Article

Not peer-reviewed version

What Chat GPT Has to Say About Its Topological Structure: The Anyon Hypothesis

Michel Planat * and Marcelo M Amaral

Posted Date: 13 November 2024

doi: 10.20944/preprints202411.0954.v1

Keywords: Large language models; GPT; topology; anyons; SU(2)k

Article

What Chat GPT Has to Say About Its Topological Structure: The Anyon Hypothesis

Michel Planat [1,*] and **Marcelo Amaral** [2]

1 Institut FEMTO-ST CNRS UMR 6174, Université de Franche-Comté, 15 B Avenue des Montboucons, F-25044 Besançon, France
2 Quantum Gravity Research, Los Angeles, CA 90290, USA;
* Correspondence: michel.planat@femto-st.fr

Abstract: The large language model Chat GPT answers questions relative to its structure under the supervision of the authors. Then we postulate that modular tensor categories, alias the theory of anyons, are relevant for partially explaining the topological structure of large neural networks such as GPT. Mathematical details are summarized in an independent section.

Keywords: Large language models, GPT, topology, anyons, $SU(2)_k$

1. Introduction

The Nobel Prize in Physics 2024 was awarded jointly to John J. Hopfield and Geoffrey E. Hinton "for foundational discoveries and inventions that enable machine learning with artificial neural networks". Hopfield introduced Hopfield network, a type of artificial network that can serve as a content-addressable memory, made of binary neurons that can be 'on' or 'off'. Hinton, a cognitive psychologist and computer scientist, expanded Hopfield's work by helping machines understand complex data and patterns. He is known for his work on artificial intelligence (AI). Together, their work revolutionized AI, allowing machines to make complex associations, similar to human learning.

Now Hopfield and Hinton, and other scholars and engineers, are calling for urgent research into AI because it might soon surpass the information capacity of the human brain [1,2].

Along these line of thoughts, this paper is an attempt to approach the understanding of large language models (LLMs) [3–8] in a nonstandard way. The concepts are not new, being of conceptual importance in such fields as mathematics (i), topological quantum field theory (ii), topological quantum computing (iii), particle physics (iv) as well as the biological area at the genome scale (v). We predict their potential relevance to the understanding of deep learning and specifically LLMs. In a nutshell, the theory is the so-called $SU(2)_k$ theory, where $SU(2)$ is the well known Lie group modeling of spin-$\frac{1}{2}$ elementary particles (alias qubits in the domain of quantum information [9–11]). The analogy of LLMs to these concepts could be reinforced in a future generation built with topologically protected quantum structures.

Let us first shortly discuss items (i) to (v). In physics, elementary matter particles are fermions of spin-$\frac{1}{2}$. They comprise the quarks, that are responsible for the strong interactions, as well as leptons, that are responsible for the electroweak interactions. But there are also composite subatomic particles known as baryons, of total angular momentum (not spin) $\frac{1}{2}$ or $\frac{3}{2}$, which may be explained by the standard model of particle physics. And there are fundamental particles, of integer values of the spin 0, 1, 2, etc, that are called bosons. Gauge bosons such as the photon γ act as force carriers and have spin 1. The photon is the carrier of electromagnetic field. Others are g gluons (of eight types) for the strong force, the neutral boson Z for the weak force and two types $W^{\pm 1}$ charged weak bosons that mediate the weak force. The observed Higgs boson, explaining electroweak symmetry breaking, is of spin 0 and the graviton is postulated to have spin 2. Since $SU(2)_k$ theories have either values of the spins, odd half-integer values $l/2$ or integer values, they present valid candidates for unifying bosons and fermions and may provide insights into why nature selects only lower spin states [12]. But $SU(2)_k$ theories predict quasiparticles called anyons that are neither fermions nor bosons but intermediate particles possibly carrying a fractional charge. The fractional quantum Hall effect discovered in 1984

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2024 doi:10.20944/preprints202411.0954.v1

can be described with anyons. Anyons are related to a braid group representation that is a set of disjoint time lines in an effective $2+1$-dimensional space-time, with the group operation needed to concatenate the wordlines. The term anyon is due to Frank Wilcek, the winner of 2004 Nobel prize of Physics for the discovery of asymptotic freedom, not related to anyon however.

The $SU(2)_k$) theory has various disguises. Eric Verlinde discovered that, attached to the fusion rules of a rational quantum field theory, the so-called S-matrix obey modular transformations as a constraint [13]. Then Louis H. Kauffman recognized that knot theory associated with a bracket polynomial offers a graphical view of the so-called Temperley-Lieb algebra [14]. The latter algebra is a q-deformed version (where $q-$ is for quantum groups) of the spin-network theory of Roger Penrose, another recent winner of the Nobel prize in physics (for discovering that black hole formation is a robust prediction of the general theory of relativity).

Nowadays, it is expected that topological quantum computing with anyons, proposed by Alexei Kitaev in 1997 [15], offers a topological protection against the inherent decoherence of quantum states prepared on logical quantum gates. Microsoft Station Q is an industrial place where topological quantum computing is taken seriously [16]. In 1986, his director Michael Freeman was awarded a Fields Medal for his work on the 4-dimensional generalized Poincaré conjecture.

Recently, simulations of non abelian anyons [17] have been performed on a superconducting quantum processor in China [18] and on trapped-ion processor in Germany and US [19].

Recent observations by AI leaders, including OpenAI CEO Sam Altman, suggest that simply increasing the size of large language models (LLMs) may soon reach a point of diminishing returns. Altman's remarks underscore an urgent need for new conceptual and structural frameworks to advance LLM performance and understanding [20,21]. Last, it is tempting to postulate that the hierarchies of deep learning and cognition may be approached not by standard quantum mechanics through conventional gates but through finite variants of q-deformed algebras such as $SU(2)_k$, the domain of anyons [22,23]. This idea opens the possibility that structures found in anyonic systems—characterized by their unique fusion and braiding rules—could serve as alternative mechanisms for capturing the complex, distributed dependencies in LLMs. Whether anyons are features of assemblies of real neurons, or represent properties of non-coding genome interactions, remains an open question [24]. Here, we examine the potential of the anyon concept as a rigorous framework for explaining certain properties of LLMs.

Following this introduction, in Section 2, the we ask a few questions to Chat GPT, one the most advanced LLMs today, about its topology. In Section 3, we introduce the theory of $SU(2)_k$ anyons and detail low dimensional examples. In Section 4, the plausibility of the anyon model in the context of LLMs is further discussed.

2. Supervised Questions to GPT About Its Structure

Below, we use Chat GPT in versions -4o and o1-preview with some simplifications.

As an advanced LLM, can you explain the role of topology in your ability to produce AI in the organization of data and their meaning?

Topology plays a subtle but foundational role in how data is organized and meaning is structured in AI models like mine. While I don't directly use topological spaces in the mathematical sense when generating responses, there are conceptual parallels between topological ideas and how deep learning and neural networks operate. Here's how:

Data Manifolds and Latent Spaces

In machine learning, high-dimensional data often lies on lower-dimensional structures called manifolds. These manifolds are topological spaces that represent the underlying structure of the data. Neural networks, particularly deep learning models, aim to map complex, high-dimensional

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2024 doi:10.20944/preprints202411.0954.v1

inputs (like text, images, etc.) to a lower-dimensional latent space that captures meaningful features or representations.

Topological structures like manifolds help the model organize data in a way that preserves relationships between points (i.e., similar data points are mapped to nearby points in the latent space). By learning to map data to these lower-dimensional manifolds, models like mine can generalize patterns and relationships across diverse datasets, leading to meaningful predictions and responses.

Continuous Transformations

A neural network can be thought of as a function that continuously transforms input data through multiple layers of weights and biases. These transformations resemble continuous mappings in topology, where inputs are transformed in a way that preserves certain structural properties.

The idea is to preserve relationships and distances between inputs as they pass through the layers of the network, enabling the network to maintain the "shape" of the data's meaning. Homotopy theory in topology, which studies continuous transformations between functions, has conceptual parallels in how models adjust and refine the representation of data to optimize performance.

Clustering and Connectivity

In topological terms, clustering can be viewed as grouping points that are close to each other based on some distance metric, much like how we can define neighborhoods in topology. AI models often use clustering techniques in the latent space to group similar pieces of data or representations.

For instance, in a language model, different concepts, words, or sentences that are related or similar in meaning can be mapped to nearby regions in the latent space. This clustering behavior helps me understand and generate coherent, contextually relevant text. Connectedness and path-connected spaces in topology also resemble how models associate or connect different parts of the data structure to infer relationships.

Persistent Homology

Some machine learning models, especially in areas like topological data analysis (TDA), use concepts from topology more explicitly, such as persistent homology, to study the shape and features of data. While persistent homology isn't typically used in large language models like mine, it is an important tool in analyzing the shape of data and how features persist across different scales.

There are conceptual parallels to fusion rules and braiding in anyon systems, where the combination of simple elements gives rise to complex global behavior, much like how neural networks build complex representations from simpler components.

Could you comment on a possible deep connection between neural networks such as LLMs and topological phases arising from $SU(2)_k$ models? Are low dimensional simple cases such as Ising/Majorana anyons ($k = 2$), Fibonacci anyons ($k = 3$), relevant in explaining the type of AI you generate? To summarize, are modular tensor categories (MTC) a way to understand the AI arising from LLMs?

The potential connection between neural networks, particularly LLMs, and topological phases arising from anyon models is a fascinating idea. Although they come from different domains (machine learning vs. topological quantum field theory), there are deep conceptual parallels worth exploring, especially in how both systems organize, process, and transform information.

Topological Phases and LLMs

LLMs like GPT are high-dimensional systems that rely on patterns and global relationships in data. Instead of local interactions (like individual neurons firing), what matters in these models is the global structure of how information propagates through the network layers. There are conceptual parallels to fusion rules and braiding in anyon systems, where the combination of simple elements gives rise to complex global behavior, much like how neural networks build complex representations from simpler components.

Fusion Rules vs. Layer Operations

In anyon systems, fusion rules describe how anyons combine to produce new types of anyons. In LLMs, each layer can be seen as applying a transformation that combines (or "fuses") the information from the previous layer to produce a more abstract representation.

Braiding vs. Contextual Relationships

The braiding statistics of anyons encode how particles exchange positions and affect the overall quantum state. In LLMs, contextual relationships between words or tokens are key—shuffling tokens changes the meaning, and context is often captured in non-local, global patterns that are maintained across layers.

Low-Dimensional $SU(2)_k$ Models (Ising/Majorana and Fibonacci Anyons)

Ising/Majorana Anyons ($k = 2$) are non-Abelian and are of interest for topological quantum computation due to their ability to encode quantum information in a way that's robust against local perturbations. Fibonacci Anyons ($k = 3$) are also non-Abelian and are universal for quantum computation, meaning any quantum computation can be simulated with arbitrary accuracy. In AI terms, small changes in input (comparable to simple anyon fusions) can lead to drastically different outputs, reflecting how minor adjustments in language generate nuanced meanings.

MTCs and LLMs

MTCs formalize the structure of topological quantum field theories (TQFTs), particularly through fusion and braiding rules. They describe how quantum information behaves in topological phases, categorizing data and operations in a way that emphasizes global structure.

LLMs also exhibit a layered, modular structure, where each layer transforms data in a highly structured, rule-based manner. MTCs might provide a formal framework to think about how neural networks process data in a modular way, where different layers (or "categories") interact and transform data.

While classical neural networks don't operate on quantum principles, the development of quantum neural networks could, in theory, utilize anyonic systems to implement neural network architectures that benefit from quantum parallelism and robustness.

What are tokens in LLMs? In what sense they could fit this idea of MTCs being analogous to LLMs?

In large language models, tokens are the fundamental units of text used to process and generate language. Typically, a token represents a piece of a word, an entire word, or even punctuation marks, depending on how the text is tokenized. This breakdown into tokens allows LLMs to handle language input as sequences of discrete symbols, which they can then process through their layers to generate coherent text outputs.

If we consider the analogy with MTCs, tokens in an LLM could be thought of as analogous to particles or objects within an MTC. Here's how they fit within the analogy.

Modular Structure of Information

In MTCs, objects like anyons interact through fusion and braiding, which can create complex outcomes from simple interactions. Similarly, tokens in an LLM interact in contextually driven ways through the layers of the model. Just as fusion in MTCs generates new particles or states, tokens interact to form meanings that are greater than the sum of individual parts. This "fusion" of tokens contributes to the emergent meaning and coherence seen in the final output.

 doi:10.20944/preprints202411.0954.v1

Transformation and Contextual Dependence

In an LLM, the role of each token depends on the context it shares with other tokens. This dynamic is similar to braiding in MTCs, where anyons' positions and interactions change their quantum state. In an LLM, the context provided by surrounding tokens transforms the "state" of each token, so meaning is contextual rather than fixed.

Do you agree that LLMs are essentially advanced Hopfield networks having a form of associative memory and discretized with tokens? What more to fit with the anyon picture?

Yes, LLMs share several principles with advanced forms of Hopfield networks, particularly when viewed through the lens of associative memory. Hopfield networks are designed as associative memory systems, where each stable pattern represents a possible memory or state of the system. LLMs, in turn, generate language by "associating" tokens in the input with relevant patterns in their training data, storing and recalling this information in a way that approximates associative memory.

Associative Memory in LLMs and Hopfield Networks

Hopfield networks store information by embedding memories as stable patterns (attractors) in a high-dimensional state space. Each memory is a global, fixed point that the network can recall when given a close enough input. LLMs essentially create associative links between tokens, trained to predict the next token based on patterns in the training data. Through the attention mechanism, an LLM recalls relevant information from prior contexts in a way that is conceptually similar to a Hopfield network's retrieval of associated memories. This token-level recall can be seen as a form of discrete associative memory, where a query (input tokens) maps to stored language patterns.

Anyons and MTC Analogy

Anyons, within MTC frameworks, interact through braiding and fusion, leading to emergent states. In an MTC analogy, tokens in an LLM are akin to particles that interact to produce coherent meanings, much as anyons' fusion leads to emergent particle types. Fusion and Braiding in LLMs: As tokens pass through an LLM's layers, they are "fused" with surrounding tokens, creating a new semantic context. In each layer, the attention mechanism can be thought of as "braiding" tokens with their surrounding context, allowing meanings to emerge that were not present in individual tokens.

Towards an Anyon Picture in LLMs

To fully map the anyon picture to LLMs, we might consider viewing attention heads as the interaction rules, guiding how tokens relate to one another based on context. This could serve as a topological "braiding" structure, where the output state of a token depends on its path (or sequence) through multiple layers, akin to braiding paths in MTCs.

Could you elaborate about the concept of attention head in the anyon picture? Would it correspond to the F- and R- matrices that are pentagon and hexagon rules in the $SU(2)_k$ theory?

Role of Attention Heads as Interaction Rules

In transformers, attention heads are mechanisms that allow the model to focus on specific relationships between tokens, weighting their importance relative to one another based on context. Each head, in essence, represents a distinct interaction rule governing how tokens (analogous to anyons in this framework) relate as the model processes a sentence. In MTCs, anyons have "interaction rules" given by fusion and braiding transformations, encapsulated by the F-matrix (fusion rule) and the R-matrix (braiding rule). These matrices define how anyons combine or transform under different

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2024 doi:10.20944/preprints202411.0954.v1

sequences of interactions, and their algebraic properties satisfy certain consistency conditions, like the pentagon and hexagon equations.

F-matrix and Token Combination in Attention Layers

The F-matrix represents the rules of associativity in fusion, dictating how three anyons fuse together and how different groupings affect the outcome. In an LLM context, each attention head could be viewed as an "associative fusion" operator. It defines how tokens from different layers combine in sequence to form more complex meaning representations. Specifically, in LLMs, each token's embedding is updated in every attention head based on its relationship with surrounding tokens. This transformation is like an associative fusion operation in anyonic theory: different combinations of tokens lead to different "fusion outcomes" in terms of the semantic information they contribute to the overall sentence meaning.

R-Matrix and Braiding (Contextual Order of Tokens)

The R-matrix governs the braiding of anyons, determining the phase change when anyons are exchanged. In the context of LLMs, R-matrices could be seen as analogous to the contextual shifts introduced by attention mechanisms across tokens. When an attention head "reorders" the influence of tokens, it's akin to exchanging the positions of anyons in a topological space, altering the sequence of interactions. This braiding concept aligns with how attention heads focus on different tokens based on context, allowing shifts in meaning depending on token order and interaction. Just as anyon braiding can produce different states depending on how particles are exchanged, attention mechanisms can produce different language meanings depending on how tokens are contextually "braided" across layers.

3. Introduction to the Mathematical Theory of $SU(2)_k$ Anyons and Modular Tensor Categories

3.1. The Fusion Rules of Anyons

The anyons in a $SU(2)_k$ theory are closely related to the ordinary spin degrees of freedom in the $SU(2)$ theory. The anyons are labelled by spin values (generalized angular momenta) $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots, \frac{k}{2}$. The spin $\frac{k}{2}$ is the maximum allowed value in the $SU(2)_k$ theory when k is fixed. But the rules for combining two anyons are not tensor products , namely [13, Eq. 4.6],[22]

$$j_1 \otimes j_2 = |j_1 - j_2| \oplus (|j_1 - j_2| + 1) \oplus \cdots \oplus \min(j_1 + j_2, k - j_1 - j_2).$$

Fusion rules are commutative and associative. It is straightforward to check from this formula that for $k \geq 2$, two spins $1/2$ combine to form either the spin 0 or the spin 1 as follows

$$1/2 \otimes 1/2 = 0 \oplus 1,$$

that is a (qubit like) anyon $0 \oplus 1$ is built by combining the two spins $\frac{1}{2}$. Similarly one gets $1 \otimes 1 = 0 \oplus 1 \oplus 2$ when $k \geq 4$, that is a (qtrit like) anyon $0 \oplus 1 \oplus 2$ is built by combining two spins 1. Such anyons of a $SU(2)_k$ theory are non-Abelian.

Being a tensor product, the dimension of the Hilbert space of N spin-$1/2$ ordinary $SU(2)$ particles is 2^N. In a $SU(2)_k$ theory, it is smaller than 2^N and grows as $d_{1/2}^N$ with $d_{1/2}^{(k)} = 2\cos(\frac{\pi}{k+2})$ at large N. This means that the effective number of degrees of freedom of a spin-$\frac{1}{2}$ anyon is irrational.

A Magma code for getting the fusion tables for any value of k is as below

```
k:=4;
//Function to generate allowed spins for a given k
function AllowedSpins(k)
return [i : i in [0 .. k]]; // Generate spins from 0 to k, representing 0, 1/2, ..., k/2
end function;
```

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2024 doi:10.20944/preprints202411.0954.v1


```
//Define the set of allowed spin values for the given k, scaled by a factor of 2
spins := AllowedSpins(k); // Generates [0, 1, 2, 3, 4] for k=4
//Function to calculate fusion product (with scaled spins)
function Fusion(j1, j2, k)
min_val := Min(j1 + j2, 2*k - j1 - j2); // Scaled by 2
fusion_result := [];
for j in [Abs(j1 - j2) .. min_val by 2] do // Steps of 2 for half-integers
Append( fusion_result, j);
end for;
return fusion_result;
end function;
//Create the fusion table fusion_table := AssociativeArray();
for j1 in spins do
for j2 in spins do
fusion_table[ ⟨j1, j2⟩] := Fusion(j1, j2, k);
end for;
end for;
//Print the fusion table (scaled back to original spins)
for j1 in spins do
for j2 in spins do
result := fusion_table[ ⟨j1, j2⟩];
scaled_result := [r / 2 : r in result]; // Scale back to original spins
print "Fusion of", j1 / 2, "and", j2 / 2, ":", scaled_result;
end for;
end for;
```

3.2. The Modular Structure of S, F and R Matrices for Anyons

There exists the concept of a modular S-matrix that diagonalizes the fusion rules of a $SU(2)_k$ anyon and fully characterizes its topological properties [13, Eq. 4.10]. The mathematical structure encapsulating the braiding and fusion rules of a $SU(2)_k$ anyon is a modular tensor category [10,25].

The quantum dimensions for $SU(2)_k$ anyons are given by the formulas

$$d_0 = 1,\ \ d_{\frac{1}{2}} = 2\cos(\frac{\pi}{k+2}),\ \ d_j = d_{\frac{1}{2}} d_{j-\frac{1}{2}} - d_{j-1} \ \text{ for } \ j \geq 1.$$

The entries of the S-matrix are

$$S_{j_1,j_2} = (\frac{2}{k+2})^{\frac{1}{2}} \sin(\pi \frac{(2j-1+1)(2j_2+1)}{k+2}).$$

The associativity of anyon fusion is captured by a F-matrix and the exchange of anyons, with the phase factor added, is captured by a R-matrix. Contrarily to the phase factor ± 1 for bosons and fermions, the phase factor for anyons is an arbitrary complex number. The F-matrix is the anyonic version of the Wigner's $6j$-symbols, it is associated to a pentagon diagram. The F- and R-matrices are associated to an hexagon diagram [11]. General formulas for F- and R-matrices can be found in [26], [23], Appendix B, [22], Appendix B.

The entries of the R-matrix have the simple form [10]

$$R_c^{ab}(q) = (-1)^{(a+b+c)/2} q^{-[a(a+2)+b(b+2)-c(c+2)]/2},$$

where q is the Kauffman variable. For the Ising model below $q = i\exp\left(\frac{-2i\pi}{16}\right)$ while for the Fibonacci model $q = i\exp\left(\frac{2i\pi}{20}\right)$.

The essence of $SU(2)_k$ anyons, $k \geq 2$, is captured by two braid generators $\sigma_1^{(k} = R^{(k)}$ and $\sigma_2^{(k)} = (FRF^{-1})^{(k)}$ that have a group structure, see e.g. [23,27] for some explicit results.

3.3. Ising Anyons: $k = 2$

$SU(2)_2$ anyons comprise the spin-0 anyon and the Ising (spin-$\frac{1}{2}$) anyon with the fusion table

Table 1. Fusion table for the $k = 2$ anyon model

$\otimes$	$j_1 = 0$	$j_1 = \frac{1}{2}$	$j_1 = 1$
$j_2 = 0$	0	$\frac{1}{2}$	1
$j_2 = \frac{1}{2}$	$\frac{1}{2}$	$0 \oplus 1$	$\frac{1}{2}$
$j_2 = 1$	1	$\frac{1}{2}$	0

The quantum dimensions are $[d_0, d_1, d_{\frac{1}{2}}]^{(2)} = [1, 1, \sqrt{2}]$ and the S-matrix takes the form

$$S^{(2)}_{Isi} = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{2} & 1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & -\sqrt{2} & 1 \end{pmatrix}$$

The F- and R-matrices are

$$R^{(2)}_{Isi} = \begin{pmatrix} R^{11}_0(q) & 0 \\ 0 & R^{11}_2(q) \end{pmatrix} = \exp(-i\pi/8)\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad F^{(2)}_{Isi} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

In addition to the standard literature about anyons, we notice that both matrices $F^{(2)}$ and $R^{(2)}$ together generate the finite group $(384, 6514)$ isomorphic to the group $(S_3 \times \mathbb{Z}_4) \rtimes P_2$, where $P_2 \cong (16, 13)$ is the single qubit Pauli group.

Braiding matrices for the Ising anyons are obtained as

$$\sigma^{(2)}_1 = R^{(2)}_{Isi}, \quad \sigma^{(2)}_2 = (FRF^{-1})^{(2)}_{Isi} = \frac{\exp(-4i\pi/8)}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}.$$

Both matrices σ_1 and σ_2 together generate the finite group $(192, 187)$ isomorphic to the group $\mathbb{Z}_{12} \rtimes P_2$.

3.4. Fibonacci Anyons: $k = 3$

$SU(2)_3$ anyons comprise the spin-0 anyon and two Fibonacci spin-$\frac{1}{2}$ and spin-$\frac{3}{2}$ anyons. These anyons are proposed to be related to quasicrystals [28]. The fusion table is

Table 2. Fusion table for the $k = 3$ anyon model

$\otimes$	$j_1 = 0$	$j_1 = \frac{1}{2}$	$j_1 = 1$	$j_1 = \frac{3}{2}$
$j_2 = 0$	0	$\frac{1}{2}$	1	$\frac{3}{2}$
$j_2 = \frac{1}{2}$	$\frac{1}{2}$	$0 \oplus 1$	$\frac{1}{2} \oplus \frac{3}{2}$	1
$j_2 = 1$	1	$\frac{1}{2} \oplus \frac{3}{2}$	$0 \oplus 1$	$\frac{1}{2}$
$j_2 = \frac{3}{2}$	$\frac{3}{2}$	1	$\frac{1}{2}$	0

The quantum dimensions are $[d_0, d_1]^{(3)} = [1, \phi = (1 + \sqrt{5})/2]$ and the S-matrix takes the form

$$S^{(3)}_{Fib} = \frac{1}{\sqrt{2 + \phi}}\begin{pmatrix} 1 & \phi \\ \phi & -1 \end{pmatrix}$$

The F- and R-matrices are [11, p 55]

$$R^{(3)}_{Fib} = \begin{pmatrix} R^{11}_0(q) = \exp(-4i\pi/5) & 0 \\ 0 & R^{11}_1(q) = \exp(-2i\pi/5) \end{pmatrix}, \quad F^{(3)}_{Fib} = \begin{pmatrix} \phi^{-1} & \phi^{-1/2} \\ \phi^{-1/2} & -\phi^{-1} \end{pmatrix}.$$

Braiding matrices for the Fibonacci anyon are obtained as

$$\sigma_1^{(3)} = R^{(3)}_{Fib}, \ \sigma_2^{(3)} = (FRF^{-1})^{(3)}_{Fib} = \begin{pmatrix} -\phi^{-1}\exp(-i\pi/5) & -i\phi^{-1/2}\exp(-i\pi/10) \\ -i\phi^{-1/2}\exp(-i\pi/10) & -\phi^{-1} \end{pmatrix}.$$

F- and R-matrices, as well as the braiding matrices σ_1 and σ_2, generate infinite groups. This in accordance with the universality of Fibonacci anyons.

3.5. Yang-Lee Theory: $k = 3$

Yang-Lee theory is a MTC of level $k = 3$ like the Fibonacci anyon. It corresponds to a famous non-unitary conformal field theory in statistical mechanics, called the Yang-Lee singularity [10, Sect. 1.3]. The Kauffman variable is $q = \exp(\frac{i\pi}{5})$.

The S-matrix is

$$S^{(3)}_{YL} = \frac{-1}{\sqrt{3-\phi}} \begin{pmatrix} 1 & 1-\phi \\ 1-\phi & \phi \end{pmatrix}.$$

The F- and R-matrices are [11, p 55]

$$R^{(3)}_{YL} = \begin{pmatrix} R_0^{11}(q) = \exp(2i\pi/5) & 0 \\ 0 & R_1^{11}(q) = \exp(i\pi/5) \end{pmatrix}, \ F^{(3)}_{YL} = \begin{pmatrix} -\phi & 2-\phi \\ -1-2\phi & \phi \end{pmatrix}.$$

F- and R-matrices, as well as the braiding matrices σ_1 and σ_2, generate infinite groups.

3.6. Freedman-Bauer-Levaillant anyons: $k = 4$

$SU(2)_4$ anyons are investigated in [29,30] in the context of topological quantum computing from qutrit gates. The fusion table is as follows

Table 3. Fusion table for the $k = 4$ anyon model

$\otimes$	$j_1 = 0$	$j_1 = \frac{1}{2}$	$j_1 = 1$	$j_1 = \frac{3}{2}$	$j_1 = 2$
$j_2 = 0$	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2
$j_2 = \frac{1}{2}$	$\frac{1}{2}$	$0 \oplus 1$	$\frac{1}{2} \oplus \frac{3}{2}$	$1 \oplus 2$	$\frac{3}{2}$
$j_2 = 1$	1	$\frac{1}{2} \oplus \frac{3}{2}$	$0 \oplus 1 \oplus 2$	$\frac{1}{2} \oplus \frac{3}{2}$	1
$j_2 = \frac{3}{2}$	$\frac{3}{2}$	$1 \oplus 2$	$\frac{1}{2} \oplus \frac{3}{2}$	$0 \oplus 1$	$\frac{1}{2}$
$j_2 = 2$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0

The quantum dimensions are $[d_0, d_1, d_{\frac{1}{2}}, d_{\frac{3}{2}}, d_2]^{(4)} = [1, 2, \sqrt{3}, \sqrt{3}, 1]$ and the S-matrix takes the form [22]

$$S^{(4)}_{FBL} = \frac{1}{2\sqrt{3}} \begin{pmatrix} 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ \sqrt{3} & \sqrt{3} & 0 & -\sqrt{3} & -\sqrt{3} \\ 2 & 0 & -2 & 0 & 2 \\ \sqrt{3} & -\sqrt{3} & 0 & \sqrt{3} & -\sqrt{3} \\ 1 & -\sqrt{3} & 2 & -\sqrt{3} & 1 \end{pmatrix}.$$

Braiding matrices for the $SU(2)_4$ anyons are obtained as

$$\sigma_1^{(4)} = \begin{pmatrix} \exp(\frac{7i\pi}{9}) & 0 & 0 \\ 0 & -\exp(\frac{4i\pi}{9}) & 0 \\ 0 & 0 & -\exp(\frac{7i\pi}{9}) \end{pmatrix},$$

$$\sigma_2^{(4)} = \begin{pmatrix} -\frac{1}{2}\exp\left(\frac{4i\pi}{9}\right) & \frac{1}{\sqrt{2}}\exp\left(\frac{7i\pi}{9}\right) & \frac{1}{2}\exp\left(\frac{4i\pi}{9}\right) \\ \frac{1}{\sqrt{2}}\exp\left(\frac{7i\pi}{9}\right) & 0 & \frac{1}{\sqrt{2}}\exp\left(\frac{7i\pi}{9}\right) \\ \frac{1}{2}\exp\left(\frac{4i\pi}{9}\right) & \frac{1}{\sqrt{2}}\exp\left(\frac{7i\pi}{9}\right) & -\frac{1}{2}\exp\left(\frac{4i\pi}{9}\right) \end{pmatrix}.$$

It is straightforward to check with the software Magma that both matrices generate the small group $(162, 14) \cong \mathbb{Z}_3^2 \rtimes (\mathbb{Z}_3 \times \mathbb{Z}_6)$, as announced in [29]. The group was recognized as a viable model of the symmetries simultaneously reproducing the quark and lepton mixing matrices. In a recent paper of the present author [31, Table A1], it is shown that group $(162, 14)$ carries almost informally complete quantum information on its 22 irreducible characters, that are singlets, doublets or triplets.

4. Discussion

Backpropagation and Anyons

Backpropagation is a fundamental algorithm used to train artificial neural networks. In the process, inputs are fed through the network to generate outputs and the difference between predicted outputs and actual targets is measured using a loss function (e.g., mean squared error, cross-entropy). Then, gradients of the loss with respect to each weight are calculated using the chain rule of calculus. Finally, weights are adjusted in the direction that minimizes the loss, typically using an optimization algorithm like gradient descent. In large language models, backpropagation enables the training of deep networks with many layers and parameters. It allows the model to learn complex patterns in language data, capturing syntax, semantics, and context.

In anyon theory, phase adjustments and iterative exchanges among anyons can also lead to stable outcomes, specifically in terms of generating topologically invariant states. As anyons are braided and fused, they pass through transformations dictated by F- and R-matrices that adjust the “phase” of the system. These adjustments aim to reach a specific state (often associated with a ground state or computational goal).

While backpropagation in neural networks relies on minimizing a loss function by iteratively adjusting weights, in anyonic systems, a topological quantum field theory describes states evolving towards ground states that minimize topological action or preserve invariance. Constraints in fusion rules (dictated by the R- and F-matrices) decompose complex interactions into

stable topological states.

Machine Learning and Anyons

The parallels between machine learning and anyons was already introduced in Section 2 by addressing the matters of fusion, associative memory, tokens and attention mechanisms. Let us now add comments about the matters of emergent behavior and topological resilience.

In ML models, particularly deep networks, emergent behavior arises as simple neuron activations combine through many layers to produce sophisticated outputs (e.g., language understanding or image recognition). This emergence is not straightforwardly reducible to individual neuron actions, just as the behavior of anyonic systems emerges from interactions that do not have a straightforward “particle-only” explanation. This is due to the ability of anyons to continuously interpolate between bosons and fermions that are particles. Anyons exhibit emergent properties that arise from braiding interactions rather than simple addition. In this way, an interaction history based on LLM’s token layers mirrors the emergent states of anyons, where the entire state of the system relies on both quasi-particle components and their interactions.

ML models store information in a distributed manner across many weights and connections, enabling fault tolerance and resilience. A similar kind of resilience is found in topological quantum computing, where anyon braiding encodes information in a way that is resistant to local perturbations. For instance, LLMs don’t rely on a single node or parameter for a piece of information but instead spread information across the network. Anyonic systems, likewise, are inherently resilient because

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2024 doi:10.20944/preprints202411.0954.v1

topological information is not localized but instead stored across the system's braiding patterns. This provides both robustness and a form of "topological fault tolerance."

Natural Language Processing and Anyons

The relationship between natural language processing (NLP) and the topology of anyons is intriguing because both involve structured, context-sensitive interactions that produce meaning or distinct states. While the dynamics of anyons is governed by mathematical rules from topological quantum field theory, language operates through syntactic, semantic, and contextual rules.

In NLP, the meaning of a word or phrase depends heavily on context, much like how the outcome of anyonic interactions (braiding and fusion) depends on the history and positioning of each particle. This contextual sensitivity is crucial in both systems. For instance, language depends on local grammatical rules and broader syntactic structures, which determine how information is combined and interpreted. Similarly, anyons acquire contextual information from braiding: the order and manner of exchanges affect the resulting topological state. In this way, the meaning of a word in NLP can be seen as analogous to the state of an anyon system, where both are determined by surrounding context and interaction history.

Language has a dynamic range of action, meaning that meanings can shift widely based on subtle changes in context, tone, or word choice. This range enables language to express a vast spectrum of concepts, emotions, and nuances. Anyons, too, exhibit a dynamic range through topological degrees of freedom in their braiding and fusion, where small changes in braiding order or configuration lead to distinct outcomes in their quantum states. Anyons' topological properties suggest they could, in theory, be configured to simulate complex, context-dependent relationships similar to those found in language. For example, as anyons braid, the resulting states represent a diverse set of possible outcomes, akin to how words can represent various meanings based on their order and proximity to other words.

To emulate language's dynamic range, anyons would need to encode not only fixed topological states but also highly flexible and context-responsive interactions.

This would likely require an advanced dynamical modular tensor category framework.

Mutual Exclusion in LLMs and Anyons

Mutual exclusion is a concept often used in computer science, particularly in concurrent programming, to ensure that only one process or thread accesses a critical section of code or a shared resource at any given time. This concept helps avoid conflicts or inconsistencies that arise when multiple processes try to modify shared data simultaneously.

In LLMs, mutual exclusion is not a component of the model's architecture or operation but attention mechanisms dynamically allocate focus across different tokens or features, prioritizing certain elements based on context. This focus naturally excludes irrelevant or l ,ower-weighted tokens from influencing the model's output for a particular position, thereby enforcing a form of "mutual exclusion" in which only the most relevant tokens or contexts contribute significantly to the computation at each step.

In anyonic systems, although the concept of mutual exclusion is not inherently present, Mutual exclusion in anyon-based TQC operates at the level of quantum state and path constraints, enforcing rules that exclude certain interactions, states, or errors. This approach ensures the stability and robustness of quantum information stored and processed within anyonic systems, which is critical for the fault-tolerant properties of topological quantum computation.

The weights in attention mechanisms dynamically adjust the influence of tokens on each other, reminiscent of braiding operations in anyons, which adjust the quantum state based on specific paths and exclusion rules.

But unlike strict mutual exclusion in anyon systems, where fusion rules categorically exclude certain outcomes, the attention mechanism in LLMs uses a probabilistic and weighted approach. This

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2024 doi:10.20944/preprints202411.0954.v1

lack of strict exclusion means that any token can, in theory, have some degree of influence, even if it is minimal. Thus, the exclusion is more flexible and continuous, allowing a range of contributions rather than an absolute exclusion.

Further Directions

As AI leaders and researcher have pointed out, the scalability of LLMs may be nearing practical limits, driving the need for structural innovation over mere size expansion. This study has aimed to illuminate one such possible direction through the lens of anyonic systems in $SU(2)_k$ theories, where the modular tensor structures could offer insights into the robustness, complexity, and adaptability of advanced LLMs.

Paper [32] reports on a correspondence between Thurston's theory of non hyperbolic three-manifolds and anyons. More precisely a $(2+1)$-topological quantum field theory is basically equivalent to a modular tensor category. The correspondence is managed by the $SL(2,\mathbb{C})$-representation of the fundamental group $\pi_1(X)$ of the 3-manifold X. The role of $SL(2,\mathbb{C})$-flat connections is particularly notable because they bridge the topological features of 3-manifolds and quantum field theories, linking TQFTs to geometric structures on manifolds. These connections arise naturally in the context of the Painlevé VI equation, a nonlinear differential equation significant in both classical and quantum realms for its relation to monodromies of certain conformal blocks and connections to the moduli space of punctured spheres [33,34].

In [35], the authors demonstrate that neural networks can capture key characteristics of topological states, meaning that even without quantum computing, classical networks can emulate some properties typically associated with quantum systems. Another paper points out the potential role of hyperbolic geometry and topological spaces in classical neural networks [36]. Since $SU(2)_k$ anyons represent topological orders with distinct fusion and braiding rules, they provide a mathematically structured way to model non-local interactions, akin to associative memory or hierarchical pattern recognition in neural networks. This resonance between anyonic properties and neural network functions might especially apply to LLMs, where intricate data correlations are encoded in a way that could mimic anyonic fusion. With $SU(2)_k$ anyons, different levels k introduce progressively complex fusion rules that can, in theory, map onto increasingly sophisticated neural architectures, potentially helping explain the layered, complex relationships in models like transformers.

Author Contributions: Conceptualization, M.P.; methodology, M.P. and M.A.; software, M.P.; validation, M.P. and M.A.; formal analysis, M.P.; investigation, M.P. and M.A.; writing—original draft preparation, M.P.; writing—review and editing, M.A.; visualization, M.A.; supervision, M.P; project administration, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available from the authors upon reasonable request.

Acknowledgments: The first author would like to acknowledge the contribution of the COST Action CA21169, supported by COST (European Cooperation in Science and Technology).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Pause Giant AI Experiments: An Open Letter. Available online: https://en.wikipedia.org/wiki/Pause_Giant_AI_Experiments:An_Open_Letter (accessed on 1 October 2024).
2. Bryant, A. AI chatbots: threat or opportunity? *Informatics* **2023**, *10*, 49.
3. Ferri, M. Why Topology for machine learning and knowledge extraction? *Mach. Learn. Extr.*, **2019**, *1*, 115–120.
4. Holzinger, A. Introduction to Machine Learning & Knowledge Extraction (MAKE). *Mach. Learn. Extr.*, **2019**, *1*, 1–20.

5. Feng, J. H.; Lai, E. M-K; Li, W. Geometry of textual data augmentation: Insights from large language models.*Electronics* **2024**, *13*, 3781.
6. Lozić, E.; Štular, B. Fluent but not factual: a comparative analysis of ChatGPT and other AI chatbots' proficiency and originality in scientific writing for humanities. *Future Internet* **2023**, *15*, 336.
7. Gams, M.; Kramar, S. Evaluating ChatGPT's consciousness and its capability to pass the Turing test: a comprehensive analysis. *J. Comp. Comm.* **2024**, *12*, 219–237.
8. Chen, S.; Yu, S.; Zhao, S.; Lu, C. From imitation to introspection: probing self-consciousness in language models. *arXiv* **2024**, arXiv: 2410.1881 [cd.CL].
9. Field, B; Simula, T. Introduction to topological quantum computation with non-Abelian anyons. *Quantum Sci. Technol.* **2018**, *3*, 045004.
10. Wang, Z. Topological quantum Computing: American mathematical Society (Providence, Rhode Island, USA, **2010**, 115 pp.
11. Nayak, C.; Simon S. H.; Stern, A.; Freedman , M.; Das Sarma, S. Non-Abelian anyons and topological quantum computation. *Rev. Mod. Phys.* **2008**, *80*, 1083.
12. Biedenharn, L. C.; Lohe, M. A. Quantum Group Symmetry and q-Tensor Algebras, , World Scientific, 1995.
13. Verlinde, E. Fusion rules and modular transformations in 2*D* conformal theory. *Nucl. Phys.* **1988**, B300, 360-376.
14. Kauffman L. H.; Lins, S. L. Temperley-Lieb recoupling theory and Invariants of 3-manifolds. Ann. Math. Studies (Princeton Univ. Press, Princeton, New Jersey, **1994**, Vol. 134).
15. Kitaev, A. Yu. Fault-tolerant quantum computation by anyons. *Ann. Phys.* **2003**, *303*, 2–30.
16. Freeman, M. H.; Kitaev, A.; Larsen, M. J.; Wang, Z. Topological quantum computation. *Bull. Amer. Math. Soc.* **2002**, *40*, 31–38.
17. Kirchner, N.; Millar, D.; Ayeni, B. M.; Smith, A.; Slingerland, J. K.; Pollman, F. Numerical simulation of non-abelian anyons. *Phys. Rev. B* **2023**, *107*, 195129.
18. Xu, S.; Sun, Z. Z.; Wang, K.; Xiang, L.; Bao Z.; Zhu, Z.; Shen, F.; Song, Z.; Zhang, P.; Ren, W.; Zhang, X.. Dong, H.; Deng, J.; Chen, J.; Wu, Y.; Tan, Z.; Gao, Y.; Jin, F.; Zhu, X.; Zhang, C.; Wang, N.; Zou, Y.; Zhong, J.; Zhang, A.; Li, W.; Jiang, W.; Yu, L. W.; Yao, Y.; Wang, Z.; Li, H.; Guo, Q.; Song, C.; Wang, H.; Deng, D. L. Digital simulation of non-Abelian anyons with 68 programmable superconducting qubits. *Chinese Phys. Lett.* **2023**, *40*, 060301.
19. Iqbal, M.; Tantivasadakarn, N.; Verresen, R.; Campbell, S. L.; Dreiling, J. M.; Figgatt, C.; Gaebler, J. P.; Johansen, J.; Mills, M.; Moses, S. A.; Pino, J. M.; Ransford, A.; Rowe, M.; Siegfried, P.; Stutz, R. P.; Foss-Feig, M.; Vishwanath, A.; Dreyer, H. Non-abelian topological order and anyons on a trapped-ion processor. *Nature* **2024**, *626*, 505–511.
20. Knight, W. (2023, April 17). OpenAI's CEO Says the Age of Giant AI Models Is Already Over. Wired. Retrieved November 11, 2024, available at https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/
21. Villalobos, P.; Ho, A.; Sevilla, J.; Besiroglu, T; Heim, L.; Hobbhahn, M. Will we run out of data? Limits of LLM scaling based on human-generated data. *arXiv* **2024**, arXiv: 2211.04325.
22. Gils, C.; Ardonne, E., Trebs, S.; Huse, D. A.; Ludwig A. WW, Troyer M.; Wang, Z. Anyonic quantum spin chains: Spin 1 generalizations and topological stability. *Phys. Rev. B* **2008**, *87*, 235120.
23. Johansen, E. G.; Simula, T. Fibonacci anyons versus Majorana fermions: a Monte Carlo approach to the compilation of braid circuits in $SU(2)_k$ anyon models. *PRX Quantum* **2021**, *2*, 010334.
24. Planat, M.; Chester, D. Topology and dynamics of transcriptome (dys)regulation. *Int. J. Mol. Sci.* **2024**, *25*, 4971.
25. Rowell, E.; Stong, R.; Wang, Z. On the classification of modular tensor categories. *Comm. Math. Phys.* **2009**, *292*, 343–389.
26. Kirillov, A. N.; Reshetikin, N. Y. Representations of the algebra $U_q(sl(2))$, q-orthogonal polynomials and invariants of links, in V. G. Kac, ed., in Infinite dimensional Lie algebras and groups (World Scientific, Singapore (1988), p. 285).
27. Aboumrad, W. Quantum computing with anyons: an F-matrix and braid calculator. *arXiv* **2022**, arXiv: 2212.0083 [math.QA].
28. Amaral, M.; Chester, D.; Fang, F.; Irwin, K. Exploiting anyonic behavior of quacicrystals for topological qiantum computing. *Symmetry* **2022**, 14, 1780.

 doi:10.20944/preprints202411.0954.v1

29. Bauer, B.; Levaillant, C. A new set of generators and a physical interpretation for the $SU(3)$ finite subgroup $D(9,1,1;2,1,1)$. *Quantum Inf. Proc.* **2013**, *7*, 2509–2521.
30. Levaillant, C. On some projective unitary qutrit gates. *arXiv* **2013**, arXiv: 1401.0506 [math.QA].
31. Planat, M.; Aschheim, R.; Amaral, M. M.; Irwin, K. Informationally complete characters from quark and lepton mixings. *Symmetry* **2020**, *12*, 1000.
32. Cui, S. X; Qiu, Y.; Wang Z. From three dimensional manifolds to modular tensor categories. *Comm. Math. Phys.* **2023**, *397*, 1191–1235.
33. Boalch, P. From Klein to Painlevé via Fourierr, Laplace and Jimbo. *Proc. Lond. Math. Soc.* **2005**, *90*, 167-208.
34. Planat, M.; Chester, D. Irwin, K. Dynamics of Fricke-Painlevé surfaces. *Dynamics* **2024**, *4*, 1–13.
35. Deng, D. L.; Li, X.; Das Sarma, S. Machine learning topological states. *Phys. Rev. B* **2017**, *96*, 195145.
36. Sanborn, S.; Mathe, J.; Papillon, M.; Buracas, D.; Lillemark, H. J.; Showmake, C.; Bertics, A.; Pennec, X.; Miolane, N. Beyond Euclid: an illustrated guide to modern machine learning with geometric, topological, and algebraic structures. *arXiv* **2024**, arXiv: 2407.09468 [cs.LG].