Deep quantum graph dreaming: deciphering neural network insights into quantum experiments

Tareq Jaouni
Nexus for Quantum Technologies, University of Ottawa
K1N 6N5, ON, Ottawa, Canada
tjaou104@uottawa.ca

Sören Arlt
Max Planck Institute for the Science of Light
Staudtstrasse 2, 91058 Erlangen, Germany

Carlos Ruiz-Gonzalez
Max Planck Institute for the Science of Light
Staudtstrasse 2, 91058 Erlangen, Germany

Ebrahim Karimi
Nexus for Quantum Technologies, University of Ottawa
K1N 6N5, ON, Ottawa, Canada

Xuemei Gu
Max Planck Institute for the Science of Light
Staudtstrasse 2, 91058 Erlangen, Germany
xuemei.gu@mpl.mpg.de

Mario Krenn
Max Planck Institute for the Science of Light
Staudtstrasse 2, 91058 Erlangen, Germany
mario.krenn@mpl.mpg.de

Abstract

Neural networks hold great promise for advancing scientific discoveries, but their opaque nature often makes it challenging to interpret the underlying logic behind their findings. In this work, we employ an eXplainable-AI technique known as *inception* or *deep dreaming*, originally developed in the context of computer vision, to investigate what neural networks learn about quantum optics experiments. We begin by training deep neural networks on the properties of quantum systems. Once trained, we 'invert' the neural network – essentially asking it to imagine a quantum system with specific properties and to continuously modify the system to change those properties. We find that the network can shift the initial distribution of properties in the quantum system, allowing us to conceptualize the strategies it has learned. Interestingly, the network's initial layers focus on identifying simple properties, while the deeper layers uncover complex quantum structures. This reflects well-known patterns observed in computer vision, which we now identify within the context of a complex natural science task. Our approach paves the way for more interpretable AI scientific discovery techniques in quantum physics.

1 Introduction

Neural networks have demonstrated significant promise in solving various tasks in quantum science [1, 2, 3]. However, a notable frustration with neural networks lies in their inscrutability: modern architectures often contain millions of trainable parameters, making it difficult to discern the role each plays in the network's predictions. Understanding the concepts learned by the network to formulate its predictions is crucial for achieving scientific insight [4]. This challenge has driven the development of eXplainable-AI (XAI), which interprets how networks arrive at their solutions [5, 6, 7, 8]. These

Interpretable AI: Past, Present and Future Workshop at NeurIPS 2024

advancements have encouraged physicists to address interpretability, leading to the rediscovery of long-standing physics concepts [9, 10], identification of phase transitions in quantum many-body physics [11, 12, 13, 14], compression of many-body quantum systems [15], and studies on the relationship between quantum systems and their entanglement properties [16, 17].

Our work focuses on applying neural networks to the design of quantum optical experiments. The increasing complexity of quantum information tasks has motivated the development of computational methods capable of navigating the vast combinatorial space of possible experimental designs, which often involve unintuitive phenomena [18]. To this end, scientists have developed automated design and machine learning routines [19], including those that leverage genetic algorithms [20, 21], active learning approaches [22], and optimization of parameterized quantum circuits [23, 24, 25]. One may ask whether new physics can be learned from the discoveries made by these algorithms. For instance, the computer algorithm MELVIN [19], which topologically searches for arrangements of optical elements, has led to discoveries such as the generation of entanglement by path identity [26] and the creation of multipartite quantum gates [27]. However, the interpretability of these solutions is often obscured by the stochasticity of the processes that create them and the unintuitive nature of their representations. The recent invention of THESEUS [24] and its successor PYTHEUS [25] address this by topologically optimizing highly interpretable, graph-based representations of quantum optical experiments. This has already enabled new scientific discoveries, such as a new form of multi-photon interference [28] and novel experimental schemes for high-dimensional quantum measurement [29].

To date, the extraction and generalization of new concepts have largely been confined to analyzing the optimal solutions discovered by these algorithms. However, we can potentially learn more physics by probing the rationale behind the computer's discoveries. Little attention has been given to applying XAI techniques to neural networks trained on quantum experiments, which could allow us to conceptualize what our algorithm has learned. In doing so, we could guide the creation of AI-based design techniques for quantum experiments that are more reliable and interpretable.

We introduce an interpretability tool based on the inceptionism technique from computer vision, better known as Deep Dreaming [30]. This technique has been used to iteratively guide the automated design of quantum circuits [31] and molecules [32] towards optimizing a target property; it has also been used in [33] to verify the reliability of a network trained to classify the entanglement spectra of many-body quantum systems. More importantly, it allows us to visualize the physical insights the neural network has gained from the training data. This enables better discernment of the strategies applied during automated design processes and verification of physical concepts rediscovered by the network, such as the thermodynamic arrow of time [34]. We apply this approach to quantum graphs by training a deep neural network to predict properties of quantum systems and then inverting the network to optimize for specific target properties. This inverse training dramatically shifts the initial distribution of properties. Visualizing the evolution of quantum graphs during this process allows us to conceptualize the strategies the neural network has learned. Examining the network's intermediate layers reveals that it initially learns simple features and progressively constructs more complex structures. This comprehensive understanding of the network's perception aids in designing more interpretable and reliable computer-assisted schemes for quantum optics experiments

2 Methodology

2.1 The graph representation for quantum optics experiments

Recent studies have shown that many quantum optics experiments (involving nonlinear pair-sources, single photon sources, linear optics, etc.) can be abstractly represented as colored, edge-weighted graphs [35, 36, 37, 24, 25]. This representation can be extended to integrated photonics [38, 39, 40, 41] and entanglement by path identity [26, 42, 43]. In the graph, vertices represent photon paths to detectors, and edges between any two vertices, a and b, indicate correlation between these paths. Edge weights $\omega_{a,b}$ denote amplitudes, while edge colors represent the photons' internal mode numbers. Each vertex inherits its color from the connected edges, defining the photon's state. As an example in Figure 1, we consider graph representations of four-qubit, two-dimensional experiments focused on state creation. Specifically, we consider graphs with vertices $V = \{0, 1, 2, 3\}$ and mode numbers 0 and 1, represented by blue and red edge colorings, respectively. Each graph, thus, consists of 24 possible edges with real-valued edge weights between 1 and -1. The quantum state $|\Phi(\omega)\rangle$ is

determined by the graph's weight function according to Eq. (2) in [25], which is

$$\Phi(\boldsymbol{\omega}) = \sum_m \frac{1}{m!} \left(\sum_{e \in E(G)} \omega(e) x^\dagger(e) y^\dagger(e) + \text{h.c.} \right)^m , \tag{1}$$

where $E(G)$ is the set of edges of the graph G, $x^\dagger(e)$ and $y^\dagger(e)$ are creation operators of photons, which are represented as vertices x and y of edge e, and $h.c.$ denotes the Hermitian conjugate, including annihilation terms. The quantum state can then be physically realized by applying the weight function to the vacuum state: $|\Phi(\omega)\rangle = \Phi(\omega)|vac\rangle$. Overall, the neural network finds a way to decompose the quantum state into perfect matchings of a graph, which is useful because arbitrary graphs can be experimentally implemented in the laboratory, with the quantum state emerging as a coherent superposition of these perfect matchings.

2.2 Training

Figure 2 illustrates the basic workflow of the dreaming process. A feed-forward neural network is first trained on the edge weights ω of a complete, quadripartite, two-dimensional quantum graph to predict certain properties of the corresponding quantum state $|\Phi(\omega)\rangle$. The edge weights are randomly initialized over a uniform distribution $[-1, 1]$. The neural network's weights and biases are optimized for this task using mini-batch gradient descent and the mean squared error (MSE) loss function. We evaluate the state fidelity $|\langle\Phi(\omega)\,|\psi\rangle|^2$ with respect to two multipartite entangled states $|\psi\rangle$, such as including the Greenberger-Horne-Zeilinger (GHZ) state [44], where 4-qubit GHZ is $|\text{GHZ}\rangle = (|0000\rangle + |1111\rangle)/\sqrt{2}$.

We generated 20 million input-output pairs using PyTheus [25], where each input is a 24-dimensional array of edge weights for a quantum graph, and the output is the property value of the corresponding state, $|\Phi(\omega)\rangle$. Networks were trained with a mini-batch size of 5000 and a 95:5 train-test split. Learning rates started at 1×10^{-3} (or 1×10^{-5} for the $[36^{26}]$ training architecture) and decreased by 0.95 if the test MSE did not improve after 25 epochs, continuing until convergence was stable for over 400 epochs. ReLU was used in all hidden layers except $[36^{26}]$, which used ELU with $\alpha = 0.1$. PyTorch [45] and Adam [46] were used for training. A hyperparameter search was carried out on the number of neurons, N, in the generic neural network architecture $[N^4]$ towards predicting the GHZ-State fidelity. The hyperparameter search stopped once satisfactory improvements in the test MSE with respect to the simplest model considered were attained, which was achieved with $N = 400$ neurons. For all cases considered, the network is trained on examples with a property value below a threshold of 0.5 to ensure that the network is not memorizing the best solutions in each case.

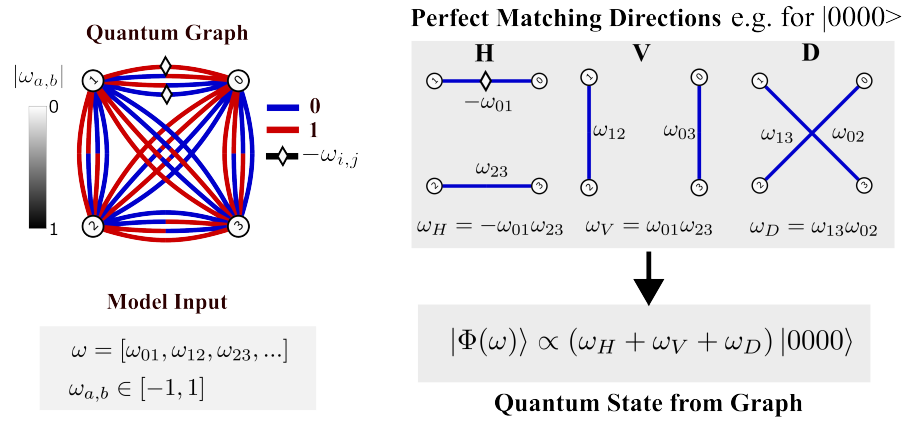

Figure 1: Overview of quantum graphs. We show a complete, two-dimensional, quadripartite quantum graph, where edge weights $\omega_{a,b}$ connect vertices a and b. Edge transparency indicates weight magnitude, and a diamond denotes negative weights. The quantum state $|\Phi(\omega)\rangle$ is formed by the coherent superposition of all perfect matchings, distinguished by direction.

This threshold remains fixed for cases involving the GHZ state fidelities. During this generation of the training data set, if examples are beyond the threshold, they are rejected.

Once training convergence is achieved, we execute the deep dreaming protocol to extract insights from the neural network. For an arbitrary input graph, we select a neuron and maximize its activation by updating the input graph via gradient ascent, while keeping the network's weights and biases frozen. The loss, defined as the negative of the neuron's activation, is calculated by evaluating the network's prediction with the intermediate input graph. Through this process, the graph evolves into a configuration that most excites the neuron. However, since neurons can recognize multiple features [47], we repeat the procedure with different input graphs to uncover all features the neuron detects. Inverse training was performed on 1 million input graphs over 100,000 iterations with a learning rate of 1×10^{-4}. This took approximately 5 minutes per graph for fidelity tasks and 4 minutes for concurrence minimization, with 15,000 iterations used for dreaming. Training was conducted on an AMD Ryzen 5 4500U @ 2.38 GHz CPU. Faster training can be achieved by increasing the learning rate and applying early stopping, as seen with GHZ fidelity maximization in 5000 iterations at a learning rate of 1×10^{-3}.

3 Results

3.1 Dreaming on the output layer

To gain insight into what the neural network has learned about the quantum state $|\Phi(\omega)\rangle$, we first apply the deep dreaming approach to the output layer. Figure 3(a) shows how an input graph mutates when deep dreaming is applied to a $[400^3,10]$ (three hidden layers of 400 neurons, one hidden layer of 10 neurons) neural network trained to predict GHZ-state fidelity. During dreaming, the network searches for configurations that maximize the property value. Notably, the optimal configuration that maximizes the GHZ-state fidelity.

We obtain $|\Phi(\omega)\rangle$ from the reconstructed, mutated graph and recompute its true property value in each step. In all cases, the graph steadily evolves toward the maximum property value. We repeat this procedure for 1000 different quantum graphs and plot the distribution of each graph's initial versus dreamed fidelities in Figure 3(b). The network consistently discovers distinct examples with property values exceeding the initial distribution's upper bounds, demonstrating the potential of our approach to uncover novel quantum graphs that optimize specific quantum state properties.

The intermediate steps of the dreaming process reveal the strategies the neural networks employ for a given optimization task. Figure 3(c) shows the evolution of different initial graphs during inverse training for various targets. We see that the neural network tries to activate the $|0000\rangle$ and $|1111\rangle$

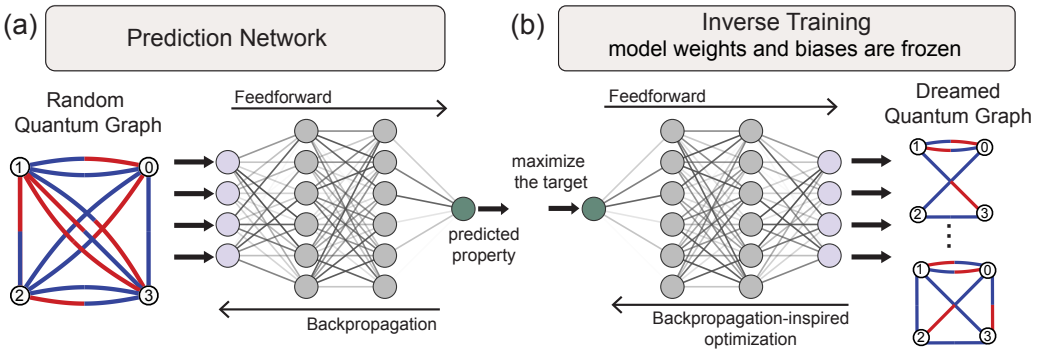

Figure 2: Quantum Graph Deep Dreaming. (a). During training, the weights and biases of a feed-forward neural network are continually updated to predict properties, such as the fidelity of a random quantum experiment represented by a graph. (b). In the deep dreaming process, the network's weights and biases are frozen, and an initial input graph is iteratively adjusted to maximize the activation of the output neuron, which predicts the specified property. The final "dreamed" graph is expected to maximize the network's output activation when fed back into the neural network.

states either by creating perfect matchings (PM) of these terms in unused directions – the input graph had no PM in that direction previously – or by completing them with the assistance of an existing PM in some direction, as is seen in particular with the $|\Phi(\omega)\rangle = |0011\rangle + |0101\rangle$ initialization. The dreaming process adjusts these PMs so their weights sum to 1. If the initial graph contains unwanted terms or if the network inadvertently creates them during dreaming, the network attempts to eliminate them by reducing the edge weights' magnitudes or introducing negative-weight PMs in different directions.

Our deep dreaming approach demonstrates that the network learns how to create states through graph representations to consistently achieve optimal values for specific quantum state properties. Notably, the network identified configurations that maximize these properties despite being trained only on configurations with values below 0.50. This suggests that the network is relying on physical insights rather than merely memorizing the best examples.

3.2 Interpretability of neural network structure

We apply the deep dreaming approach to the neurons in the hidden layers to gain insight into the neural network's internal model, which generalizes well beyond the training data. Figure 4 summarizes the insights obtained through this process. To demonstrate the universality of our approach, we consider several neural network architectures – $[400^4]$, $[49^{10}]$ and $[36^{26}]$ – each trained to predict GHZ-state fidelity. For each network, we dream on the i^{th} neuron in the j^{th} hidden layer using 20 input graphs to capture all possible structures that excite the neuron.

We focus on how the complexity of dreamed graphs evolves with network depth. The greatest insight into our quantum graphs is gained by analyzing the different ways a ket is realized through the graphs' PMs. To each dreamed graph, we assign 3×16 array, $p_{i,j}$, representing the probabilities of all possible PMs, which provides insight into the state created by the graph and the PM directions utilized. As we move deeper into the neural network, the dreamed graphs activate more PM directions and kets, reflecting the increasing complexity of structures the network recognizes. We also observe the multifaceted nature of the neurons: different input graphs result in dreamed graphs that recreate different input states. For example, as shown in the third inset of Figure 4(a), a neuron may focus on parts of the graph that best create the $|0000\rangle$ term or may interpret different PM directions for $|0000\rangle$ or parts of the graph that realize the $|1111\rangle$ term.

We may quantify the complexity of structures recognized throughout the network with the information entropy $H_{i,j}$. By averaging $p_{i,j}$ across all dreamed graphs, we compute $H_{i,j}$ for each neuron and

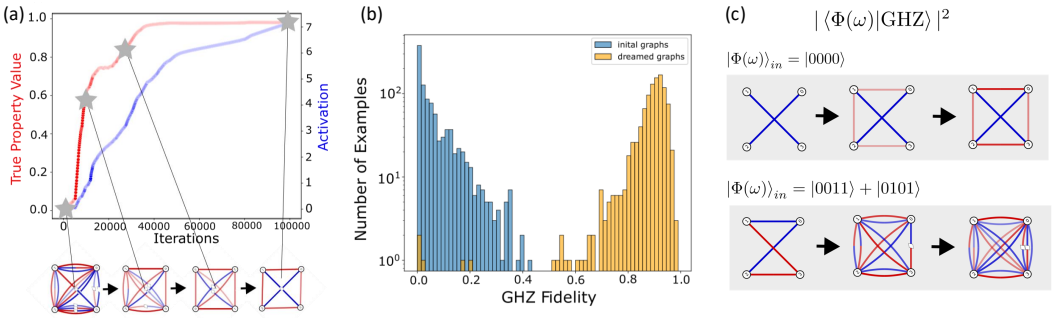

Figure 3: Results. (a), We show the evolution of an input graph's fidelity respect to the GHZ state, during the dreaming on the neural network ($[400^3, 10]$). Intermediate steps of a random graph's evolution to its dreamed counterpart are displayed. For each case, we show the intermediate steps of the input graphs' evolution to its dreamed counterpart and only show edges whose weights are above a threshold of 0.3. (b), Distribution of initial vs. dreamed fidelity respect to GHZ state. No trained initial graph has fidelity above 0.5. (c), Strategies from the inverse training are explored, using specified inputs to optimize GHZ state fidelity. Disjoint yellow-cycle graphs are generated for input $|0000\rangle$, and, edge weights are modified for the diagonal perfect matching, and new perfect matchings are created with specific weight contributions.

then determine the average entropy for the j^{th} layer, providing a general metric of the complexity of recognized structures. Figure 4(b) shows the trend of $\overline{H_{i,j}}$ across all three neural network architectures. As expected, the deep neural network initially learns to recognize simple structures, with entropy dropping to its lowest values in the early layers, then gradually increasing as more abstract features are recognized in deeper layers. This pattern confirms that the network first identifies simple features, such as edges forming one or two PMs, before forming more complex graphical structures in deeper layers that involve a greater set of PMs.

4 Outlook

We present preliminary results on adapting the deep dreaming approach to quantum optical graphs using deep neural networks for various target quantities. Our routine reveals the strategies used by the neural network by dreaming on both the output layer and hidden layers. Notably, we demonstrate that the trained neural network constructs a non-trivial model of quantum state properties and that deep dreaming effectively identifies novel examples beyond the initial dataset. Additionally,

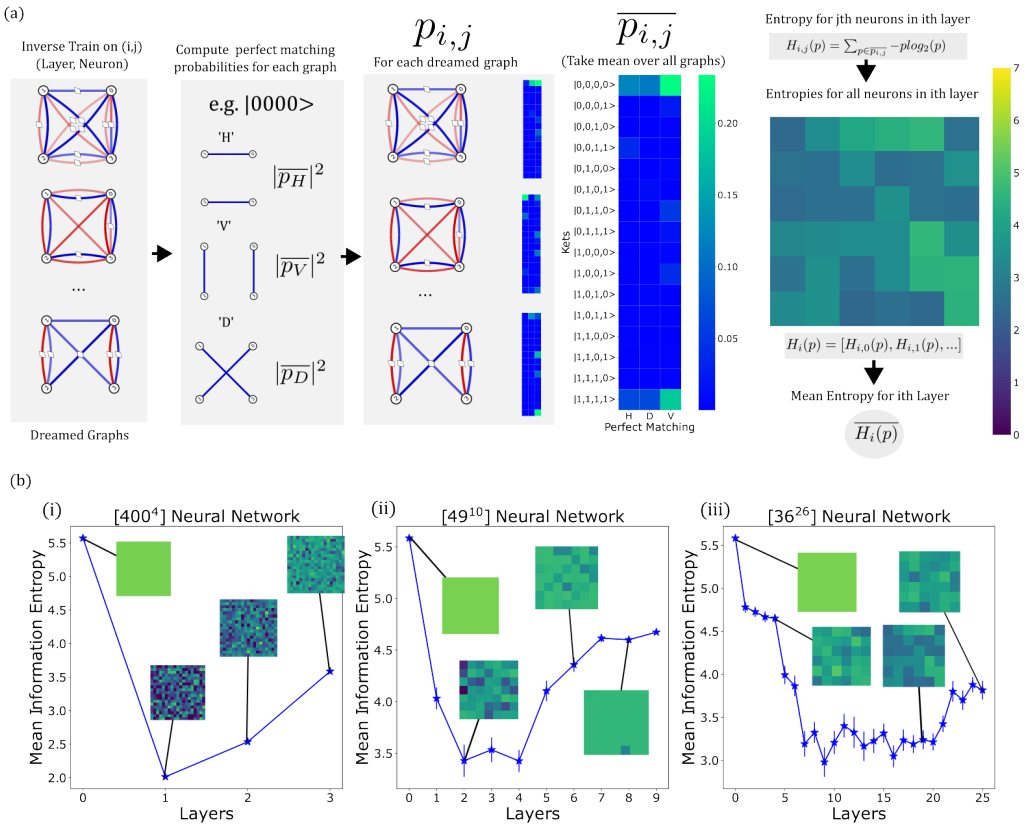

Figure 4: Information Entropy Across Neural Network Architectures. (a), Workflow behind computing the mean information entropy for each layer of a trained neural network. We perform deep dreaming on multiple input graphs for each neuron, calculating the mean probability amplitudes for all perfect matchings corresponding to each ket. This reveals the overall graph structure that a neuron most strongly recognizes. We then compute the information entropy for each neuron, $H_{i,j}(p)$ and the mean entropy for the layer $\overline{H_{i,j}(p)}$ to measure the complexity of structures identified by the network. The variety in $p_{i,j}$ for each dreamed graph highlights the multifaceted nature of the neurons. (b), Mean information entropy plots for the (i) $[400^4]$ (ii) $[49^{10}]$ and (iii) $[36^{26}]$ neural network architectures. A common trend across all cases shows that mean entropy is minimal in the lower layers and gradually increases in deeper layers, reflecting the network's progression from recognizing simpler structures to more complex ones.

applying this approach to the network's hidden layers shows that the network progressively learns to recognize increasingly complex structures, with individual neurons being multifaceted in the structures that excite them. Future work could enhance the transparency of learned representations by using regularization techniques such as α-norm [48], jitter [30], or dreaming on the mean of multiple input graphs [47]. Further insights might also be gained by directly modifying the network's weights and biases. Moreover, applying these tools to larger graphs and exploring applications beyond state creation, such as quantum measurements and quantum communication, will be valuable.

Quadripartite graphs have served as an effective test case for our approach, and the knowledge gained can be applied to other systems. Larger graphs and new targets will offer deeper insights into quantum optics experiments and inspire further research. We anticipate that this approach could extend frameworks for automated setup design [25, 19, 4] and generative molecular algorithms [32, 49], helping to decode the underlying science and strategies toward target configurations.

References

[1] Anna Dawid, Julian Arnold, Borja Requena, Alexander Gresch, Marcin Płodzień, Kaelan Donatella, Kim A Nicoli, Paolo Stornati, Rouven Koch, Miriam Büttner, et al. Modern applications of machine learning in quantum sciences. *arXiv:2204.04198*, 2022.

[2] Mario Krenn, Jonas Landgraf, Thomas Foesel, and Florian Marquardt. Artificial intelligence and machine learning for quantum technologies. *Phys. Rev. A*, 107:010101, Jan 2023.

[3] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019.

[4] Mario Krenn, Robert Pollice, Si Yue Guo, Matteo Aldeghi, Alba Cervera-Lierta, Pascal Friederich, Gabriel dos Passos Gomes, Florian Häse, Adrian Jinich, AkshatKumar Nigam, Zhenpeng Yao, and Alán Aspuru-Guzik. On scientific understanding with artificial intelligence. *Nature Reviews Physics*, 4(12):761–769, oct 2022.

[5] Derek Doran, Sarah Schulz, and Tarek R Besold. What does explainable ai really mean? a new conceptualization of perspectives. *arXiv:1710.00794*, 2017.

[6] Erico Tjoa and Cuntai Guan. A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11):4793–4813, nov 2021.

[7] Nadia Burkart and Marco F. Huber. A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70:245–317, jan 2021.

[8] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv:1708.08296*, 2017.

[9] Raban Iten, Tony Metger, Henrik Wilming, Lidia del Rio, and Renato Renner. Discovering physical concepts with neural networks. *Physical Review Letters*, 124(1):010508, jan 2020.

[10] Sebastian J. Wetzel, Roger G. Melko, Joseph Scott, Maysum Panju, and Vijay Ganesh. Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks. *Phys. Rev. Res.*, 2:033499, Sep 2020.

[11] Anna Dawid, Patrick Huembeli, Michał Tomza, Maciej Lewenstein, and Alexandre Dauphin. Hessian-based toolbox for reliable and interpretable machine learning in physics. *Machine Learning: Science and Technology*, 3(1):015002, nov 2021.

[12] Anna Dawid, Patrick Huembeli, Michal Tomza, Maciej Lewenstein, and Alexandre Dauphin. Phase detection with neural networks: interpreting the black box. *New Journal of Physics*, 22(11):115001, nov 2020.

[13] Niklas Käming, Anna Dawid, Korbinian Kottmann, Maciej Lewenstein, Klaus Sengstock, Alexandre Dauphin, and Christof Weitenberg. Unsupervised machine learning of topological phase transitions from experimental data. *Machine Learning: Science and Technology*, 2(3):035037, jul 2021.

[14] Sebastian J Wetzel and Manuel Scherzer. Machine learning of explicit order parameters: From the ising model to su (2) lattice gauge theory. *Physical Review B*, 96(18):184410, 2017.

[15] Andrea Rocchetto, Edward Grant, Sergii Strelchuk, Giuseppe Carleo, and Simone Severini. Learning hard quantum distributions with variational autoencoders. *npj Quantum Information*, 4(1):28, 2018.

[16] Daniel Flam-Shepherd, Tony C. Wu, Xuemei Gu, Alba Cervera-Lierta, Mario Krenn, and Alán Aspuru-Guzik. Learning interpretable representations of entanglement in quantum optics experiments using deep generative models. *Nature Machine Intelligence*, 4(6):544–554, jun 2022.

[17] Felix Frohnert and Evert van Nieuwenburg. Explainable representation learning of small quantum states. *arXiv:2306.05694*, 2023.

[18] Mario Krenn, Manuel Erhard, and Anton Zeilinger. Computer-inspired quantum experiments. *Nature Reviews Physics*, 2(11):649–661, sep 2020.

[19] Mario Krenn, Mehul Malik, Robert Fickler, Radek Lapkiewicz, and Anton Zeilinger. Automated search for new quantum experiments. *Physical review letters*, 116(9):090405, 2016.

[20] Rosanna Nichols, Lana Mineh, Jesús Rubio, Jonathan C F Matthews, and Paul A Knott. Designing quantum experiments with a genetic algorithm. *Quantum Science and Technology*, 4(4):045012, oct 2019.

[21] L O'Driscoll, Rosanna Nichols, and Paul A Knott. A hybrid machine learning algorithm for designing quantum experiments. *Quantum Machine Intelligence*, 1:5–15, 2019.

[22] Xavier Valcarce, Pavel Sekatski, Elie Gouzien, Alexey Melnikov, and Nicolas Sangouard. Automated design of quantum-optical experiments for device-independent quantum key distribution. *Physical Review A*, 107(6):062607, 2023.

[23] Juan Miguel Arrazola, Thomas R Bromley, Josh Izaac, Casey R Myers, Kamil Brá dler, and Nathan Killoran. Machine learning method for state preparation and gate synthesis on photonic quantum computers. *Quantum Science and Technology*, 4(2):024004, jan 2019.

[24] Mario Krenn, Jakob S. Kottmann, Nora Tischler, and Alán Aspuru-Guzik. Conceptual Understanding through Efficient Automated Design of Quantum Optical Experiments. *Physical Review X*, 11(3):031044, aug 2021.

[25] Carlos Ruiz-Gonzalez, Sören Arlt, Jan Petermann, Sharareh Sayyad, Tareq Jaouni, Ebrahim Karimi, Nora Tischler, Xuemei Gu, and Mario Krenn. Digital discovery of 100 diverse quantum experiments with pytheus. *arXiv:2210.09980*, 2022.

[26] Mario Krenn, Armin Hochrainer, Mayukh Lahiri, and Anton Zeilinger. Entanglement by path identity. *Phys. Rev. Lett.*, 118:080401, Feb 2017.

[27] Xiaoqin Gao, Manuel Erhard, Anton Zeilinger, and Mario Krenn. Computer-inspired concept for high-dimensional multipartite quantum gates. *Phys. Rev. Lett.*, 125:050501, Jul 2020.

[28] Sören Arlt, Carlos Ruiz-Gonzalez, and Mario Krenn. Digital discovery of a scientific concept at the core of experimental quantum optics. *arXiv:2210.09981*, 2022.

[29] Tareq Jaouni, Xiaoqin Gao, Sören Arlt, Mario Krenn, and Ebrahim Karimi. Experimental solutions to the high-dimensional mean king's problem. *arXiv:2307.12938*, 2023.

[30] A Mordvintsev, C Olah, and M Tyka. Inceptionism: Going Deeper into Neural Networks, jun 2015. Avaliable at https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html.

[31] Romi Lifshitz. Quantum deep dreaming: A novel approach for quantum circuit design. *arXiv:2211.04343*, 2022.

[32] Cynthia Shen, Mario Krenn, Sagi Eppel, and Alán Aspuru-Guzik. Deep molecular dreaming: inverse machine learning for de-novo molecular design and interpretability with surjective representations. *Machine Learning: Science and Technology*, 2(3):03LT02, jul 2021.

[33] Frank Schindler, Nicolas Regnault, and Titus Neupert. Probing many-body localization with neural networks. *Physical Review B*, 95(24):245134, jun 2017.

[34] Alireza Seif, Mohammad Hafezi, and Christopher Jarzynski. Machine learning the thermodynamic arrow of time. *Nature Physics*, 17(1):105–113, jan 2021.

[35] Mario Krenn, Xuemei Gu, and Anton Zeilinger. Quantum Experiments and Graphs: Multiparty States as coherent superpositions of Perfect Matchings. *Physical Review Letters*, 119(24):240403, dec 2017.

[36] Xuemei Gu, Manuel Erhard, Anton Zeilinger, and Mario Krenn. Quantum experiments and graphs II: Quantum interference, computation, and state generation. *Proceedings of the National Academy of Sciences*, 116(10):4147–4155, March 2019.

[37] Xuemei Gu, Lijun Chen, Anton Zeilinger, and Mario Krenn. Quantum experiments and graphs. iii. high-dimensional and multiparticle entanglement. *Phys. Rev. A*, 99:032338, Mar 2019.

[38] Yunhong Ding, Daniel Llewellyn, Imad Faruque, Stefano Paesani, Davide Bacco, Karsten Rottwitt, Anthony Laing, Mark Thompson, Jianwei Wang, and Leif K. Oxenløwe. Integrated quantum photonics on silicon platform. In *Optical Fiber Communication Conference (OFC) 2020*, page W4C.6. Optica Publishing Group, 2020.

[39] Lan-Tian Feng, Guang-Can Guo, and Xi-Feng Ren. Progress on integrated quantum photonic sources with silicon. *Advanced Quantum Technologies*, 3(2):1900058, 2020.

[40] Emanuele Pelucchi, Giorgos Fagas, Igor Aharonovich, Dirk Englund, Eden Figueroa, Qihuang Gong, Hübel Hannes, Jin Liu, Chao-Yang Lu, Nobuyuki Matsuda, et al. The potential and global outlook of integrated photonics for quantum technologies. *Nature Reviews Physics*, 4(3):194–208, 2022.

[41] Jueming Bao, Zhaorong Fu, Tanumoy Pramanik, Jun Mao, Yulin Chi, Yingkang Cao, Chonghao Zhai, Yifei Mao, Tianxiang Dai, Xiaojiong Chen, et al. Very-large-scale integrated quantum graph photonics. *Nature Photonics*, pages 1–9, 2023.

[42] Kaiyi Qian, Kai Wang, Leizhen Chen, Zhaohua Hou, Mario Krenn, Shining Zhu, and Xiao-song Ma. Multiphoton non-local quantum interference controlled by an undetected photon. *Nature Communications*, 14(1):1480, 2023.

[43] Lan-Tian Feng, Ming Zhang, Di Liu, Yu-Jie Cheng, Guo-Ping Guo, Dao-Xin Dai, Guang-Can Guo, Mario Krenn, and Xi-Feng Ren. On-chip quantum interference between the origins of a multi-photon state. *Optica*, 10(1):105–109, Jan 2023.

[44] Daniel M Greenberger, Michael A Horne, and Anton Zeilinger. Going beyond bell's theorem. In *Bell's theorem, quantum theory and conceptions of the universe*, pages 69–72. Springer, 1989.

[45] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[46] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014.

[47] Anh Nguyen, Jason Yosinski, and Jeff Clune. Multifaceted feature visualization: Uncovering the different types of features learned by each neuron in deep neural networks. *arXiv:1602.03616*, 2016.

[48] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv:1312.6034*, 2013.

[49] Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pages 8867–8887. PMLR, 2022.