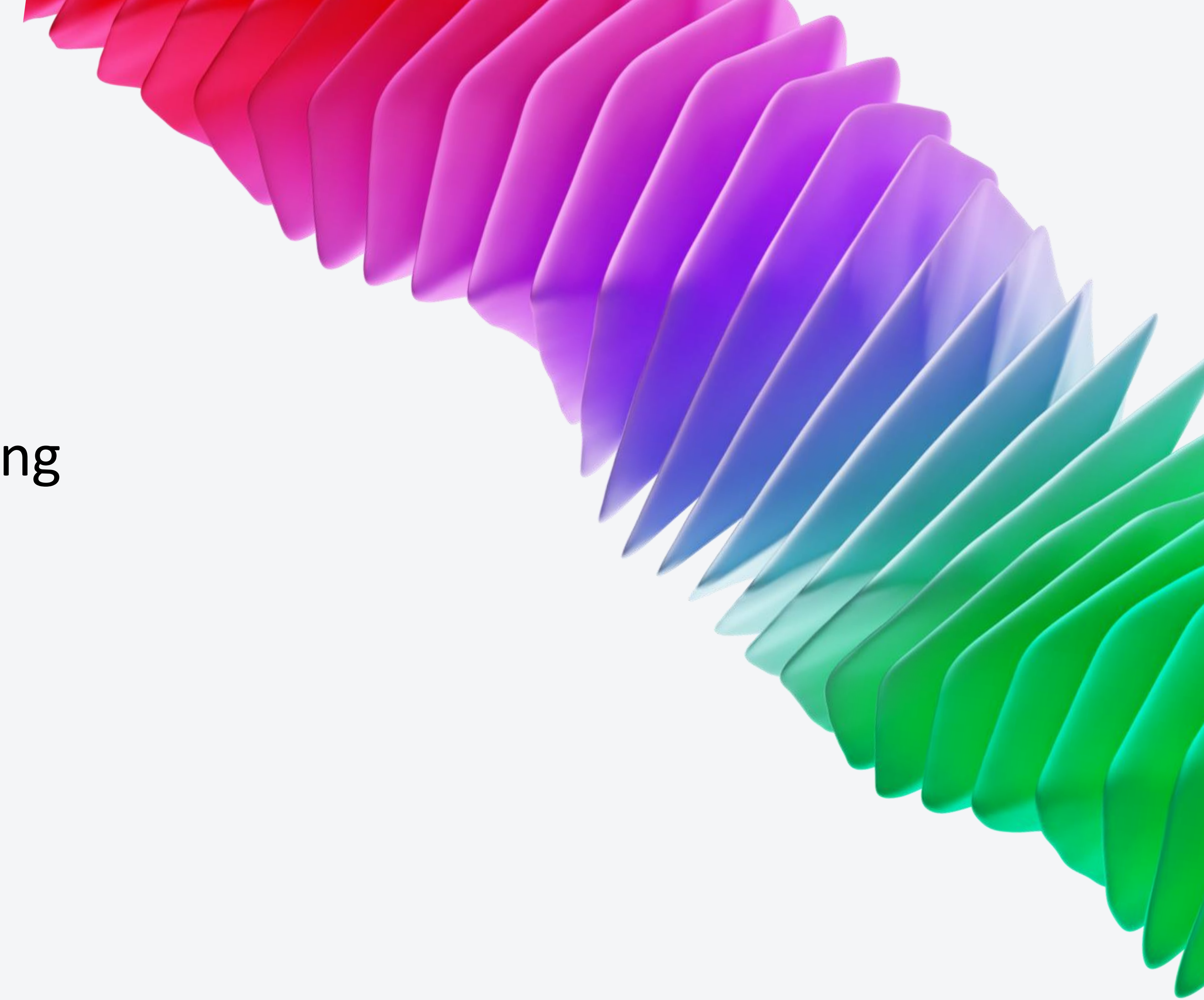
MULTIVERSE
COMPUTING

Towards Intelligent Quantum Computers

Borja Aizpurua Altuna

Multiverse Computing

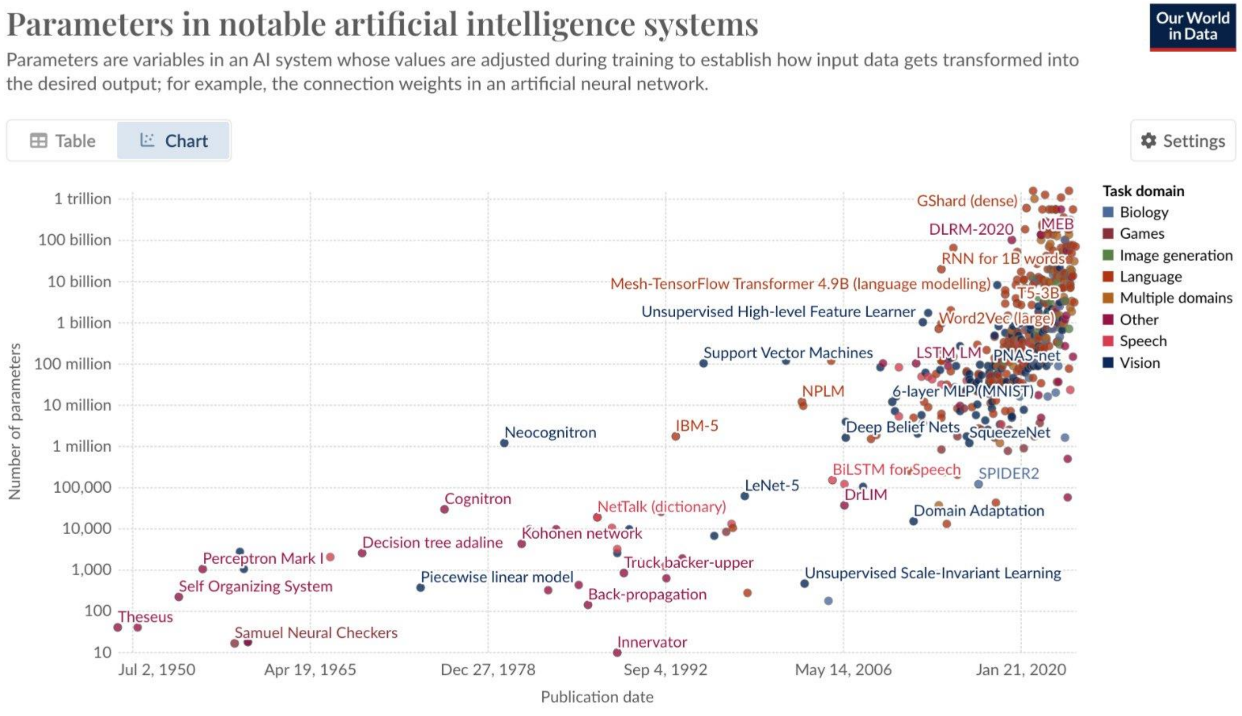


Our Journey

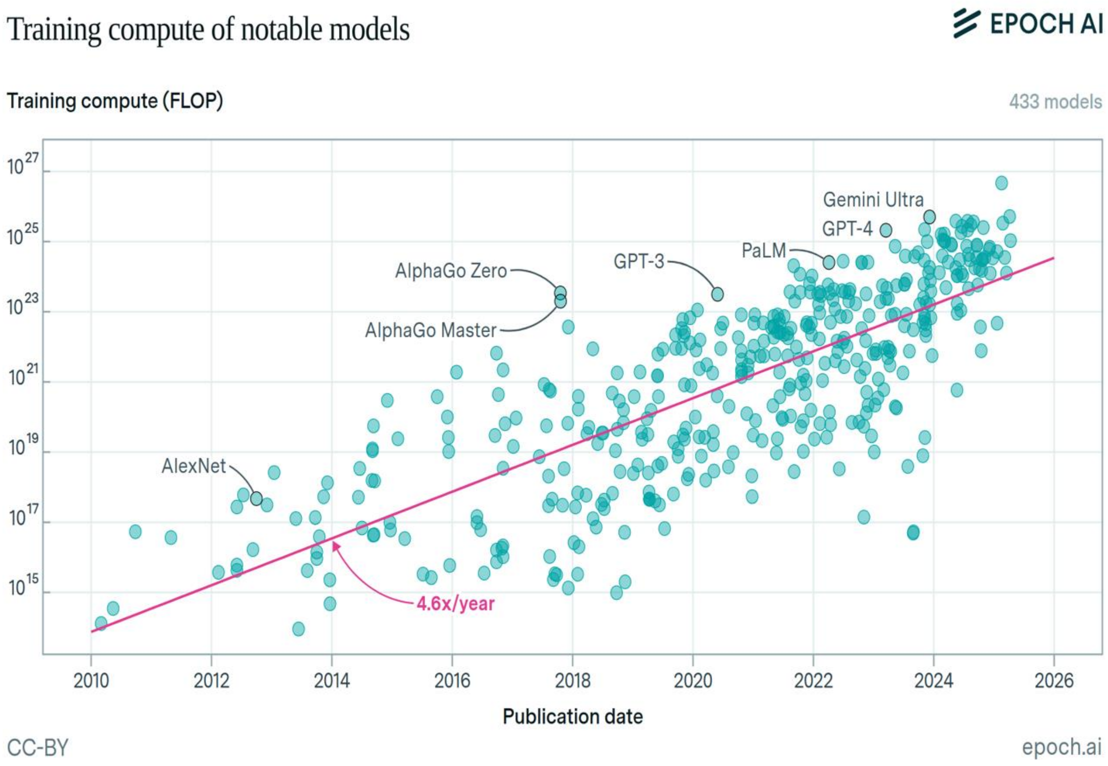
2019	2020	2021	2022	2023	2024–26
Founded San Sebastián Quantum for finance	**Singularity** BBVA, CaixaBank, Bank of Canada	**Beyond finance** Bosch, BASF, Repsol Gartner Cool Vendor	**Tensor networks** compress neural networks (vision)	**CompactifAI** LLM compression goes commercial	**$215M Series B** Compressed Llama, DeepSeek, HyperNova

Skyrocketing AI Compute Costs

Training compute of notable AI models is doubling roughly every five months. growing at a rate of **4.6x per year**,
While costs are doubling every nine months for the largest AI models, growing at a rate of **2.6x per year**



Source: Our World in Data (https://ourworldindata.org/)

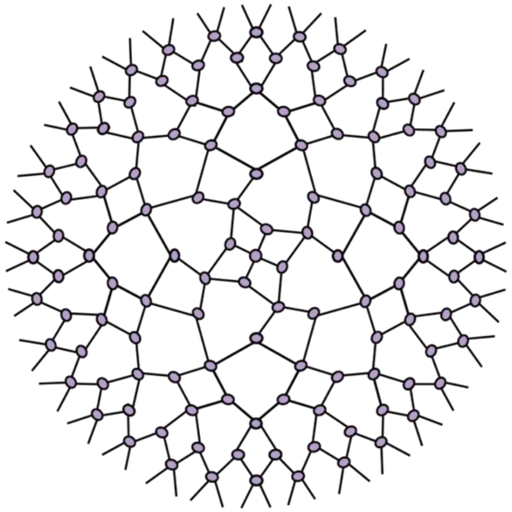


Robi Rahman and David Owen (2024), "The training compute of notable AI models is doubling roughly every five months". *Published online at epoch.ai.* Retrieved from: 'https://epoch.ai/data-insights/compute-trend-post-2010' [online resource]. Updated on June 4th , 2025

Tensor Networks – a Quantum-Inspired Approach

A functional toolset borrowed from Quantum Physics.

Allows us to understand which parameters and layers carry knowledge and which are redundant, by examining the correlation space across model parameters to extract where the information truly lies — and where it doesn't.



Real advantages and value **TODAY**

- **Efficient** – less parameters
- **Faster** – in training and inference
- **Scalable** – over uncompressed alternatives
- **Universal** – can compress all types of NN models

Our techniques produce models compatible with the main libraries and plugins in the market

Singularity Machine Learning

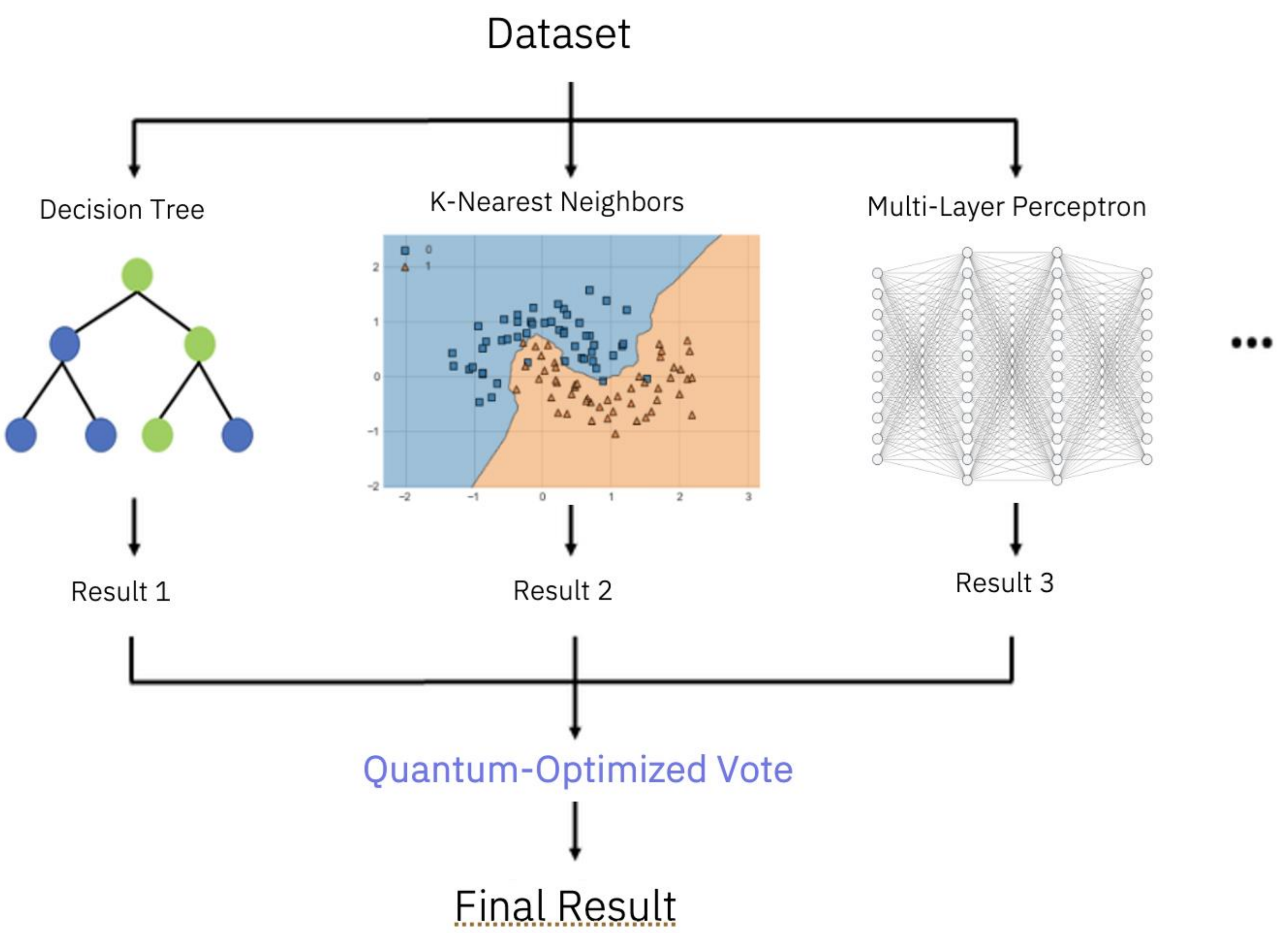
Singularity Machine Learning - Classification: A Qiskit Function by Multiverse Computing

Overview

With the "Singularity Machine Learning - Classification" function, you can solve real-world machine learning problems on quantum hardware without requiring quantum expertise. This Application function, based on ensemble methods, is a hybrid classifier. It leverages classical methods like boosting, bagging, and stacking for initial ensemble training. Subsequently, quantum algorithms such as variational quantum eigensolver (VQE) and quantum approximate optimization algorithm (QAOA) are employed to enhance the

The Classifier

- Ensemble classifier
  - Combines the predictions of various models (learners)
- Heterogeneous – can be built with different types of learners
  - Decision Trees, k-Nearest Neighbours, Multi-Layer Perceptrons + others
  - Wider perspective of the data landscape
- In *classical* ensemble models, the final result is usually obtain
  by *majority vote* of the learner predictions.
- Our Quantum approach:
  - We find the **optimal subset of learners** that yield the best results.
  - **NP-hard binary optimization problem** (need to assign 1 to learners that g
    to vote, 0 otherwise) → cannot be efficiently solved classically
  - Formulate as a **QUBO** and solve on IBM Quantum hardware → minimize th
    cost function $\mathcal{H}$.



$$\mathcal{H}(w) = \sum_{i,j} w_i w_j \mathrm{Corr}(h_i, h_j) + \sum_i w_i(\lambda - 2\mathrm{Corr}(h_i, y))$$

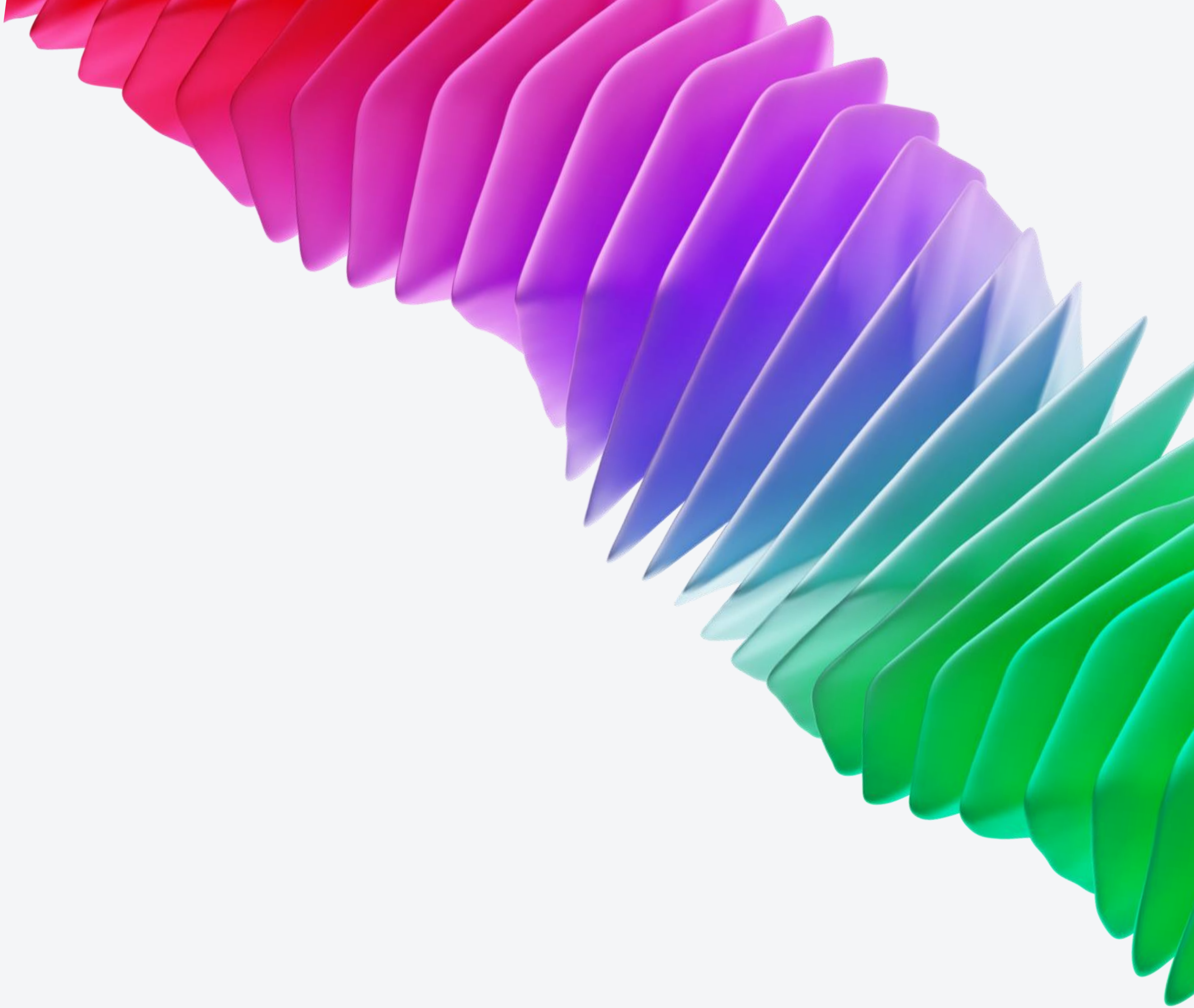
w_i - the binary weight of learner i

h_i - the prediction of learner i

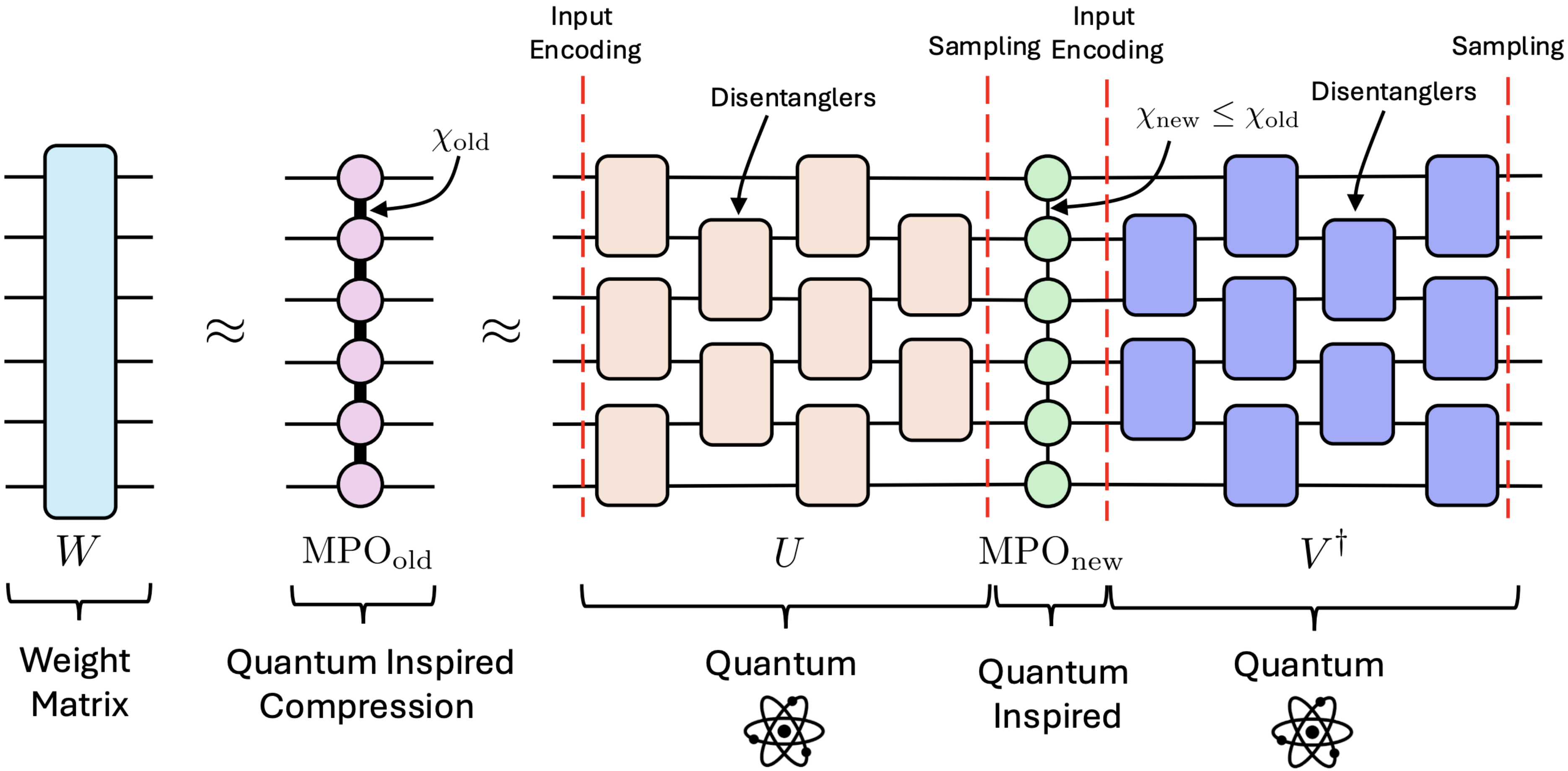
y - the classification labels

MULTIVERSE COMPUTING

Quantum LLM



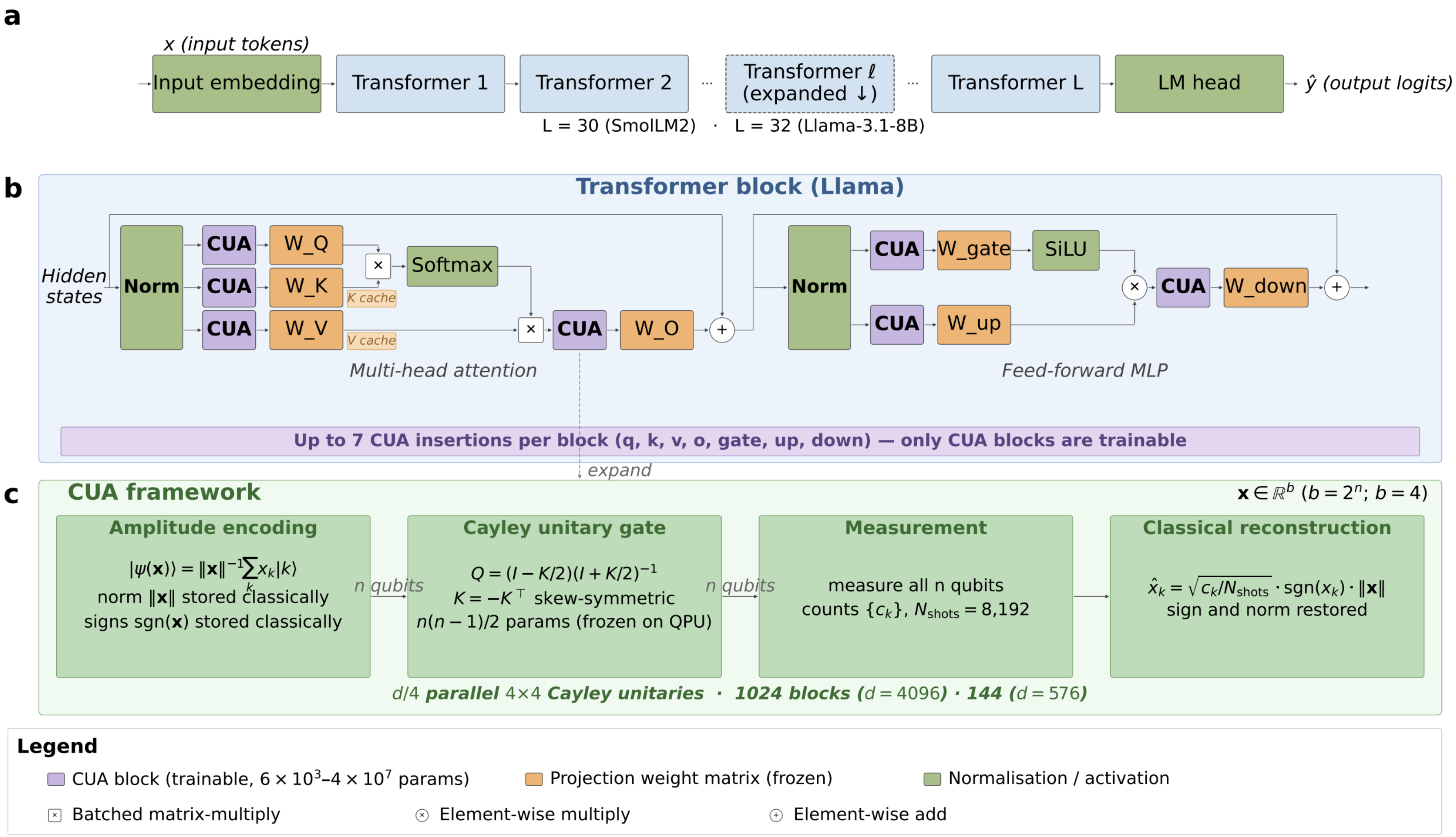
First Paper: Quantum Large Language Models via Tensor Network Disentanglers



MULTIVERSE COMPUTING

https://arxiv.org/pdf/2410.17397

Quantum-enhanced Large Language Models on Quantum Hardware via Cayley Unitary Adapters

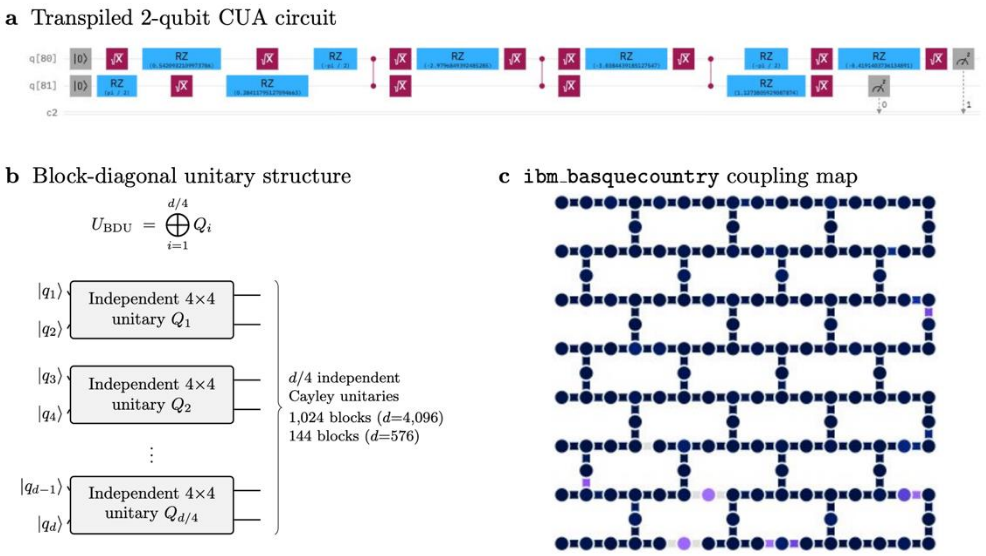


MULTIVERSE COMPUTING

https://arxiv.org/abs/2605.05914

Quantum-enhanced Large Language Models on Quantum Hardware via Cayley Unitary Adapters

- We added a quantum circuit inside Llama3.1-8B and run inference on ibm_basquecountry (Donostia’s IBM System Two)
- The circuit is trained classically, only inference on quantum
- Some questions that were wrong in original model are right in the quantum inference
- We did multiple scaling/noise analysis



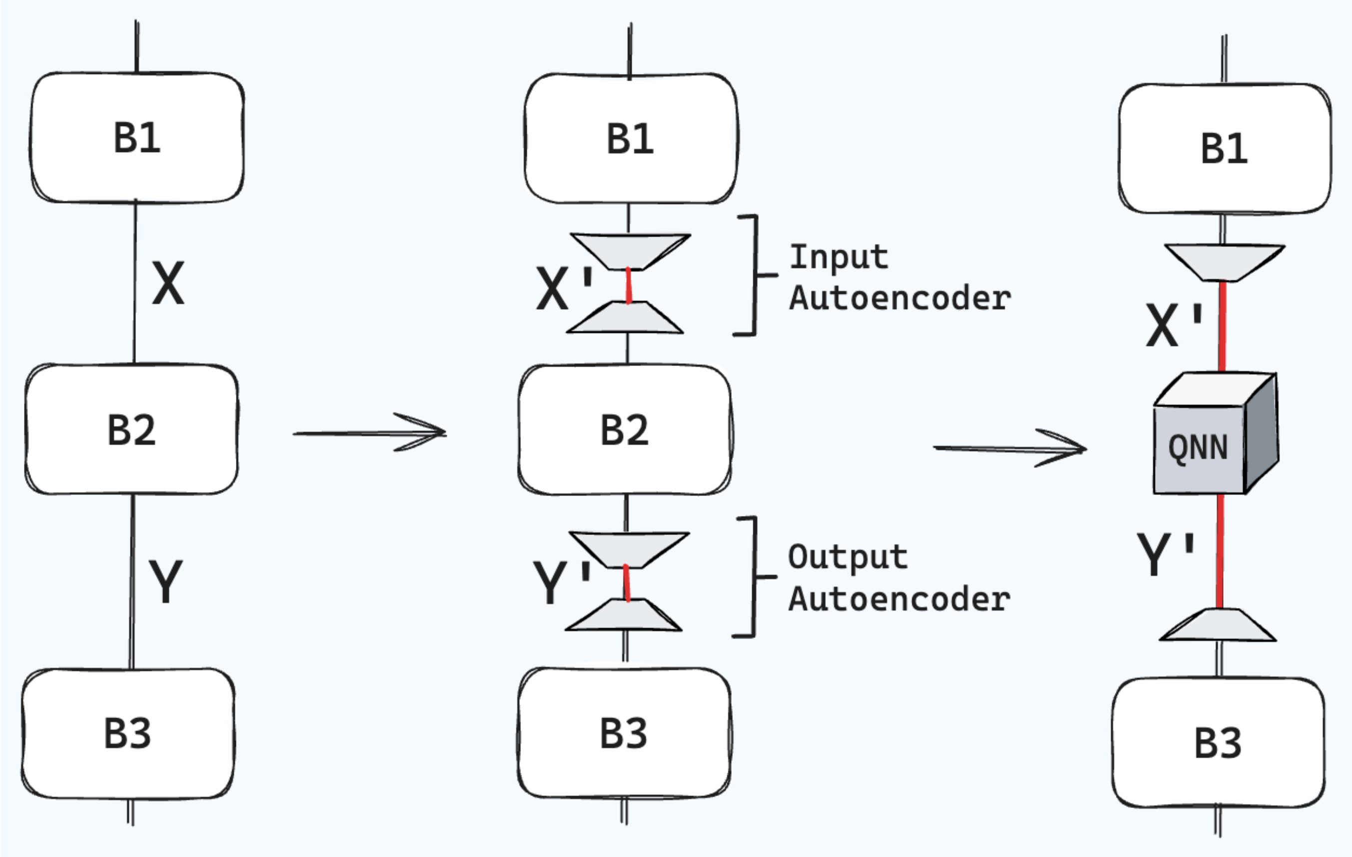
MULTIVERSE COMPUTING

https://arxiv.org/abs/2605.05914

Replacing an entire Transformer block with AE + QNN

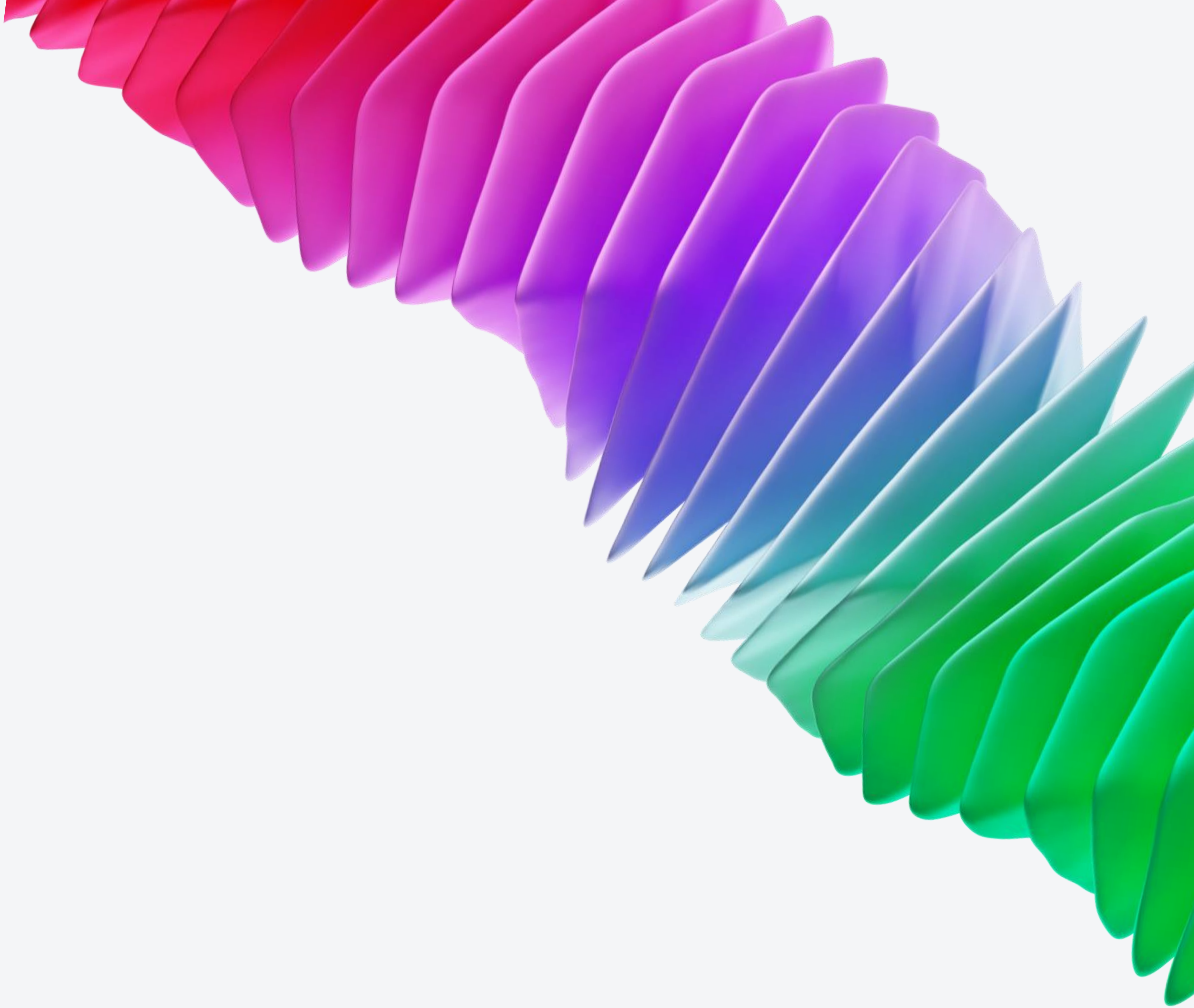
- Replace an entire block by autoencoder + QNN
- Original PPL 35.31
- Original - block PPL 41.8
- Original - block + AE + QNN

PPL 37.6 and ongoing…



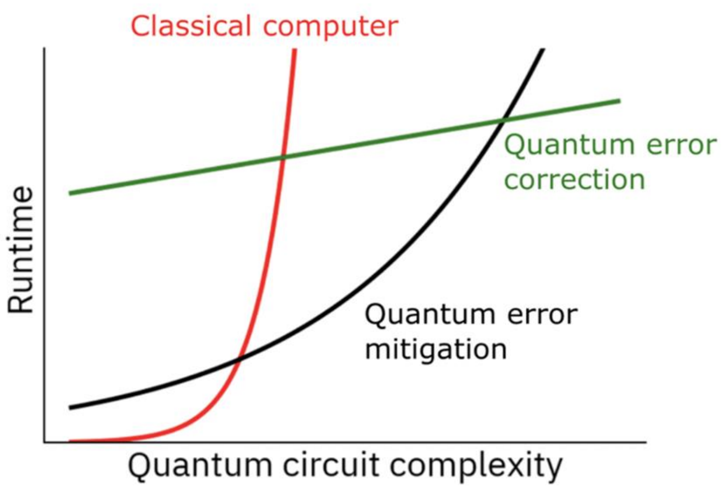
MULTIVERSE
COMPUTING

AI for Quantum



Definition of Quantum Error Mitigation

Algorithmic schemes that reduce the bias in expectation values caused by noise in the hardware by using additional circuit executions and classical post-processing, without requiring full quantum error correction



Qin, D., Xu, X., & Li, Y. (2022). An overview of quantum error mitigation formulas. *Chinese Physics B*, *31*(9), 090306. https://doi.org/10.1088/1674-1056/ac7b1e

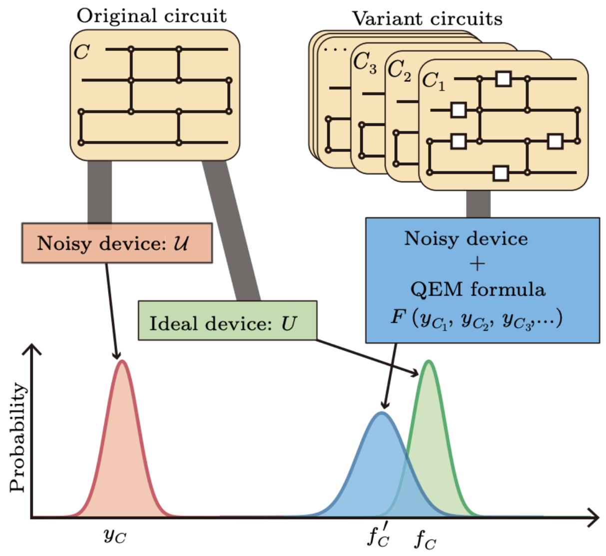
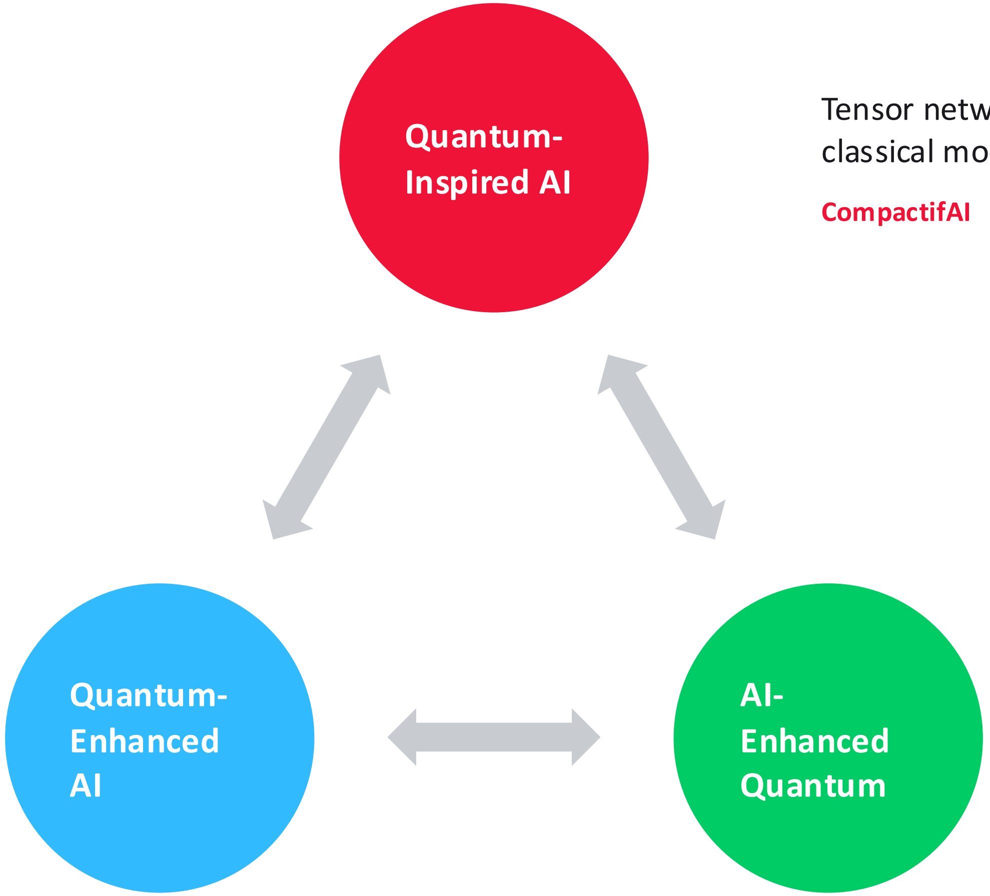


Fig. 1. Schematic diagram of quantum computing with and without QEM. The output of the original circuit is f_C on an ideal quantum device and y_C on a noisy device. Quantum error mitigation performs computations on a set of variant circuits to predict the error-free result with an increased accuracy. A general cost of quantum error mitigation is the enlarged variance.

Idea: train an AI model to learn the difference between the original circuit and the circuit + error mitigation's outputs. Given a noisy circuit, predict the output of the error mitigated circuit

MULTIVERSE COMPUTING

Towards Intelligent Quantum Computers



Tensor networks compress classical models

CompactifAI

Quantum hardware lifts machine learning

Singularity · Quantum LLM

Classical AI corrects and controls qubits

Error Mitigation / Correction

The substrate is silicon. The qubits are fabricated objects. This convergence is built on the work of the people in this room.

The Real Question

MULTIVERSE
COMPUTING

Not *when* will quantum computers become **intelligent** — but *who* will build them.