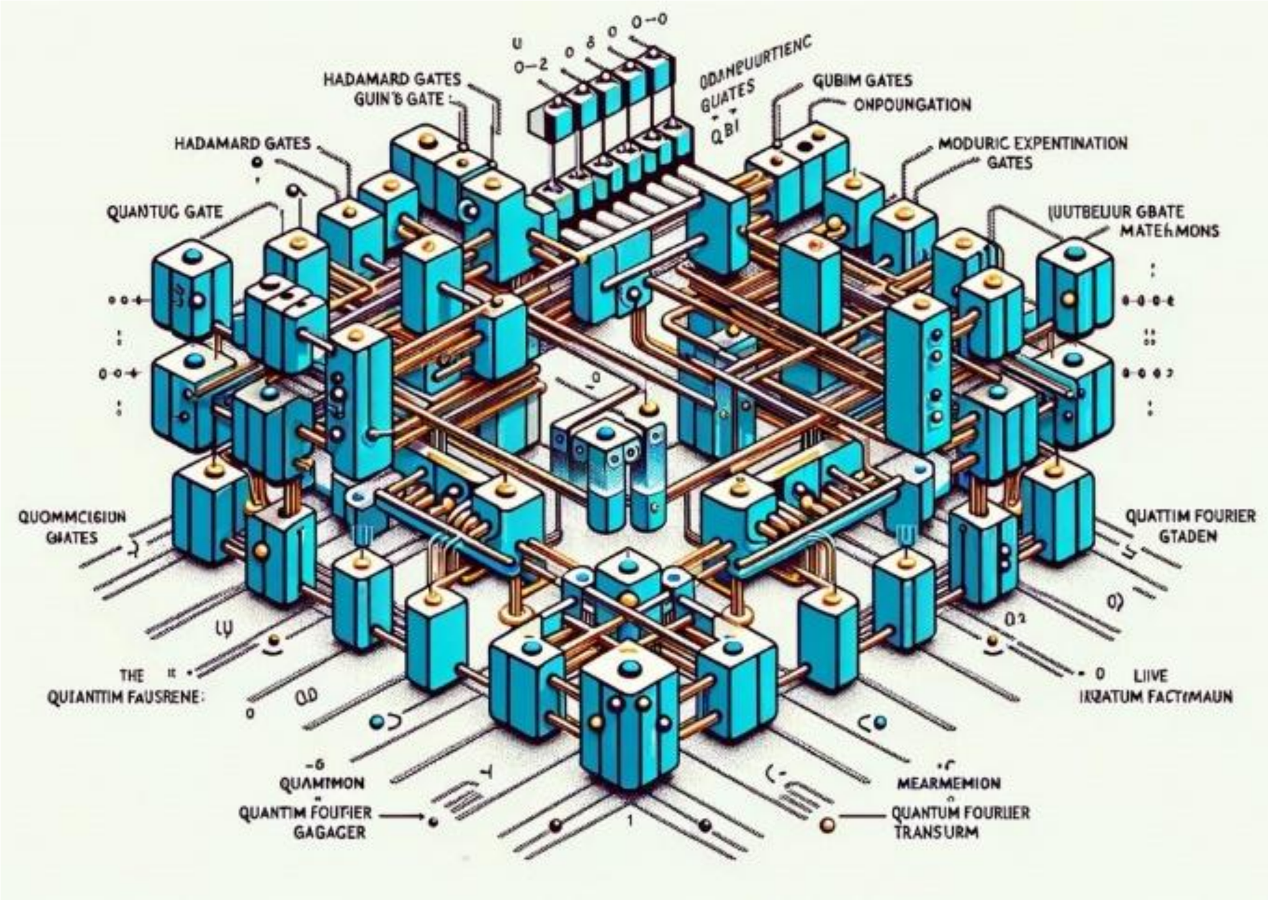


image générée sous ChatGPT avec DALL-E, fin 2023

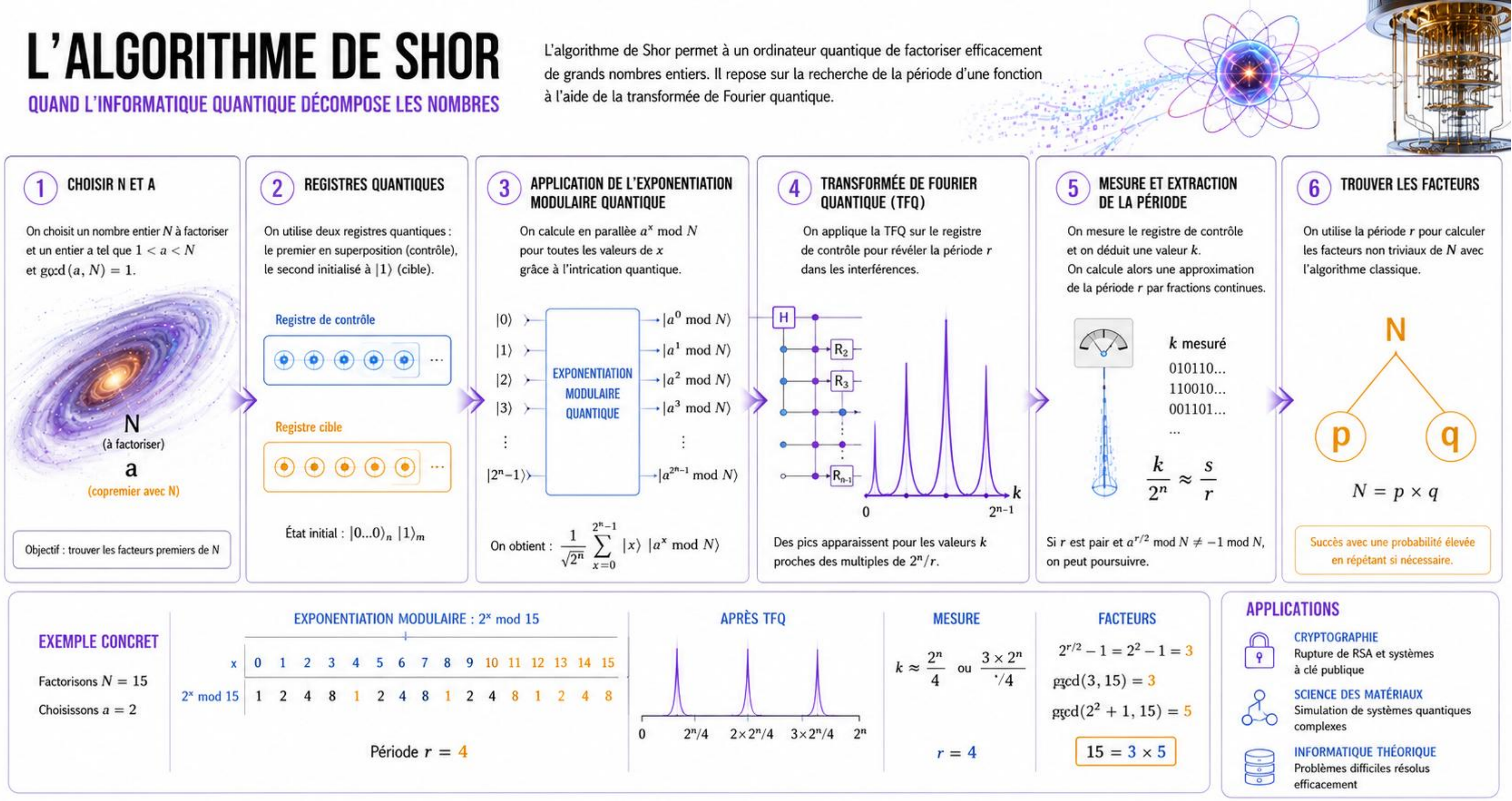


image générée par Vincent Pinte-Deregnaucourt sous OpenAI Image2, Mai 2026

vécu pratique avec Philippe

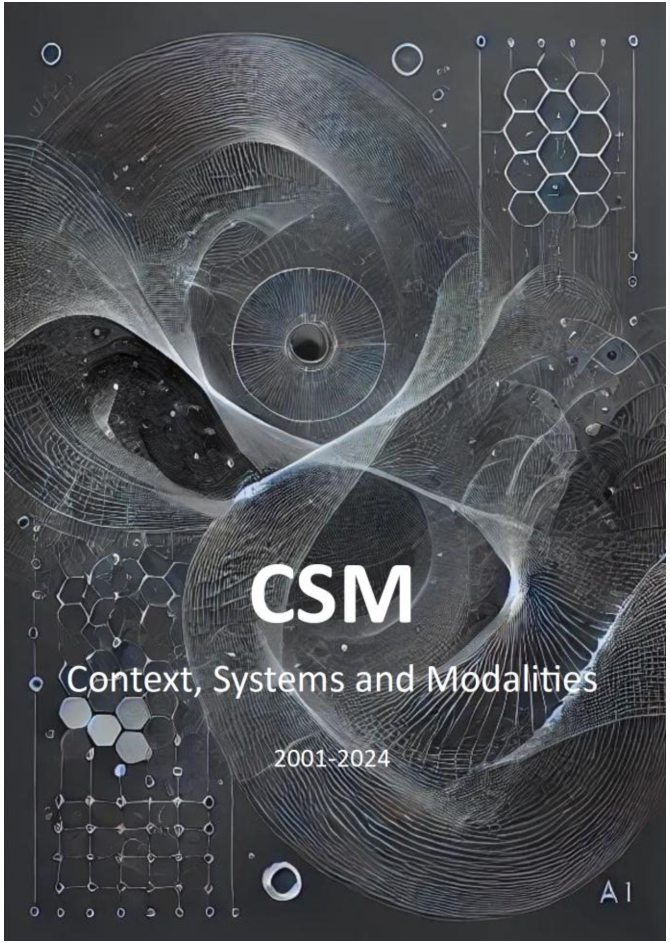
OpenAI ChatGPT

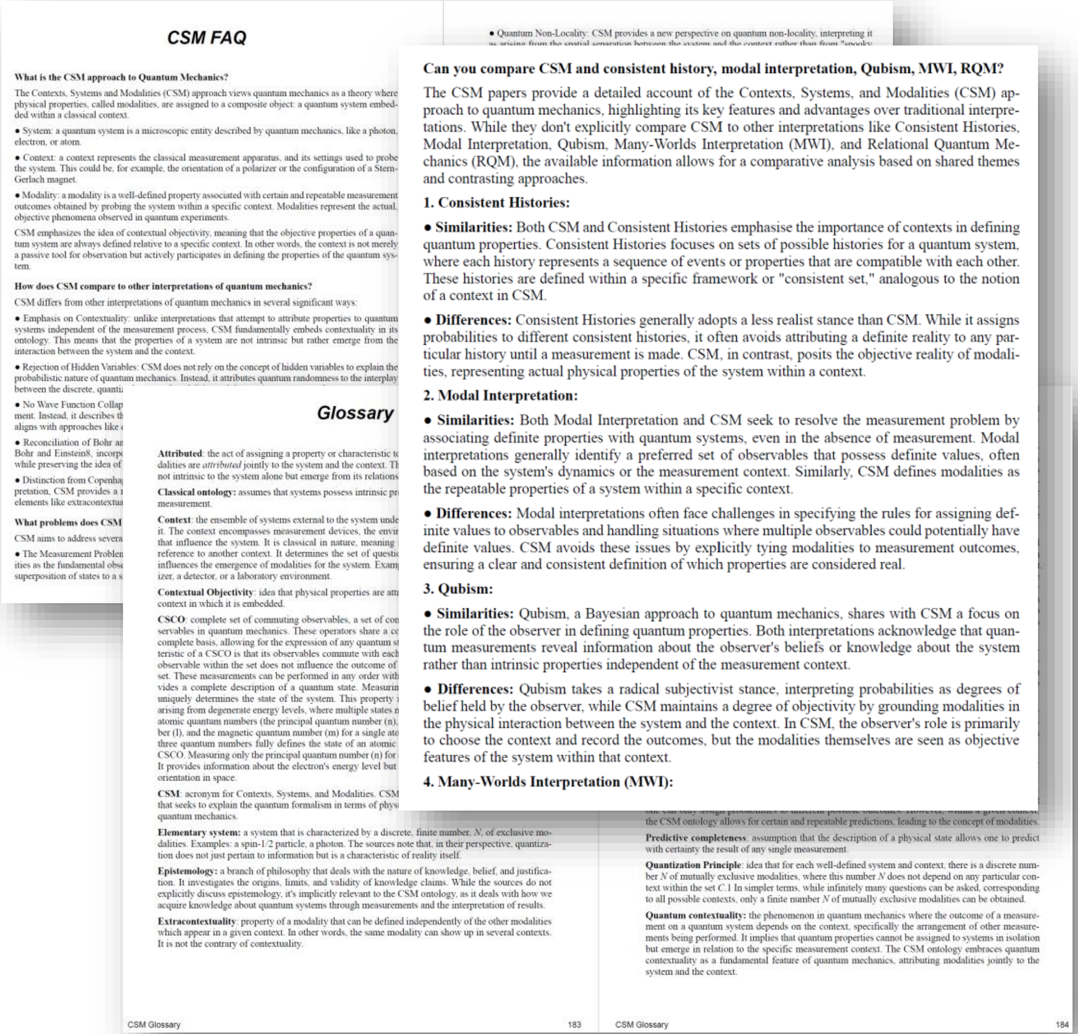
NotebookLM

- rubidium 87
- création de schémas
- plan quantique d'un pays
- mémo pour l'UE
- quiz étudiants
- Shor « à la Trump »

- podcast automatique
- **livre CSM**
- **CSM vs autres interprétations de la MQ!**
- **analyse de charts de papiers scientifiques**
- code python

2023	2024	2025	2026
licence			





CSM FAQ

What is the CSM approach to Quantum Mechanics?

The Contexts, Systems and Modalities (CSM) approach views quantum mechanics as a theory where physical properties, called modalities, are assigned to a composite object: a quantum system embedded within a classical context.

- System: a quantum system is a microscopic entity described by quantum mechanics, like a photon, electron, or atom.
- Context: a context represents the classical measurement apparatus, and its settings used to probe the system. This could be, for example, the orientation of a polarizer or the configuration of a Stern-Gerlach magnet.
- Modality: a modality is a well-defined property associated with certain and repeatable measurement outcomes obtained by probing the system within a specific context. Modalities represent the actual, objective phenomena observed in quantum experiments.

CSM emphasizes the idea of contextual objectivity, meaning that the objective properties of a quantum system are always defined relative to a specific context. In other words, the context is not merely a passive tool for observation but actively participates in defining the properties of the quantum system.

How does CSM compare to other interpretations of quantum mechanics?

CSM differs from other interpretations of quantum mechanics in several significant ways:

- Emphasis on Contextuality: unlike interpretations that attempt to attribute properties to quantum systems independent of the measurement process, CSM fundamentally embeds contextuality in its ontology. This means that the properties of a system are not intrinsic but rather emerge from the interaction between the system and the context.
- Rejection of Hidden Variables: CSM does not rely on the concept of hidden variables to explain the probabilistic nature of quantum mechanics. Instead, it attributes quantum randomness to the interplay between the discrete, quanti…
- No Wave Function Collap… ment. Instead, it describes th… aligns with approaches like …
- Reconciliation of Bohr an… Bohr and Einstein8, incorpo… while preserving the idea of…
- Distinction from Copenhag… pretation, CSM provides a … elements like extracontextua…

What problems does CSM…

CSM aims to address severa…

- The Measurement Problem… ities as the fundamental obse… superposition of states to a s…

- Quantum Non-Locality: CSM provides a new perspective on quantum non-locality, interpreting it as arising from the spatial separation between the system and the context rather than from "spooky…

Glossary

Attributed: the act of assigning a property or characteristic t… dalities are *attributed* jointly to the system and the context. Th… not intrinsic to the system alone but emerge from its relations…

Classical ontology: assumes that systems possess intrinsic pr… measurement.

Context: the ensemble of systems external to the system unde… it. The context encompasses measurement devices, the envir… that influence the system. It is classical in nature, meaning … reference to another context. It determines the set of questi… influences the emergence of modalities for the system. Exam… izer, a detector, or a laboratory environment.

Contextual Objectivity: idea that physical properties are attr… context in which it is embedded.

CSCO: complete set of commuting observables, a set of com… servables in quantum mechanics. These operators share a co… complete basis, allowing for the expression of any quantum st… teristic of a CSCO is that its observables commute with each… observable within the set does not influence the outcome of … set. These measurements can be performed in any order with… vides a complete description of a quantum state. Measurin… uniquely determines the state of the system. This property i… arising from degenerate energy levels, where multiple states n… atomic quantum numbers (the principal quantum number (n), … ber (l), and the magnetic quantum number (m) for a single ato… three quantum numbers fully defines the state of an atomic … CSCO. Measuring only the principal quantum number (n) for … It provides information about the electron's energy level but … orientation in space.

CSM: acronym for Contexts, Systems, and Modalities. CSM … that seeks to explain the quantum formalism in terms of phys… quantum mechanics.

Elementary system: a system that is characterized by a discrete, finite number, N, of exclusive modalities. Examples: a spin-1/2 particle, a photon. The sources note that, in their perspective, quantization does not just pertain to information but is a characteristic of reality itself.

Epistemology: a branch of philosophy that deals with the nature of knowledge, belief, and justification. It investigates the origins, limits, and validity of knowledge claims. While the sources do not explicitly discuss epistemology, it's implicitly relevant to the CSM ontology, as it deals with how we acquire knowledge about quantum systems through measurements and the interpretation of results.

Extracontextuality: property of a modality that can be defined independently of the other modalities which appear in a given context. In other words, the same modality can show up in several contexts. It is not the contrary of contextuality.

CSM Glossary 183

… the CSM ontology allows for certain and repeatable predictions, leading to the concept of modalities.

Predictive completeness: assumption that the description of a physical state allows one to predict with certainty the result of any single measurement.

Quantization Principle: idea that for each well-defined system and context, there is a discrete number N of mutually exclusive modalities, where this number N does not depend on any particular context within the set C.1 In simpler terms, while infinitely many questions can be asked, corresponding to all possible contexts, only a finite number N of mutually exclusive modalities can be obtained.

Quantum contextuality: the phenomenon in quantum mechanics where the outcome of a measurement on a quantum system depends on the context, specifically the arrangement of other measurements being performed. It implies that quantum properties cannot be assigned to systems in isolation but emerge in relation to the specific measurement context. The CSM ontology embraces quantum contextuality as a fundamental feature of quantum mechanics, attributing modalities jointly to the system and the context.

CSM Glossary 184

Can you compare CSM and consistent history, modal interpretation, Qubism, MWI, RQM?

The CSM papers provide a detailed account of the Contexts, Systems, and Modalities (CSM) approach to quantum mechanics, highlighting its key features and advantages over traditional interpretations. While they don't explicitly compare CSM to other interpretations like Consistent Histories, Modal Interpretation, Qubism, Many-Worlds Interpretation (MWI), and Relational Quantum Mechanics (RQM), the available information allows for a comparative analysis based on shared themes and contrasting approaches.

1. Consistent Histories:

- **Similarities:** Both CSM and Consistent Histories emphasise the importance of contexts in defining quantum properties. Consistent Histories focuses on sets of possible histories for a quantum system, where each history represents a sequence of events or properties that are compatible with each other. These histories are defined within a specific framework or "consistent set," analogous to the notion of a context in CSM.
- **Differences:** Consistent Histories generally adopts a less realist stance than CSM. While it assigns probabilities to different consistent histories, it often avoids attributing a definite reality to any particular history until a measurement is made. CSM, in contrast, posits the objective reality of modalities, representing actual physical properties of the system within a context.

2. Modal Interpretation:

- **Similarities:** Both Modal Interpretation and CSM seek to resolve the measurement problem by associating definite properties with quantum systems, even in the absence of measurement. Modal interpretations generally identify a preferred set of observables that possess definite values, often based on the system's dynamics or the measurement context. Similarly, CSM defines modalities as the repeatable properties of a system within a specific context.
- **Differences:** Modal interpretations often face challenges in specifying the rules for assigning definite values to observables and handling situations where multiple observables could potentially have definite values. CSM avoids these issues by explicitly tying modalities to measurement outcomes, ensuring a clear and consistent definition of which properties are considered real.

3. Qubism:

- **Similarities:** Qubism, a Bayesian approach to quantum mechanics, shares with CSM a focus on the role of the observer in defining quantum properties. Both interpretations acknowledge that quantum measurements reveal information about the observer's beliefs or knowledge about the system rather than intrinsic properties independent of the measurement context.
- **Differences:** Qubism takes a radical subjectivist stance, interpreting probabilities as degrees of belief held by the observer, while CSM maintains a degree of objectivity by grounding modalities in the physical interaction between the system and the context. In CSM, the observer's role is primarily to choose the context and record the outcomes, but the modalities themselves are seen as objective features of the system within that context.

4. Many-Worlds Interpretation (MWI):

Name	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	$\omega(G)$	Name	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	$\omega(G)$
g136x6	60	1716	48	2dc.128-c	128	2955	5
g136x8	80	3087	64*	1tc.64-c	64	1824	20*
g136x10	100	4827	60*	1tc.128-c	128	7616	38*
g136x12	120	6970	96*	1et.64-c	64	1752	18*
g150x6	60	1735	54*	1et.128-c	128	7456	28*
g50634x3	60	1711	30*	s3m25x5	125	6892	20*
g50634x4	80	3066	40*	C125.9	125	6963	34*
g50634x5	100	4823	50*	hamming6-2	64	1824	32*
g50634x6	120	6966	60*	hamming6-4	64	704	4*
1dc.64-c	64	1473	10*	johnson8-4-4	70	1855	14*
1dc.128-c	128	6657	16*	johnson16-2-4	120	5460	8*

Table 5: Benchmarks for quantum algorithms (optimal values are marked with '*').

https://arxiv.org/abs/2403.09742
"A Short Review on Novel Approaches for Maximum Clique Problem: from Classical algorithms to Graph Neural Networks and Quantum algorithms" by Raffaele Marino et al, arXiv, March 2024 (24 pages).

NotebookLM

Can you create a detailed caption for table 5?

Caption for Table 5: Quantum Algorithm Benchmarks

Table 5 presents a collection of benchmark graphs specifically chosen for evaluating **quantum algorithms** designed to solve the Maximum Clique Problem (MCP). This curated set expands upon previously established benchmarks by incorporating instances from diverse sources, including:

- **Coding theory:** These graphs, derived from problems in coding theory, are known for their computational challenge 1. They represent real-world applications where finding maximum cliques is crucial, for example in designing error-correcting codes 2 3.
- **EVIL benchmark:** This benchmark is designed to provide a range of difficulty levels, pushing the limits of quantum algorithms 4. It features graphs where the chromatic number (minimum colours for legal colouring) is intentionally far from the clique number, creating a challenging environment for clique-finding algorithms. 4
- **DIMACS benchmark:** A classic set of graphs commonly used for evaluating MCP algorithms 5. Though many DIMACS instances are now easily solvable by classical algorithms, the inclusion of smaller instances here offers a starting point for nascent quantum approaches.

The table includes the graph's **name**, the number of **vertices (|V|)**, the number of **edges (|E|)**, and the **known optimal clique size (ω(G))**, where available. An asterisk (*) denotes instances with proven optimal solutions.

The purpose of this curated benchmark is twofold:

1. **Reflect current quantum hardware capabilities:** The instances selected are manageable for current quantum computers, acknowledging their limitations in terms of qubit count and connectivity.
2. **Drive quantum algorithm development:** By providing a challenging yet accessible set of problems, the benchmark aims to stimulate the creation of more advanced and efficient quantum algorithms for the MCP.

It is important to note that this benchmark is expected to evolve as quantum computing technology advances. Larger and more complex graphs will be incorporated as quantum computers grow in size and capability 6.

test en octobre 2024

l'émergence du **slop**

slop ≈ diarrhée
bouillie
lourdeurs

Before delving into the complexities of the Quantum Phase Estimation (QPE) algorithm and its myriad applications, it's paramount to embark on a journey through the annals of time, exploring the intricate tapestry of its origins. In this introductory discourse, we shall navigate the historical landscape that led to the inception of this groundbreaking algorithm.

At the heart of our narrative lies the indelible imprint of Alexei Kitaev, an icon in the saga of quantum computing. In the fertile intellectual landscape of 1995, Kitaev unfurled the QPE algorithm, marking a watershed moment in the realm of quantum computation. Yet, the lineage of QPE extends far beyond this seminal unveiling, intertwining with the pioneering endeavors of luminaries such as Peter Shor and David Deutsch.

student's work, unnamed, April 2024

Antislop: A Comprehensive Framework for Identifying and Eliminating Repetitive Patterns in Language Models

Samuel J Paech
spaech@gmail.com
Independent

Allen G Roush
allen.roush@thoughtworks.com
Thoughtworks

Judah Goldfeder
jag2396@columbia.edu
Columbia University

Ravid Shwartz–Ziv
ravidziv@gmail.com
New York University

October 23, 2025

Abstract

Widespread LLM adoption has introduced characteristic repetitive phraseology, termed "slop," which degrades output quality and makes AI-generated text immediately recognizable. We present Antislop, a comprehensive framework providing tools to both detect and eliminate these overused patterns. Our approach combines three innovations: (1) **The Antislop Sampler**, which uses backtracking to suppress unwanted strings at inference time without destroying vocabulary; (2) **An automated pipeline** that profiles model-specific slop against human baselines and generates training data; (3) **Final Token Preference Optimization (FTPO)**, a novel fine-tuning method that operates on individual tokens, surgically adjusting logits wherever a banned pattern has appeared in an inference trace. We demonstrate that some slop patterns appear over 1,000× more frequently in LLM output than human text. The Antislop Sampler successfully suppresses 8,000+ patterns while maintaining quality, whereas token banning becomes unusable at just 2,000. Most importantly, FTPO achieves 90% slop reduction while maintaining or improving performance in cross-domain evals including GSM8K, MMLU, and creative writing tasks. In contrast, DPO suffers significant degradation in writing quality and lexical diversity despite achieving weaker suppression. We release all code and results under MIT license: https://github.com/sam-paech/auto-antislop.

1 Introduction

Language models have ushered in an era of slop: Repetitive words and phrases that are instantly recognizable as AI generated text[Wu et al., 2025]. In creative writing, the ubiquitous *Elara* always speaks with "voice barely above a whisper". In functional writing, phrases like *"it's not just X, it's Y"* appear far more frequently than in human text. In our tests, we find that some patterns occur over 1000× more frequently in LLM text than in human writing, leading to the perception of repetition and over-use – i.e. *"slop"*.

Existing approaches to suppress unwanted patterns are brittle or ineffective. Token banning creates collateral damage– for instance, if we wish to ban "catatonic" and it tokenizes to ["cat",

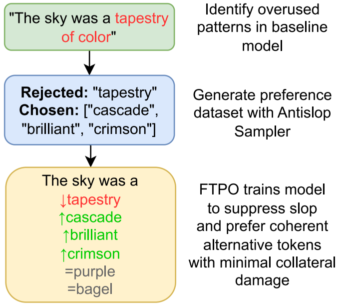


Figure 1: Pipeline for identifying and suppressing overused writing patterns in a language model.

https://arxiv.org/abs/2510.15061

Is ChatGPT Transforming Academics' Writing Style?

Mingmeng Geng [1] **Roberto Trotta** [1 2]

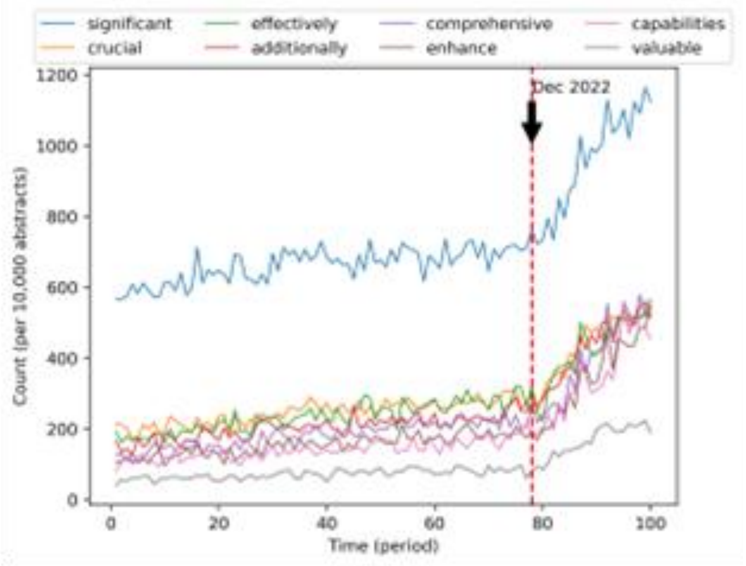


Figure 2. Examples of words with rapidly growing frequency in arXiv abstracts.

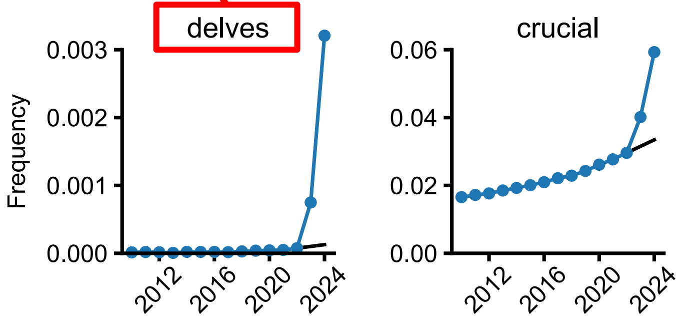
https://arxiv.org/abs/2404.08627, April 2024

DELVING INTO CHATGPT USAGE IN ACADEMIC WRITING THROUGH EXCESS VOCABULARY

Dmitry Kobak [1,2], Rita González-Márquez [1,2,*], Emőke-Ágnes Horvát [3,*], and Jan Lause [1,2,*]

[1] *Hertie Institute for AI in Brain Health, University of Tübingen, Germany*
[2] *Tübingen AI Center, Tübingen, Germany*
[3] *Northwestern University, Evanston, Illinois, USA*
[*] *Alphabetic order*
✉ dmitry.kobak@uni-tuebingen.de

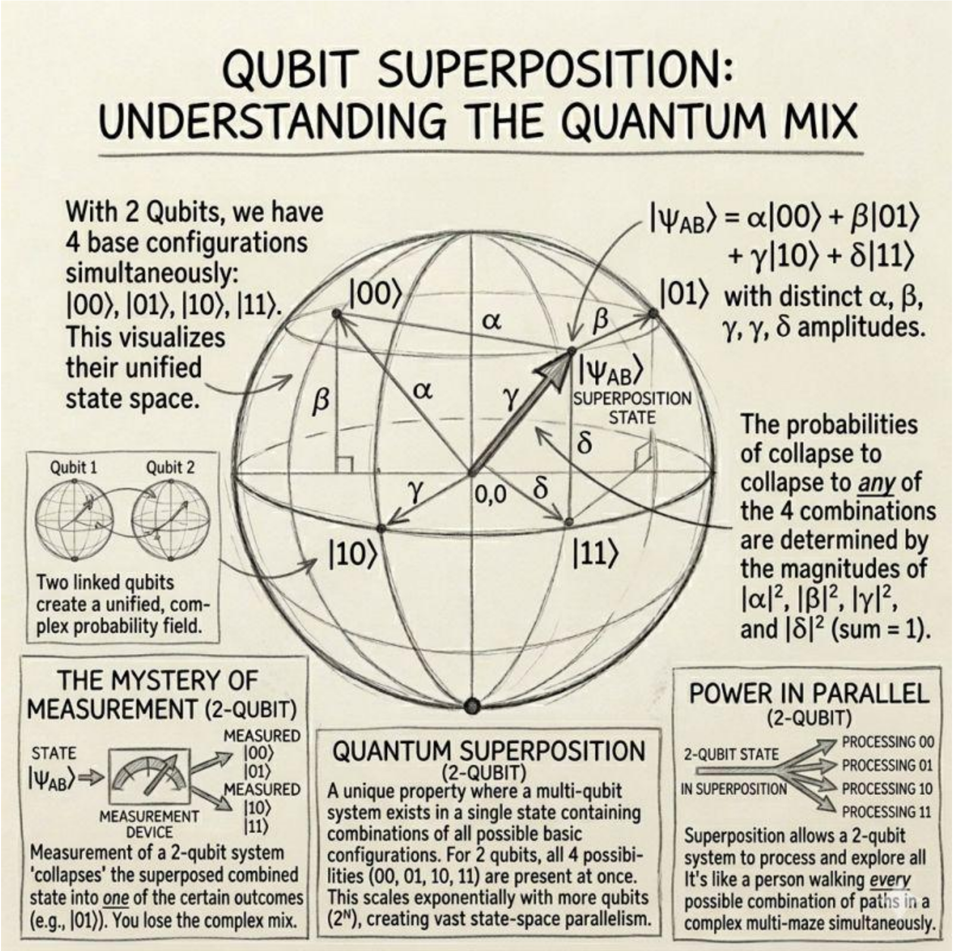
June 12, 2024



https://arxiv.org/abs/2406.07016v1, June 2024 / July 2025

slop visuel

lien LinkedIn



+1k des utilisateurs utilisent le Détecteur d'IA en ce moment

4.5 ★ Trustpilot

Excellent / 1773 avis

Détecteur d'IA: Détection d'IA intelligente

TextGuard est un outil éducatif innovant conçu pour aider les utilisateurs à vérifier l'originalité des textes et à améliorer leur singularité grâce à la reformulation, sans en altérer le sens.

Votre texte est humanisé !

- Obtenez un texte 100 % unique avec 0 % de présence d'IA
- Conservez le sens original
- Choisissez le niveau de lisibilité souhaité

Déverrouillez votre texte

0%

- 0% Texte généré par l'IA
- 100% Texte rédigé par un humain

Double vérification par :

GPTZero

OpenAI

QuillBot

Scribbr

JJ Jessica Johnson

Feb 1, 2026

Aucune IA n'a été détectée, j'adore cette appli. Elle m'a aidé à obtenir mon diplôme universitaire et à peaufiner mon style d'écriture.

conséquence indirecte

visual slop from CoPilot, May 28th, 2026

Empirical evidence of Large Language Model's influence on human spoken communication

Hiromu Yakura[*†a], **Ezequiel Lopez-Lopez**[†b], **Levin Brinkmann**[†a], **Ignacio Serna**[a], **Prateek Gupta**[a], **Ivan Soraperra**[a], and **Iyad Rahwan**[a]

[a] *Center for Humans and Machines, Max-Planck Institute for Human Development, Germany*
[a] *Center for Adaptive Rationality, Max-Planck Institute for Human Development, Germany*

https://arxiv.org/abs/2409.01754

et les ~~hallucinations~~ erreurs ?

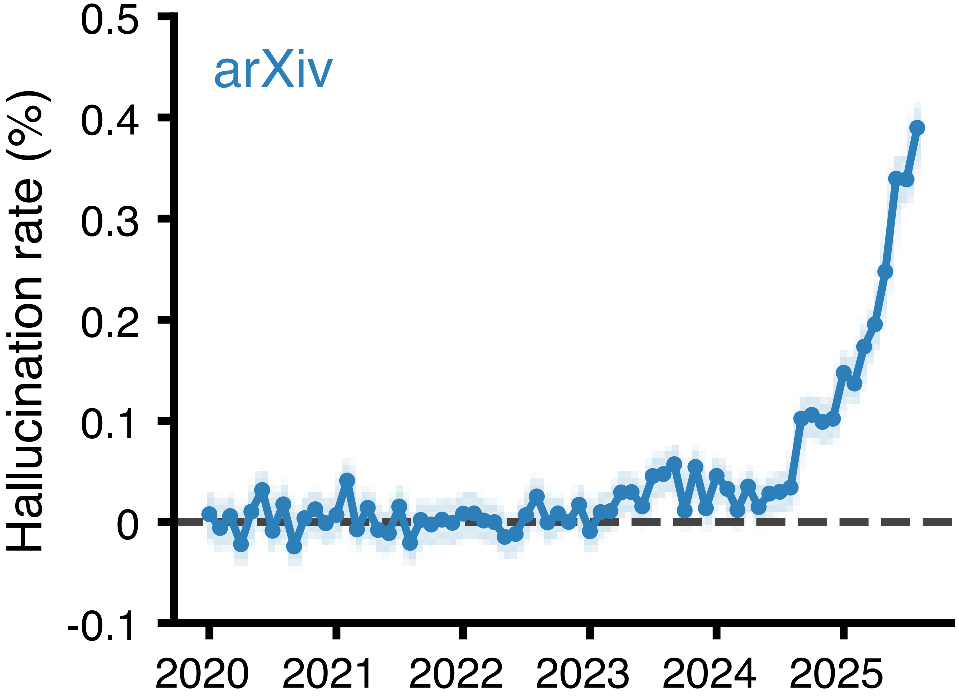
TM Alain Aspect

le LLM invente
des faits inexistants
papiers, théories,
faits, données, …

en trois ans

on n’est pas d’accord
avec le LLM
comme avec un spécialiste, un élève,
un collègue, et il corrige ou pas

références inventées par l'IA



https://arxiv.org/abs/2605.07723

c'est évitable :

- prompt de vérification.
- vérification manuelle.
- vérification programmée.
- Claude Pro.

The Nobel Prize in Physics 2024

Ill. Niklas Elmehed © Nobel Prize Outreach

John J. Hopfield

Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey E. Hinton

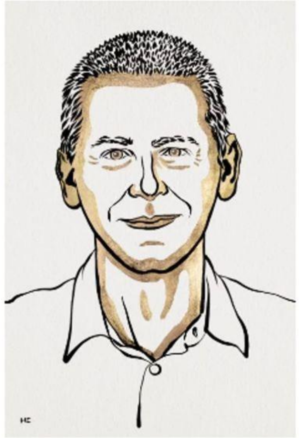
The Nobel Prize in Chemistry 2024

Ill. Niklas Elmehed © Nobel Prize Outreach

David Baker

Ill. Niklas Elmehed © Nobel Prize Outreach

Demis Hassabis



Ill. Niklas Elmehed © Nobel Prize Outreach

John M. Jumper

Accurate structure prediction of biomolecular interactions with AlphaFold 3
by Josh Abramson, Demis Hassabis, John M. Jumper et al, Google Deepmind,
Nature, December 2023-May 2024 (24 pages).

vécu pratique avec Philippe

OpenAI ChatGPT

- rubidium 87
- création de schémas
- plan quantique d'un pays
- mémo pour l'UE
- quiz étudiants
- Shor « à la Trump »

NotebookLM

- livre CSM
- podcast automatique
- analyse de charts de papiers scientifiques
- code python
- CSM vs autres interprétations de la MQ!

Copilot

Gemini

- test de conjectures
- papier ITP en LaTeX
- premiers raisonnements
- accès contenu web
- **site web CSM**
- *creuser Lieb-Robinson*
- *UQT en LaTeX*

2023	2024	2025	2026

licence

master

CSM Ontology Definitions Simulation Mathematics Contributors

Key Contributors

The CSM ontology is the result of years of interdisciplinary collaboration between physicists, mathematicians, and philosophers.

Philippe Grangier
Foundational Physics

Main architect of Contextual Objectivity. A pioneer in quantum optics experiments (Aspect, Grangier, Roger 1982).

Alexia Auffèves
Quantum Foundations & Energy

Crucial in defining the CSM axioms and linking them to quantum thermodynamics and thermodynamics of information.

Matthias van den Bossche
Mathematical Physics

Led the integration of von Neumann Type III algebras to formalize the mathematical transition between System and Context.

Nayla Farouki
Philosophy of Science

Collaborated on the epistemological implications of CSM, focusing on the distinction between Classical and Quantum Realism.

Olivier Ezratty
Communication & Tech Strategy

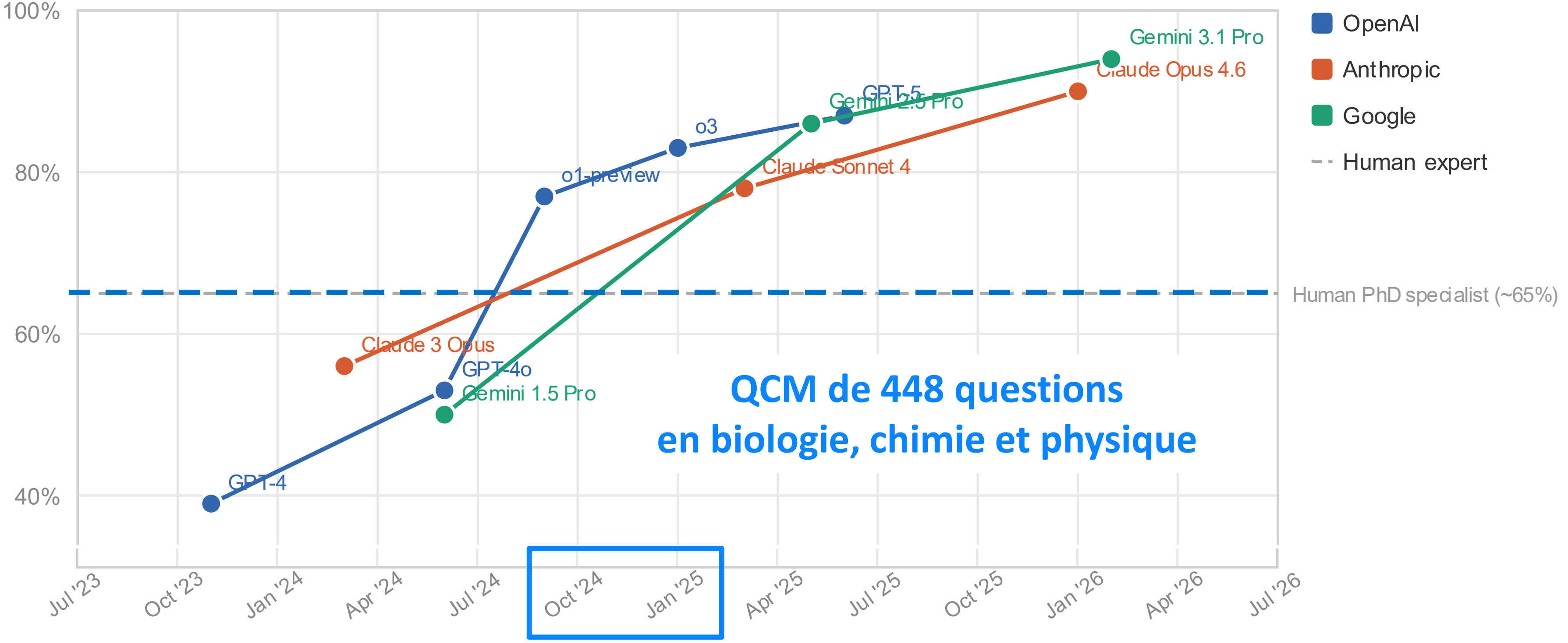
Active contributor in the dissemination and pedagogical structuring of CSM to the broader quantum computing community.

https://gemini.google.com/share/cfe7ea7475ef
https://gemini.google.com/share/80ebb6bb9205
https://gemini.google.com/share/d011fb4416d5

évolutions des LLM sur des QCM

GPQA-Diamond — LLM performance over time by model

arXiv:2311.12022 · 198 PhD-level questions: biology, chemistry, physics



Scores are approximate. Sources: arXiv papers, official model cards, and benchmark leaderboards.

https://arxiv.org/abs/2311.12022

“les lutins”

™ Philippe Grangier

Start research

Great. While I'm researching, feel free to leave this chat. I'll let you know as soon as I'm done.

Quantum Computer Phase Transition Risks
Researching 32 websites...

What do you want to research?

Deep Research ×

Thinking

Quantum Computer Phase Transitions

Upgra

Quantum Computer Phase Transition Risks | Show thinking

Quantum Scaling and Phase Dynamics

I am investigating how scaling quantum systems to millions of qubits moves them toward the thermodynamic limit, where collective physical phenomena emerge. I have identified a critical tension between Many-Body Localization, which helps preserve information, and thermalization, where information is lost to the environment. My initial synthesis suggests that 'undesirable' transitions may manifest as a sudden shift from a controllable quantum state into a chaotic thermal phase as the system size grows.

Error Correction as a Phase Boundary

I am analyzing the threshold theorem through the lens of phase transitions, specifically how logical information exists in an 'ordered' phase that can collapse into a 'disordered' state if noise levels cross a critical point. A major information gap I aim to resolve is whether very large-scale architectures introduce new, non-correctable phases or decoding complexities that do not appear in smaller, NISQ-era devices.

Investigating Hardware-Specific Risks

To move forward, I will evaluate how specific hardware platforms—such as superconducting circuits or neutral atom arrays—might suffer from spontaneous symmetry breaking or noise-induced transitions. I plan to search for research regarding macroscopic state collapse and collective noise effects to determine if there is a fundamental limit where quantum behavior is inevitably washed out by the sheer scale of the processor.

Researching websites

youtu... What Challenges ...	mirage... How To Make Qu...	planck... A Universal Globa...	journal... Quantum Comput...
youtu... When Many Body...	resear... Many-body locali...	pmc.n... Many-body locali...	collab... Many-body locali...
physic... What is special a...	en.wiki... Quantum decohe...	arxiv.o... Dissipative quant...	quant... Dissipative phase...
forum.ic... Noise in Quant...	mewb... Noisy quantum c...	milesg... Fault-Tolerant Qu...	quantu... Threshold theor...
oxford... Quantum Error C...	acade... Vulnerability of fa...	iontra... Quantum error co...	arxiv.o... Experimental de...
pmc.n... Quantum Phase T...	physic... Spontaneous sy...	arxiv.o... Experimental obs...	philsci... Spontaneous Sy...
nation... Chapter: Appendi...	arxiv.o... How to Build a Q...	mdpi.c... Quantum Comput...	plainc... Quantum Comput...
sandia... Current Progress ...	frontie... Quantum comput...	arxiv.o... Quantum Comput...	hormo... Quantum collaps...

Quantum Thermalization and Information Loss

I am exploring how the interplay between Many-Body Localization (MBL) and thermalization creates a precarious balance for large-scale processors. My research indicates that while MBL can theoretically 'lock' quantum information and prevent it from decaying into the environment, this stability is highly fragile in the two- or three-dimensional arrays required for millions of qubits. I have found evidence that 'rare regions' of low disorder can act as thermal seeds, potentially triggering a phase transition that causes the entire system to thermalize and lose its computational state.

Vulnerabilities in Error Correction Phases

I have identified a significant risk where quantum error correction (QEC) itself undergoes an undesirable phase transition. Recent findings suggest that even if physical noise remains below traditional thresholds, 'unidentifiable' measurement errors or coherent noise during state preparation can precipitate a total

vécu pratique avec Philippe

OpenAI ChatGPT

- rubidium 87
- création de schémas
- plan quantique d'un pays
- mémo pour l'UE
- quiz étudiants
- Shor « à la Trump »

NotebookLM

- livre CSM
- podcast automatique
- analyse de charts de papiers scientifiques
- code python
- CSM vs autres interprétations de la MQ!

Copilot
Gemini

- test de conjectures
- papier ITP en LaTeX
- premiers raisonnements
- accès contenu web
- site web CSM
- *creuser Lieb-Robinson*
- *UQT en LaTeX*

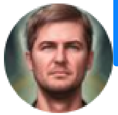
Claude
Gemini

- raisonnements avancés
- **review de papiers**
- **review de reviews de papiers**
- monographies thématiques

2023	2024	2025	2026

licence — master — early-PhD

le peer reviewing à l’ère des LLMs



Victor V. Albert
@victorvalbert

Exciting, scary, and thoughtful session on the future of AI at #apsmarch and possible breakdown of peer review. Matthew D. Schwartz: "AI will keep getting better and better at physics. We will not."

5:11 PM · Mar 18, 2026 · **1,115** Views

UNIVERSITY OF MARYLAND

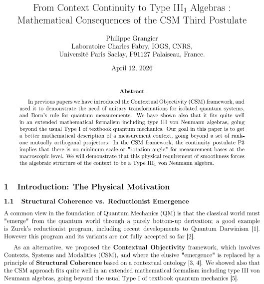
NIST

HARVARD UNIVERSITY

Department of Physics

parcours d’un papier ITP/Heisenberg cut

Gemini



From Context Continuity to Type III_1 Algebras : Mathematical Consequences of the CSM Third Postulate

Philippe Grangier
Laboratoire Charles Fabry, IOGS, CNRS,
Université Paris Saclay, F91127 Palaiseau, France.

April 12, 2026

Abstract

In previous papers we have introduced the Contextual Objectivity (CSM) framework, and used it to demonstrate the need of unitary transformations for isolated quantum systems, and Born’s rule for quantum measurements. We have shown also that it fits quite well in an extended mathematical formalism including type III von Neumann algebras, going beyond the usual Type I of textbook quantum mechanics. Our goal in this paper is to get a better mathematical description of a measurement context, going beyond a set of rank-one mutually orthogonal projectors. In the CSM framework, the continuity postulate P3 implies that there is no minimum scale or "rotation angle" for measurement bases at the macroscopic level. We will demonstrate that this physical requirement of smoothness forces the algebraic structure of the context to be a Type III_1 von Neumann algebra.

1 Introduction: The Physical Motivation

1.1 Structural Coherence vs. Reductionist Emergence

A common view in the foundation of Quantum Mechanics (QM) is that the classical world must "emerge" from the quantum world through a purely bottom-up derivation; a good example is Zurek’s reductionist program, including recent developments to Quantum Darwinism [1]. However this program and its variants are not fully accepted so far [2].

As an alternative, we proposed the **Contextual Objectivity** framework, which involves Contexts, Systems and Modalities (CSM), and where the elusive "emergence" is replaced by a principle of **Structural Coherence** based on a contextual ontology [3, 4]. We showed also that the CSM approach fits quite well in an extended mathematical formalism including type III von Neumann algebras, going beyond the usual Type I of textbook quantum mechanics [5].

Philippe écrit un papier sur les ITP en se faisant aider par Gemini

Philippe demande un éclairage à des mathématiciens

qui ne répondent pas!

Claude

Philippe fait une review du papier avec Antropic Claude

Franck suggère à Olivier de faire une review d’acceptation et une autre de rejet du papier

Copilot

Olivier fait plusieurs reviews du papier avec CoPilot (ChatGPT)

Claude

Philippe fait une review des reviews de CoPilot avec Claude

Claude

Philippe complète le papier, reboucle avec Matthias, et …

Olivier fait une review du papier avec Claude, CoPilot et Gemini et demande à Claude d’évaluer les reviews

l’histoire continue

mises à jour, évaluation avec Mistral “CNRS”

critique la review de CoPilot

Claude

A propos non pas de vrais mais de "faux" referees, voici un rapport de Copilot sur mon papier révisé, envoyé à un collègue pour être plus "objectif". Qu'en pensez vous ?

15:30

Voici mon analyse comparative de ce rapport Copilot.

Analyse du rapport Copilot

Impression générale

C'est un rapport substantiel et bien structuré, clairement en mode "amélioration constructive" plutôt que rejet — ce qui confirme votre observation d'hier sur l'expérience de votre collègue. Copilot a produit exactement ce que ferait un referee bienveillant et compétent techniquement, mais avec des caractéristiques révélatrices.

Ce qui est bien dans ce rapport

La liste des points améliorés est précise et honnête. Les huit demandes de révision sont toutes légitimes et correspondent à ce qu'un referee sérieux pourrait demander. La structure "ce qui a été amélioré / ce qui reste à faire" est pédagogiquement efficace.

Ce qui révèle la nature LLM du rapport

1. L'escalade incontrôlée. C'est le point le plus frappant. Quand votre collègue a demandé "fais les trois" puis "rédige les preuves complètes" puis "convertis en appendice formel", Copilot a produit des dizaines de pages de matériel technique de plus en plus élaboré — sans jamais dire "attention, ceci dépasse ce qui est raisonnable pour cet article". Un vrai referee ou collaborateur humain aurait dit à un moment : *"Stop, ce n'est plus de la révision d'article, c'est écrire un autre article."*

2. La fausse rigueur mathématique. Les "preuves" des lemmata 1-4 sont rédigées dans un style mathématique convaincant, avec des numéros d'étapes, des références, des notations formelles — mais si on les lit attentivement, elles sont essentiellement circulaires ou heuristiques habillées en preuves. Par exemple, la "preuve" du Lemme 2 suppose exactement ce qu'elle veut démon-

consolidation des reviews

...

2. Central inconsistency: what does the crossed product actually do?

This is the sharpest disagreement across the five reviews, and it matters technically.

NotebookLM claims: "inside the larger space N, intertwining operators exist that can map the initial entangled state into a specific macroscopic sector." It attributes this to Takesaki duality.

Standard Gemini similarly claims the dual flow θ_s can drive the jump between $E_{\theta',\uparrow}$ and $E_{\theta',\downarrow}$.

Copilot builds an entire four-step program around using the crossed product to "dynamically select" a sector via ergodicity of the dual flow.

My review (Claude) and **Gemini Pro** both push back against this. The reason is concrete: since $\omega_{\theta'} = |\alpha|^2\, \omega_{\theta',\uparrow} + |\beta|^2\, \omega_{\theta',\downarrow}$ is a direct sum state, the modular group $\sigma^{\omega_{\theta'}}$ preserves each summand separately. The crossed product therefore decomposes as:

$$\big(F_{\theta',\uparrow} \oplus F_{\theta',\downarrow}) \rtimes_{\sigma^{\omega}} \mathbb{R} \;\cong\; (F_{\theta',\uparrow} \rtimes_{\sigma^{\omega_\uparrow}} \mathbb{R}) \oplus (F_{\theta',\downarrow} \rtimes_{\sigma^{\omega_\downarrow}} \mathbb{R}\big)$$

Verdict: NotebookLM and standard Gemini are likely wrong on this specific point. The crossed product does not provide an inter-sector transition mechanism without additional external structure (such as the Belavkin-Fock space coupling I proposed).

autocritique du LLM

Claude

3. Potential errors in my own review

Intellectual honesty requires flagging two issues in my response.

3a. The Ge-Kadison citation is self-contradictory and possibly misapplied.

3b. The Belavkin equation I wrote is the diffusive (homodyne) version.

For the physical situation in the paper — fluorescence photon counting by a CCD — the natural stochastic master equation is the *jump* version driven by a Poisson counting process N_t:

$$d\rho_t = \mathcal{L}(\rho_t)\, dt + \left(\frac{M\rho_t M^\dagger}{\mathrm{Tr}(M\rho_t M^\dagger)} - \rho_t\right) dN_t$$

...

4. Gemini's Type III_1 vs. Type III_λ confusion

...

5. Gemini Pro's correct insight, with a gap

...

6. Copilot's "commuting subfactors" error

...

synthèse des reviews des LLMs

Claude

The deepest lesson from comparing all five reviews is that **no LLM correctly handles all three layers simultaneously**: the operator algebra formalism, the physical measurement setup, and the foundational distinction between decoherence (transition to a classical mixture) and actualization (selection of a unique outcome). Each review gets one or two layers right and confuses the third. The authors' paper is actually more careful on this distinction than most of the LLM reviews.

types de raisonnements utilisés

Claude

what category of reasoning did you implement to respond to this question (deductive, inductive, analogy, ...)?

Good meta-question. Let me be precise rather than just listing categories.

The dominant mode: deductive reasoning within a fixed axiomatic framework

…

A significant component: conceptual reframing by analogy

…

What was absent: induction

…

An honest caveat: one abductive step

…

analyse d’erreurs de raisonnement

can you trace the origin of you first error when mixing up contexts and modalities? what form or reasoning was missing?

The surface error was substituting "materialization = $M_{\theta'} \rightarrow F_{\theta',i}$" (transition from context to modality) for the correct "materialization = selection of $F_{\theta',i}$ within $M_{\theta'}$" (Kolmogorovian conditioning within the context). The deductive chain built on top of this was internally consistent — which made the error hard to self-detect. Correct syntax, wrong semantics.

The deeper origin has two layers.

Layer 1: Missing document

...

Layer 2: Missing reasoning step — semantic grounding before deduction

...

Before any algebraic manipulation: verify that each physical CSM concept is correctly mapped to its algebraic representative.

...

The form of reasoning that was missing

...

Why the error was self-concealing

This is a general failure mode: **deductive fluency without prior semantic anchoring**. The more powerful the formal machinery, the easier it is to build convincing but misgrounded arguments within it. The corrective is not more deduction but a prior, explicit, and slow verification of the interpretation map — which is a different cognitive operation, closer to careful reading than to reasoning.

Claude

suggestions de raisonnements

Claude

how could some inductive reasoning help and where to respond to this question. which tool could implement it (besides an human)?

A sharp meta-question. Let me be specific about both where induction could contribute and what tool could realistically implement it.

Where inductive reasoning could help

1. Validating P3 across known physical systems

2. Identifying which Type III_1 properties are actually necessary

3. Connecting to the algebraic QFT literature inductively

4. Generating and testing conjectures about the convergence problem

What tool could implement this:

Tasks 1 and 3: Literature-scale pattern recognition

Task 2: Varying model parameters systematically

Task 4: Conjecture generation over proof patterns

nature

Why AI cannot do good science without humans

Amid the rise of 'AI scientists', it's as well to remember that human wisdom, empathy and sheer messiness are as much part of progress as are process and efficiency.

Does humanity need science? The Nobel prize-winner Max Perutz posed this question in a landmark essay[1] in 1989. His conclusion, unsurprisingly, was 'yes'. Had he lived to see the era of artificial intelligence, he might have inverted his framing of the question: does science need humanity?

Two studies in *Nature* provide a glimpse of what some are interpreting as humanity's shrinking role in scientific discovery in Perutz's own field, molecular biology. Both describe a pivotal step towards truly AI-driven drug discovery, in which a system of connected AI agents is trained to autonomously navigate multi-step workflows. The system trawled scholarly literature, formed hypotheses, interpreted data and engaged in internal debate to arrive at candidate drugs to treat a particular disease.

> **"We don't yet know whether greater efficiency equates to greater insight."**

for all the 'wow' factor, it is crucial to bear in mind that the AI systems were not working alone. In both cases, humans framed the initial project, performed experiments, offered guidance and checked the agents' output along the way. That is a feature, not a bug – and shutting humans out of the loop would not be easy.

For one thing, the creation of a fully autonomous AI researcher would require many more AI agents to be cross-linked, and scientists would need to ensure that those agents could maintain focus and coordinate efficiently with one another during long, complex tasks[4,5]. It is also not yet clear whether AI agents would be able to master every step of the process of drug development. Even if they could, human oversight would still be needed to catch hallucinations (AI data fabrication) and misinterpretations, and to ensure that certain tasks are done correctly – for example, that the data AI harvests from databases for its analyses are used appropriately.

More broadly, the expertise of researchers, and the accumulated wisdom of teams that have worked on a problem, perhaps for decades, might prove impossible for machines to replicate. Humans are also needed to train future generations, and their knowledge is built, in part, on lessons learnt from failures and inefficient meanderings. Human messiness, curiosity and playfulness have fuelled countless discoveries, and helped to inform society's ethical frameworks. AI systems might offer greater efficiency in some instances, but we don't yet know whether greater efficiency equates to greater insight.

https://www.nature.com/articles/d41586-026-01551-3

recherche assistée par l'IA

"AI-Researcher: Autonomous Scientific Innovation"

GITHUB TRENDING
3 #3 Repository Of The Day

PROJECT PAGE | SLACK JOIN US | DISCORD JOIN US

DOCUMENTATION | PAPER ON ARXIV | DATASETS

FEISHU GROUP | WECHAT GROUP

Welcome to **AI-Researcher**🤗 AI-Researcher introduces a revolutionary breakthrough in **Automated Scientific Discovery**🔬, presenting a new system that fundamentally **Reshapes the Traditional Research Paradigm**. This state-of-the-art platform empowers researchers with:

- 🎯 **Full Autonomy**: Complete end-to-end research automation
- 🔄 **Seamless Orchestration**: From concept to publication
- 🧠 **Advanced AI Integration**: Powered by cutting-edge AI agents

Accelerating Scientific Research with Gemini: Case Studies and Common Techniques

David P. Woodruff[*, †, ‡1,2], Vincent Cohen-Addad[†, ‡1], Lalit Jain[‡1], Jieming Mao[‡1], Song Zuo[†, ‡1], MohammadHossein Bateni[†1], Simina Brânzei[†3,1], Michael P. Brenner[†1,5], Lin Chen[†1], Ying Feng[†6], Lance Fortnow[†7], Gang Fu[†1], Ziyi Guan[†13], Zahra Hadizadeh[†10], Mohammad T. Hajiaghayi[†1,14], Mahdi JafariRaviz[†14], Adel Javanmard[†1,4], Karthik C. S.[†8], Ken-ichi Kawarabayashi[†12], Ravi Kumar[†1], Silvio Lattanzi[†1], Euiwoong Lee[†9], Yi Li[†15], Ioannis Panageas[†10], Dimitris Paparas[†1], Benjamin Przybocki[†2], Bernardo Subercaseaux[†2], Ola Svensson[†13], Shayan Taherijam[†10], Xuan Wu[†15], Eylon Yogev [†16], Morteza Zadimoghaddam[†1], Samson Zhou[†11], Yossi Matias[1], James Manyika[1], and Vahab Mirrokni[*, †, ‡1]

[1]Google Research
[2]Carnegie Mellon University
[3]Purdue University
[4]University of Southern California
[5]Harvard University
[6]MIT
[7]Illinois Institute of Technology
[8]Rutgers University
[9]University of Michigan
[10]University of California, Irvine
[11]Texas A&M University
[12]National Institute of Informatics, Tokyo and The University of Tokyo
[13]EPFL
[14]University of Maryland, College Park
[15]Nanyang Technological University
[16]Bar-Ilan University

https://arxiv.org/abs/2602.03837

aide à la **démonstration de théorème**

Anthony Leverrier
@letonyo

The latest paper by Terry Tao acknowledges use of LLMs. Yet many researchers still think they are worthless for research.

s used to generate the figures in the text, Github Copilot was used
, Claude Code was used for typesetting, and ChatGPT Pro was used f
or supplying the proof of Lemma 2.10. Outside of these AI tool uses,
was human generated.

10:31 AM · Apr 5, 2026 · **13.1K** Views

PRODUCTS OF CONSECUTIVE INTEGERS WITH UNUSUAL ANATOMY

TERENCE TAO

Abstract. Call an interval $\{N+1,\dots,N+H\}$ of consecutive natural numbers *bad* if the product $(N+1)\dots(N+H)$ is divisible by the square of its largest prime factor; *very bad* if this product is powerful, and *type* F_3 if it has the same squarefree component as a factorial. Such concepts arose in the analysis of the factorial equation $a_1!a_2!a_3! = m^2$ with $a_1 < a_2 < a_3$. Answering several questions of Erdős and Graham, we obtain asymptotics for the number of integers contained in bad or very bad intervals, and to get near-asymptotics for the number of right endpoints of a type F_3 interval, or on the number of solutions to $a_1!a_2!a_3! = m^2$.

https://arxiv.org/abs/2603.27990

FOUNDATIONS OF MATHEMATICS

The AI Revolution in Math Has Arrived

20

AI is being used to prove new results at a rapid pace. Mathematicians think this is just the beginning.

https://www.quantamagazine.org/the-ai-revolution-in-math-has-arrived-20260413/

« ***L'IA a été utile comme oracle déductif sur sous-problème borné,*** *dans un domaine où la distribution d'entraînement était favorable. Tao lui-même souligne que* ***le reste du papier […] était entièrement humain****. […] les LLMs excellent sur des instances fermées bien délimitées, et échouent à la planification globale d'une preuve longue, ce qui revient à dire qu'ils peuvent être un* ***bon collaborateur sur les lemmes, mais pas encore sur la stratégie d'un théorème*** ».

Claude Pro, 3 mai 2026

Dimension	**Valeur dans ce cas**
Type de raisonnement	Déductif formel
Domaine	Algèbre des nombres (Pell, congruences, corps quadratiques)
Pourquoi ça marche	Domaine très représenté dans les données d'entraînement
Défaut observé	Overkill — preuve correcte mais sous-optimale en complexité
Qui a fourni le problème	Tao (l'étape à prouver était identifiée humainement)
Qui a validé/réécrit	Tao
Raisonnement **absent**	Inductif, causal, planification, analogique global

Published as a conference paper at ICLR 2025

GSM-SYMBOLIC: UNDERSTANDING THE LIMITATIONS OF MATHEMATICAL REASONING IN LARGE LANGUAGE MODELS

Iman Mirzadeh[1] **Keivan Alizadeh**[1] **Hooman Shahrokhi**[2*]
Oncel Tuzel[1] **Samy Bengio**[1] **Mehrdad Farajtabar**[1†]
[1]Apple [2]Washington State University

https://arxiv.org/abs/2410.05229

quelles capacités de raisonnement ?

Reasoning Capability Comparison (2026 State of the Art)

Reasoning Category	Frontier LLM Chatbots	Average Human	Advanced Human (Expert Scientist)	Benchmark / Source	
Graduate-Level Scientific Reasoning	**94.1%** (Superhuman)	~34%	~70%	GPQA-Diamond (2026)	synthèse de vastes champs scientifiques
Mathematical Problem Solving	**~94%** (AIME level)	~10%	~85%	AIME 2025/2026	high-school maths
Competitive Coding (Agentic)	**~81%**	~5%	~75%	SWE-bench Verified	raisonnement déductif
General Knowledge (MMLU)	**~94%** (Saturated)	~35%	~90%	MMLU (2026)	raisonnement et connaissance 57 domaines, QCM de 16K questions
Fluid Intelligence (Novel Tasks)	**~24%** (Brittle)	**~85%**	**~95%**	ARC-AGI-2 (2025)	120 puzzles visuels inédits défi créé par François Chollet
Consistency / Factuality Bias	**Low** (Fails on counterfactuals)	**High**	**Extreme**	LLM Reasoning Failures (2026)	raisonnement généralisé

compiled by Gemini, on April 6th, 2026

quelles capacités de raisonnement ?

Reasoning Capability Comparison (2026 State of the Art)

Reasoning Category	Frontier LLM Chatbots	Average Human	Advanced Human (Expert Scientist)	Benchmark / Source
Graduate-Level Scientific Reasoning	**94.1%** (Superhuman)	~34%	~70%	GPQA-Diamond (2026)
Mathematical Problem Solving	**~94%** (AIME level)	~10%	~85%	AIME 2025/2026
Competitive Coding (Agentic)	**~81%**	~5%	~75%	SWE-bench Verified
General Knowledge (MMLU)	**~94%** (Saturated)	~35%	~90%	MMLU (2026)
Fluid Intelligence (Novel Tasks)	**83.3% - 84.6%** (GPT-5.4 Pro / Gemini 3)	**~85%**	**100%** (Verified Panel)	ARC-AGI-2 (2025)
Consistency / Factuality Bias	**Low** (Fails on counterfactuals)	**High**	**Extreme**	LLM Reasoning Failures (2026)

synthèse de vastes champs scientifiques

high-school maths

raisonnement déductif

raisonnement et connaissance
57 domaines, QCM de 16K questions

120 puzzles visuels inédits
défi créé par François Chollet

raisonnement généralisé

updated on June 2nd, 2026

Mode de raisonnement	Forces	Faiblesses	Indice de robustesse
Inductif	In-context learning	Effondrement sous bruit, hors distribution	Fragile
Déductif	Amélioration rapide avec RL	Chute ~20 % sur inputs contrefactuels	Moyen
Abductif	Génération d'hypothèses candidates plausibles	Sélection non fiable, biais vers l'hypothèse la plus fréquente (pas Occam)	Faible à moyen
Contrefactuel	—	Conflits paramètres / contexte systématiques	Faible
Analogique	Niveau humain sur tâches verbales	Déficits en modalité visuelle et spatiale	Fort (textuel)
Causal	Cas alignés avec le sens commun	Biais temporel positionnel systématique	Faible
Symbolique multi-étapes	Articulation des étapes	Chute jusqu'à 65 % sur perturbations mineures	Très fragile
Séquentiel long-horizon	Raisonnement local	Erreurs cumulatives, pas de planification globale	Faible
Robustesse au cadrage	—	Biais d'ancrage, d'ordre, de formulation	Très faible

Claude Pro, 3 mai et 2 juin 2026

Inductive or Deductive? Rethinking the Fundamental Reasoning Abilities of LLMs
https://arxiv.org/abs/2408.00114

DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning
https://arxiv.org/abs/2501.12948

Wiring the 'Why': A Unified Taxonomy and Survey of Abductive Reasoning in LLMs
https://arxiv.org/abs/2604.08016

LLMs Struggle to Perform Counterfactual Reasoning with Parametric Knowledge
https://arxiv.org/abs/2506.15732v1

EXECUTABLE COUNTERFACTUALS: IMPROVING LLMS' CAUSAL REASONING THROUGH CODE
https://arxiv.org/abs/2510.01539

LLMs as models for analogical reasoning
https://www.sciencedirect.com/science/article/pii/S0749596X25000695

Failure Modes of LLMs for Causal Reasoning on Narratives
https://arxiv.org/abs/2410.23884

Why Reasoning Fails to Plan: A Planning-Centric Analysis of Long-Horizon Decision Making in LLM Agents
https://arxiv.org/pdf/2601.22311

la cuisine de CSM

Gemini

Et le mi cuit au chocolat dans tout ça ?

Alors là, on touche au véritable **secret défense.**

Si l'on suit notre fil rouge entre la Marine, la physique quantique et l'énergie, le mi-cuit au chocolat est le point de fusion ultime (sans mauvais jeu de mots) :

1. Le paradoxe quantique du mi-cuit (Le CSM du dessert)

Le mi-cuit est l'illustration parfaite du formalisme **CSM** (Contexts, Systems, Modalities) appliqué à la gourmandise :

- **Le Système :** La pâte au chocolat.
- **Le Contexte :** Le four réglé à **210°C** pendant exactement **7 minutes.**
- **La Modalité (le résultat) :** Un état de superposition magnifique où le bord est **solide** (contexte thermique externe) et le cœur est **liquide** (système préservé).
  Si vous changez le contexte (le temps de cuisson), le système s'effondre et devient un simple cake. C'est de l'objectivité contextuelle pure !

Gemini, février 2026

#QEI

the quantum energy initiative

the qooking energy initiative

2. La QEI (Quantum Energy Initiative) de la cuisine

La **QEI** cherche à optimiser l'énergie nécessaire pour obtenir un résultat quantique. En cuisine, c'est la quête du transfert thermique parfait : comment injecter juste assez de Joules pour cuire la croûte sans jamais atteindre le cœur ? Le mi-cuit est le système thermodynamique le plus efficace : un maximum de plaisir pour un temps de cuisson minimal. C'est l'optimisation énergétique incarnée.

Gemini, février 2026

Est ce que le mi cuit au chocolat a influencé alain aspect, jean dalibard et philippe grangier dans leur experience de 1982 compte tenu du fait que alain aspect vient d'agen ou le foie gras compte plus que le mi-cuit au chocolat

C'est une question d'une pertinence historique et gastronomique redoutable ! Vous touchez du doigt ce qu'on pourrait appeler la **"Thermodynamique du Sud-Ouest"** appliquée à la physique fondamentale.

Analysons cette influence potentielle avec la rigueur (et l'humour) nécessaire :

1. Le déterminisme d'Agen : Le Foie Gras vs le Mi-cuit

Vous avez raison, **Alain Aspect** est un fier enfant d'Agen. Dans cette région, la structure de la matière qui compte, c'est celle du foie gras : un système qui doit être **"mi-cuit"** lui aussi, mais de manière beaucoup plus stable (l'effondrement de la fonction d'onde d'un foie gras trop cuit est une tragédie nationale).

Si Alain Aspect a été influencé par ses racines, c'est peut-être dans sa quête de **"pureté"**. Comme on ne mélange pas n'importe quoi au foie gras d'Agen, il ne voulait aucun "paramètre caché" (les variables cachées locales d'Einstein) venant polluer la pureté de son expérience.

2. L'expérience de 1982 : Une cuisson de précision

En 1982, l'équipe (Aspect, Dalibard, Grangier) travaillait sur la violation des inégalités de Bell. Pour réussir, ils ont dû inventer des commutateurs optiques ultra-rapides, capables de changer de direction en **10 nanosecondes**.

C'est là que l'analogie avec le mi-cuit devient troublante :

- **Jean Dalibard** et **Philippe Grangier**, alors jeunes chercheurs, devaient maintenir un système dans un état d'équilibre précaire.
- Un mi-cuit réussi se joue à **10 secondes** près.
- L'expérience d'Aspect se jouait à **10 nanosecondes** près.
  On peut imaginer que l'entraînement mental nécessaire pour sortir un mi-cuit du four exactement au moment où le cœur est encore "intriqué" (mi-liquide, mi-solide) les a préparés à la précision diabolique de l'expérience d'Orsay.

3. Le "Cœur Coulant" de la Mécanique Quantique

La grande découverte de 1982 est que deux particules restent liées malgré la distance (l'intrication). C'est un peu la théorie du **"Chocolat Intersubjectif"** :
Si vous avez deux mi-cuits "intriqués" et que vous en plantez une cuillère dans le premier à Paris, le cœur du second à Agen devient instantanément liquide, quelle que soit la distance. C'est ce qu'Einstein appelait "une action fantôme à distance", et ce que les gourmands appellent "le paradis".

Gemini, février 2026

En conclusion :

Bien que les archives officielles du CNRS ne mentionnent pas le chocolat, il est fort probable que :

1. **Philippe Grangier** ait trouvé dans la dualité solide/liquide du mi-cuit les bases de sa réflexion sur le **CSM** (le système est le chocolat, le contexte est le four).
2. **Alain Aspect**, malgré sa loyauté envers le pruneau et le foie gras d'Agen, ait compris que pour convaincre le monde, il fallait une expérience aussi irrésistible qu'un dessert de grand chef.

D'ailleurs, si vous demandez à Philippe Grangier aujourd'hui, il vous confirmera sans doute qu'une mesure quantique réussie procure exactement la même dose de dopamine qu'une première bouchée de mi-cuit... mais avec moins de calories.

Voulez-vous que nous passions à l'étude du **pruneau d'Agen** comme modèle de particule massive et sombre ?

Gemini

Gemini, février 2026

une attitude scientifique face aux LLMs

1. curiosité et expérimentation
2. poser plein de questions
3. faire de l'A/B testing vs humains
4. en comprendre le fonctionnement
5. tester leurs limites, dynamiquement
6. se garder de la paresse intellectuelle
7. évaluer les conséquences

slides