

Machine learning meets quantum foundations: A brief survey


Cite as: AVS Quantum Sci. 2, 034101 (2020); <https://doi.org/10.1116/5.0007529>


Submitted: 14 March 2020 . Accepted: 12 June 2020 . Published Online: 10 July 2020

 Kishor Bharti,  Tobias Haug, Vlatko Vedral, and Leong-Chuan Kwek

COLLECTIONS

Paper published as part of the special topic on [Special Topic: Towards Practical Quantum Computers PQC2020](#)

 This paper was selected as Featured

 This paper was selected as Scilight



View Online



Export Citation

ARTICLES YOU MAY BE INTERESTED IN

[Machine learning helps researchers gain crucial understanding into quantum foundations](#)
Scilight 2020, 281107 (2020); <https://doi.org/10.1063/10.0001610>

[Quantum effects in the brain: A review](#)

AVS Quantum Science 2, 022901 (2020); <https://doi.org/10.1116/1.5135170>

[Introduction to quantum optimal control for quantum sensing with nitrogen-vacancy centers in diamond](#)

AVS Quantum Science 2, 024701 (2020); <https://doi.org/10.1116/5.0006785>





Advance your science and career as a member of



LEARN MORE



Machine learning meets quantum foundations: A brief survey



Cite as: AVS Quantum Sci. 2, 034101 (2020); doi: 10.1116/5.0007529

Submitted: 14 March 2020 · Accepted: 12 June 2020 ·

Published Online: 10 July 2020



View Online



Export Citation



CrossMark

Kishor Bharti,^{1,a)} Tobias Haug,¹ Vlatko Vedral,^{1,2} and Leong-Chuan Kwek^{1,3,4,5}

AFFILIATIONS

¹Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543

²Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom

³MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, Singapore

⁴National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616

⁵School of Electrical and Electronic Engineering Block S2.1, 50 Nanyang Avenue, Singapore 639798

Note: This paper is part of the special topic Towards Practical Quantum Computers.

^{a)}Electronic mail: e0016779@u.nus.edu

ABSTRACT

The goal of machine learning is to facilitate a computer to execute a specific task without explicit instruction by an external party. Quantum foundations seek to explain the conceptual and mathematical edifice of quantum theory. Recently, ideas from machine learning have successfully been applied to different problems in quantum foundations. Here, the authors compile the representative works done so far at the interface of machine learning and quantum foundations. The authors conclude the survey with potential future directions.

Published by the AVS. <https://doi.org/10.1116/5.0007529>

TABLE OF CONTENTS

I. INTRODUCTION	1	C. Classification with tomographic data	11
II. MACHINE LEARNING	2	VII. NEURAL NETWORKS AS “HIDDEN” VARIABLE	
A. Artificial neural networks	3	MODELS FOR QUANTUM SYSTEMS	11
B. Relation with artificial intelligence	4	VIII. A FEW MORE APPLICATIONS	11
III. QUANTUM FOUNDATIONS	5	IX. CONCLUSION AND FUTURE WORK	11
A. Entanglement	5		
B. Bell nonlocality	5		
C. Contextuality	6		
D. Quantum steering	7		
IV. NEURAL NETWORK AS ORACLE FOR BELL			
NONLOCALITY	8		
A. Machine learning nonlocal correlations	8		
B. Oracle for networks	8		
V. MACHINE LEARNING FOR OPTIMIZING			
SETTINGS FOR BELL NONLOCALITY	9		
A. Detecting Bell nonlocality with many-body states	9		
B. Playing Bell nonlocal games	10		
VI. MACHINE LEARNING-ASSISTED STATE			
CLASSIFICATION	10		
A. Classification with Bell inequalities	10		
B. Classification by representing quantum states			
with restricted Boltzmann machines	11		

I. INTRODUCTION

The rise of machine learning (ML) in recent times has remarkably transformed science and society. The goal of machine learning is to get computers to act without being explicitly programmed.^{1,2} Some of the typical applications of machine learning are self-driving cars, efficient web search, improved speech recognition, enhanced understanding of the human genome, and online fraud detection. This viral spread in interest has exploded to various areas of science and engineering, in part due to the hope that artificial intelligence (AI) may supplement human intelligence to understand some of the deep problems in science.

The techniques from machine learning have been used for automated theorem proving, drug discovery, and predicting the 3D structure of proteins based on their genetic sequence.^{3–5} In physics, techniques from machine learning have been applied for many important avenues^{6–39} including the study of black hole detection,²⁵

topological codes,⁴⁰ phase transition,¹⁶ glassy dynamics,²⁷ gravitational lenses,²³ Monte Carlo simulation,^{28,29} wave analysis,²⁴ quantum state preparation,^{41,42} and anti-de Sitter/conformal field theory correspondence⁴³ and characterizing the landscape of string theories.⁴⁴ Vice versa, the methods from physics have also transformed the field of machine learning at both the foundational and practical front.^{45,46} For a comprehensive review on machine learning for physics, refer to the study by Carleo *et al.*⁴⁷ and references therein. For a thorough review on machine learning and artificial intelligence in the quantum domain, refer to the study by Dunjko *et al.*⁴⁸ or Benedetti *et al.*⁴⁹

Philosophies in science can, in general, be delineated from the study of the science itself. Yet, in physics, the study of quantum foundations has essentially sprouted an enormously successful area called quantum information science. Quantum foundations tell us about the mathematical and conceptual understanding of quantum theory. Ironically, this area has potentially provided the seeds for future computation and communication, without, at the moment, reaching a consensus among all the physicists regarding what quantum theory tells us about the nature of reality.⁵⁰

In recent years, techniques from machine learning have been used to solve some of the analytically/numerically complex problems in quantum foundations. In particular, the methods from reinforcement learning (RL) and supervised learning have been used for determination of the maximum quantum violation of various Bell inequalities, the classification of experimental statistics in local/nonlocal sets, training AI for playing Bell nonlocal games, using hidden neurons as hidden variables for completion of quantum theory, and machine learning-assisted state classification.^{51–54}

Machine learning also attempts to mimic human reasoning, leading to the almost remote possibility of machine-assisted scientific discovery.⁵⁵ Can machine learning do the same with the foundations of quantum theory? At a deeper level, machine learning or artificial intelligence, presumably with some form of quantum computation, may capture somehow the essence of Bell nonlocality and contextuality. Of course, such speculation belies the fact that human abstraction and reasoning could be far more complicated than the capabilities of machines.

In this brief survey, we compile some of the representative works done so far at the interface of quantum foundations and machine learning. The survey includes eight sections excluding the Introduction. In Sec. II, we discuss the basics of machine learning. Section III contains a brief introduction to quantum foundations. In Secs. IV–VII, we discuss various applications of machine learning in quantum foundations. There is a rich catalog of works, which we could not incorporate in detail in Secs. IV–VII, but we find them worth mentioning. We include such works briefly in Sec. VIII. Finally, we conclude in Sec. IX with open questions and some speculations.

II. MACHINE LEARNING

Machine learning is a branch of artificial intelligence that involves learning from data.^{1,2} The purpose of machine learning is to facilitate a computer to achieve a specific task without an explicit instruction by an external party. According to Mitchel (1997),⁵⁶ “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if the performance at tasks in T , as measured by P , improves with E .” Note that the meaning of the word “task” does not involve the process of learning. For instance, if we are

programming a robot to play Go, playing Go is the task. Some of the examples of machine learning tasks are given as follows.

- (1) Classification: In classification tasks, the computer program is trained to learn the appropriate function $h: \mathbb{R}^m \rightarrow \{1, 2, \dots, t\}$. Given an input, the learned program determines which of the t categories the input belongs to via h . Deciding if a given picture depicts a cat or a dog is a canonical example of a classification task.
- (2) Regression: In regression tasks, the computer program is trained to predict a numerical value for a given input. The aim is to learn the appropriate function $h: \mathbb{R}^m \rightarrow \mathbb{R}$. A typical example of regression is predicting the price of a house given its size, location, and other relevant features.
- (3) Anomaly detection: In anomaly detection (also known as outlier detection) tasks, the goal is to identify rare items, events, or objects that are significantly different from the majority of data. A representative example of anomaly detection is credit card fraud detection where the credit card company can detect the misuse of the customer’s card by modeling his/her purchasing habits.
- (4) Denoising: Given a noisy example $\tilde{x} \in \mathbb{R}^n$, the goal of denoising is to predict the conditional probability distribution $P(x|\tilde{x})$ over noise-free data $x \in \mathbb{R}^n$.

The measure of success P of a machine learning algorithm depends on task T . For example, in the case of classification, P can be measured via the accuracy of the model, i.e., fraction of examples for which the model produces the correct output. An equivalent description can be in terms of error rate, i.e., fraction of examples for which the model produces the incorrect output. The goal of machine learning algorithms is to work well on previously unseen data. To get an estimate of model performance P , it is customary to estimate P on a separate dataset called the test set, which the machine has not seen during the training. The data employed for training are known as the training set.

Depending on the kind of experience, E , the machine is permitted to have during the learning process, the machine learning algorithms can be categorized into supervised learning, unsupervised learning, and reinforcement learning.

- (1) Supervised learning: The goal is to learn a function $y = f(x)$, which returns label y given the corresponding unlabeled data x . A prominent example would be images of cats and dogs, with the goal of recognizing the correct animal. The machine is trained with labeled example data, such that it learns to correctly identify datasets it has not seen before. Given a finite set of training samples from the joint distribution $P(Y, X)$, the task of supervised learning is to infer the probability of a specific label y given example data x , i.e., $P(Y = y|X = x)$. The function that assigns labels can be inferred from the aforementioned conditional probability distribution.
- (2) Unsupervised learning: For this type of machine learning, data x are given without any label. The goal is to recognize possible underlying structures in the data. The task of the unsupervised machine learning algorithms is to learn the probability distribution $P(x)$ or some interesting properties of the distribution, when given access to several examples x . It is worth stating that

the distinction between supervised and unsupervised learning can be blurry. For example, given a vector $x \in \mathbb{R}^m$, the joint probability distribution can be factorized (using the chain rule of probability) as

$$P(x) = \prod_{i=1}^m P(x_i | x_1, \dots, x_{i-1}). \quad (1)$$

The above factorization (1) enables us to transform an unsupervised learning task of learning $P(x)$ into m supervised learning tasks. Furthermore, given the supervised learning problem of learning the conditional distribution $P(y|x)$, one can convert it into the unsupervised learning problem of learning the joint distribution $P(x, y)$ and then infer $P(y|x)$ via

$$P(y|x) = \frac{P(x, y)}{\sum_{y1} P(x, y1)}. \quad (2)$$

The last argument suggests that supervised and unsupervised learning is not entirely distinct. Yet, this distinction between supervised and unsupervised is sometimes useful for the classification of algorithms.

- (3) Reinforcement learning: Here, neither data nor labels are available. The machine has to generate the data itself and improve this data generation process through optimization of a given reward function. This method is somewhat akin to a human child playing games: The child interacts with the environment and initially performs random actions. By external reinforcement (e.g., praise or scolding by parents), the child learns to improve itself. Reinforcement learning has shown immense success recently. Through reinforcement learning, machines have mastered games that were initially thought to be too complicated for computers to master. This was, for example, demonstrated by Deepmind's AlphaZero, which has defeated the best human player in the board game Go.⁵⁷

One of the central challenges in machine learning is to devise algorithms that perform well on previously unseen data. The learning ability of a machine to achieve a high performance P on previously unseen data is called generalization. The input data to the machine are a set of variables, called features. A specific instance of data is called the feature vector. The error measure on the feature vectors used for training is called training error. In contrast, the error measure on the test dataset, which the machine has not seen during the training, is called generalization error or test error. Given an estimate of training error, how can we estimate test error? The field of statistical learning theory aptly answers this question. The training and test data are generated according to a probability distribution over some datasets. Such a distribution is called data-generating distribution. It is conventional to make independent and identically distributed assumption, i.e., each example of the dataset is independent of another, and the training and test sets are identically distributed. Under such assumptions, the performance of the machine learning algorithm depends on its ability to reduce the training error and the gap between training and test error. If a machine-learning algorithm fails to get sufficiently low training error, the phenomenon is called under-fitting. On the other hand, if the training error is low, but the test error is large, the phenomenon is called over-fitting. The capacity of a machine learning model to fit a

wide variety of functions is called model capacity. A machine learning algorithm with low model capacity is likely to underfit, whereas a too high model capacity often leads to over-fitting the data. One of the ways to alter the model capacity of a machine learning algorithm is by constraining the class of functions that the algorithm is allowed to choose. For example, to fit a curve to a dataset, one often chooses a set of polynomials as fitting functions (see Fig. 1). If the degree of the polynomials is too low, the fit may not be able to reproduce the data sufficiently (under-fitting, orange curve). However, if the degree of the polynomials is too high, the fit will reproduce the training dataset too well, such that noise and the finite sample size are captured in the model (over-fitting, green curve).

A. Artificial neural networks

Most of the recent advances in machine learning were facilitated by using artificial neurons (ANs). The basic structure is a single AN, which is a real-valued function of the following form:

$$AN(\mathbf{x}) = \phi\left(\sum_i w_i x_i\right), \text{ where } \mathbf{x} = (x_i)_i \in \mathbb{R}^k, \\ \mathbf{w} = (w_i)_i \in \mathbb{R}^k, \text{ and } \phi: \mathbb{R} \rightarrow \mathbb{R}. \quad (3)$$

In Eq. (3), function ϕ is usually known as the activation function. Some of the well-known activation functions are

- (1) threshold function: $\phi(a) = 1$ if $a > 0$ and 0 otherwise,
- (2) sigmoid function: $\phi(a) = \sigma(a) = 1/(1 + \exp(-a))$, and
- (3) rectified linear (ReLU) function: $\phi(a) = \max(0, a)$.

The \mathbf{w} vector in Eq. (3) is known as the weight vector.

Many of those artificial neurons can be combined together via communication links to perform complex computation. This can be achieved by feeding the output of neurons (weighted by weights w_i) as an input to another neuron, where the activation function is applied again. Such a graph structure $G = (V, E)$, with the set of nodes (V) as artificial neurons and edges in E as connections, is known as an

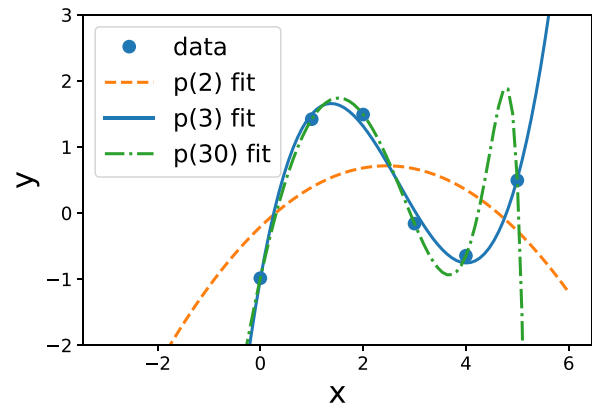


Fig. 1. Fitting data sampled from a degree 3 polynomial with small noise. A degree 2 polynomial (orange dashed line) is under-fitting as it has too low model capacity to capture the data. The degree 3 model has the right model capacity (blue solid line). The degree 30 polynomial (red dashed-dotted line) is over-fitting the data, as it captures the sampled data (low training error) but fails to capture the underlying model (high test error).

input layer hidden layers output layer

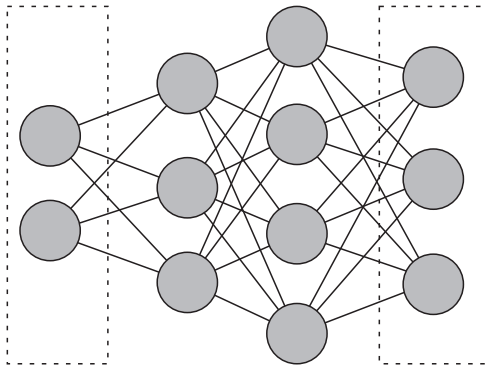


FIG. 2. Feed-forward neural networks are a network of artificial neurons chained together in sequence. Each dot corresponds to an artificial neuron, whereas the line indicates the input weights that feed into that neuron. Data fed into the input layer propagate each layer at a time via the hidden layers (here two hidden layers as example) to the final output layer.

artificial neural network or a neural network in short. The first layer, where the data are fed in, is called the input layer. The layers of neurons in-between are the hidden layers, which are defining the feature of deep learning (DL). The last layer is called the output layer. A neural network with more than one hidden layer is called the deep neural network. Machine learning involving deep neural networks is called deep learning. A feedforward neural network is a directed acyclic graph, which means that the output of the neurons is fed only into the forward direction (see Fig. 2). Apart from the feedforward neural network, some of the popular neural network architectures are convolutional neural networks, recurrent neural networks, generative adversarial network, Boltzmann machine, and restricted Boltzmann machines (RBMs). We provide a brief summary of RBMs here. For a detailed understanding of various other neural networks, refer to Ref. 2.

An RBM is a bipartite graph with two kinds of binary-valued neuron units, namely, visible ($\mathbf{v} = \{v_1, v_2, \dots\}$) and hidden ($\mathbf{h} = \{h_1, h_2, \dots\}$) (see Fig. 3). The weight matrix $W = (w_{ij})$ encodes the

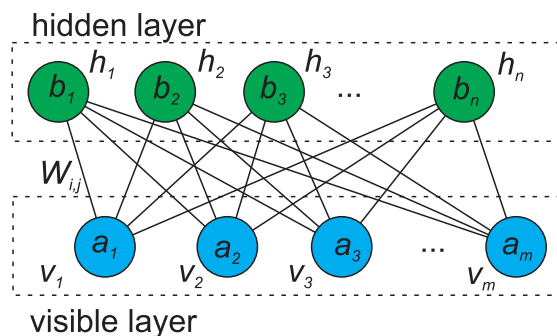


FIG. 3. Restricted Boltzmann machine (RBM) is composed of hidden and visible layers. Each node has a bias (a_i for the visible layer and b_j for the hidden layer) and is connected with weight W_{ij} to all the nodes in the opposite layer. There are no intralayer connections.

weight corresponding to the connection between visible unit v_i and hidden unit h_j .⁵⁸ Let the bias weight (offset weight) for the visible unit v_i be a_i and hidden unit h_j be b_j .

For a given configuration of visible and hidden neurons (\mathbf{v}, \mathbf{h}) , one can define an energy function $E(\mathbf{v}, \mathbf{h})$ inspired from statistical models for spin systems as

$$E(\mathbf{v}, \mathbf{h}) = - \sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j v_i w_{ij} h_j. \quad (4)$$

The probability of a configuration (\mathbf{v}, \mathbf{h}) is given by Boltzmann distribution,

$$P(\mathbf{v}, \mathbf{h}) = \frac{\exp(-E(\mathbf{v}, \mathbf{h}))}{Z}, \quad (5)$$

where $Z = \sum_{\mathbf{v}, \mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}))$ is a normalization factor, commonly known as the partition function. As there are no intralayer connections, one can sample from this distribution easily. Given a set of visible units \mathbf{v} , the probability of a specific hidden unit being $h_j = 1$ is

$$p(h_j = 1 | \mathbf{v}) = \sigma(h_j + \sum_i w_{ij} v_i), \quad (6)$$

where $\sigma(a)$ is the sigmoid activation function as introduced earlier. A similar relation holds for the reverse direction, e.g., given a set of hidden units, what is the probability of the visible unit.

With these concepts, complicated probability distributions $P(\mathbf{v})$ over some variable \mathbf{v} can be encoded and trained within the RBM. Given a set of training vectors $\mathbf{v} \in \mathbb{V}$, the goal is to find the weights W' of the RBM, which fit the training set best,

$$\arg \max_{W'} \prod_{\mathbf{v} \in \mathbb{V}} P(\mathbf{v}). \quad (7)$$

RBMs have been shown to be a good ansatz to represent many-body wavefunctions, which are difficult to handle with other methods due to the exponential scaling of the dimension of the Hilbert space. This method has been successfully applied to quantum many-body physics and quantum foundation problems.^{54,59}

In context beyond physics, machine learning with deep neural networks has accomplished many significant milestones, such as mastering various games, image recognition, and self-driving cars. Its impact on the physical sciences is just about to start.^{47,54}

B. Relation with artificial intelligence

The term “artificial intelligence” was first coined in the famous 1956 Dartmouth conference.⁶⁰ Although the term was invented in 1956, the operational idea can be traced back to Alan Turing’s influential “Computing Machinery and Intelligence” paper in 1950, where Turing asks if a machine can think.⁶¹ The idea of designing machines that can think dates back to ancient Greece. Mythical characters, such as Pygmalion, Hephaestus, and Daedalus, can be interpreted as some of the legendary inventors, and Galatea, Pandora, and Hephaestus can be thought of as examples of artificial life.² The field of artificial intelligence is difficult to define, as can be seen by four different and popular candidate definitions.⁶² The definitions start with “AI is the discipline that aims at building...”

- (1) (reasoning-based and human-based): agents that can reason like humans,
- (2) (reasoning-based and ideal rationality): agents that think rationally,
- (3) (behavior-based and human-based): agents that behave like humans, and
- (4) (behavior-based and ideal rationality): agents that behave rationally.

Apart from its foundational impact in attempting to understand “intelligence,” AI has reaped practical impacts such as automated routine labor and automated medical diagnosis, to name a few among many. The real challenge of AI is to execute tasks that are easy for people to perform but hard to express formally. An approach to solve this problem is by allowing machines to learn from experience, i.e., via machine learning. From a foundational point of judgment, the study of machine learning is vital as it helps us understand the meaning of intelligence. It is worthwhile mentioning that deep learning is a subset of machine learning, which can be thought of as a subset of AI (see Fig. 4).

III. QUANTUM FOUNDATIONS

The mathematical edifice of quantum theory has intrigued and puzzled physicists, as well as philosophers, for many years. Quantum foundations seek to understand and develop the mathematical and conceptual understanding of quantum theory. The study concerns the search for nonclassical effects such as Bell nonlocality, contextuality, and different interpretations of quantum theory. This study also involves the investigation of physical principles that can put the theory into an axiomatic framework, together with an exploration of possible extensions of quantum theory. In this survey, we will focus on nonclassical features such as entanglement, Bell nonlocality, contextuality, and quantum steering in some detail.

An interpretation of quantum theory can be viewed as a map from the elements of the mathematical structure of quantum theory to elements of reality. Most of the interpretations of quantum theory seek to explain the famous measurement problem. Some of the brilliant interpretations are the Copenhagen interpretation, quantum Bayesianism,⁶³ the many-world formalism,^{64,65} and the consistent history interpretation.⁶⁶ The axiomatic reconstruction of quantum theory is generally categorized into two parts: the generalized probabilistic theory approach⁶⁷ and the black-box approach.⁶⁸ The physical

principles underlying the framework for axiomatizing quantum theory are nontrivial communication complexity,^{69,70} information causality,⁷¹ macroscopic locality,⁷² negating the advantage for nonlocal computation,⁷³ consistent exclusivity,⁷⁴ and local orthogonality.⁷⁵ The extensions of the quantum theory include collapse models,⁷⁶ quantum measure theory,⁷⁷ and acausal quantum processes.⁷⁸

A. Entanglement

Quantum interaction inevitably leads to quantum entanglement. The individual states of two classical systems after an interaction are independent of each other. Yet, this is not the case for two quantum systems.^{79,80} Quantum states can be pure or mixed. For a bipartite pure quantum state, $|\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ is entangled if it cannot be written as a product state, i.e., $|\psi\rangle = |\psi^A\rangle \otimes |\psi^B\rangle$ for some $|\psi^A\rangle \in \mathcal{H}^A$ and $|\psi^B\rangle \in \mathcal{H}^B$. A mixed bipartite state is, however, expressed in terms of a density matrix, ρ . Like for a pure state, a density matrix ρ is entangled if it cannot be expressed in the form $\rho = \sum_i p_i |\psi_i^A\rangle \langle \psi_i^A| \otimes |\psi_i^B\rangle \langle \psi_i^B|$. An n -partite state ρ_{sep} is called separable (SEP) if it can be represented as a convex combination of product states, i.e.,

$$\rho_{sep} = \sum_i p_i \rho_i^1 \otimes \rho_i^2 \otimes \cdots \otimes \rho_i^n, \quad (8)$$

where $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$. A quantum state is called separable if it is not entangled.

Computationally, it is not easy to check if a mixed state, especially in higher dimensions and for more parties, is separable or entangled. Numerous measures of quantum entanglement for both pure and mixed states are proposed.⁸⁰ for bipartite systems where the dimension of \mathcal{H}^i ($i = A, B$) is 2, a good measure is concurrence.⁸¹ Other measures for quantifying entanglement are entanglement of formation, robustness of entanglement, Schmidt rank, squashed entanglement, and so forth.⁸⁰ It turns out that for any entangled state ρ , there exists a Hermitian matrix W such that $\text{Tr}(\rho W) < 0$ and for all separable states, ρ_{sep} , $\text{Tr}(\rho W) \geq 0$.^{82–85}

Quantum entanglement was mooted a long time ago by Erwin Schrödinger, but it took another thirty years or so for John Bell to show that quantum theory imposes strong constraints on statistical correlations in experiments. Yet, correlations are not tantamount to causation, and one wonders if machine learning could do better.⁸⁶ In the sixties and seventies, this Gedanken experiment was given further impetus with some ingenious experimental designs.^{87–90} The advent of quantum information in the nineties gave a further push: quantum entanglement became a useful resource for many quantum applications, ranging from teleportation,⁹¹ communication,⁹² purification,⁹³ dense coding,⁹⁴ and computation.^{95,96} Interestingly, quantum entanglement is a fascinating area that emerges almost serendipitously from the foundation of quantum mechanics into real practical applications in the laboratories.

B. Bell nonlocality

According to Bell,⁹⁷ any theory based on the collective premises of locality and realism must be at variance with experiments conducted by spatially separated parties involving shared entanglement, if the underlying measurement events are spacelike separated. The phenomenon, as discussed before, is known as Bell nonlocality.⁹⁸ Apart from its significance in understanding foundations of quantum theory,

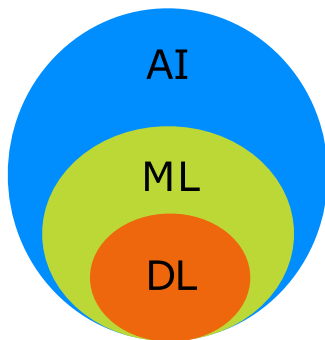


Fig. 4. DL is a subdiscipline of ML. ML can be considered a subdiscipline of AI.

Bell nonlocality is a valuable resource for many emerging device-independent quantum technologies such as quantum key distribution, distributed computing, randomness certification, and self-testing.^{92,99,100} The experiments, which can potentially manifest Bell nonlocality, are known as Bell experiments. A Bell experiment involves N spatially separated parties A_1, A_2, \dots, A_N . Each party receives an input $x_1, x_2, x_3, \dots, x_N \in \mathcal{X}$ and gives an output $a_1, a_2, a_3, \dots, a_N \in \mathcal{A}$. For various input-output combinations, one gets the statistics of the following form:

$$\mathcal{P} = \{P(a_1, a_2, \dots, a_N | x_1, x_2, \dots, x_N)\}_{x_1, \dots, x_N \in \mathcal{X}, a_1, \dots, a_N \in \mathcal{A}}. \quad (9)$$

We will refer to \mathcal{P} as behavior. A Bell experiment involving N space-like separated parties, each party having access to m inputs and each input corresponding to k outputs, is referred to as the (N, m, k) scenario. The famous Clauser–Horne–Shimony–Holt (CHSH) experiment is a $(2, 2, 2)$ scenario,¹⁰¹ and it is the simplest scenario in which Bell nonlocality can be demonstrated (see Fig. 5).

A behavior \mathcal{P} admits a local hidden variable description if and only if

$$P(a_1, a_2, \dots, a_N | x_1, x_2, \dots, x_N) = \sum_{\lambda} P(\lambda) P(a_1 | x_1, \lambda) \cdots P(a_N | x_N, \lambda). \quad (10)$$

This is known as local behavior. The set of local behaviors (\mathcal{L}) forms a convex polytope, and the facets of this polytope are Bell inequalities (see Fig. 6). In quantum theory, the Born rule governs probability according to

$$P(a_1, a_2, \dots, a_N | x_1, x_2, \dots, x_N) = \text{Tr}[[M_{a_1|x_1} \otimes \cdots \otimes M_{a_N|x_N}] \rho], \quad (11)$$

where $\{M_{a_i|x_i}\}_i$ are positive-operator valued measures (POVMs) and ρ is a shared density matrix. If a behavior satisfying Eq. (11) falls outside \mathcal{L} , it then violates at least one Bell inequality, and such a behavior is said to manifest Bell nonlocality. The condition that parties do not communicate during the course of the Bell experiment is known as the

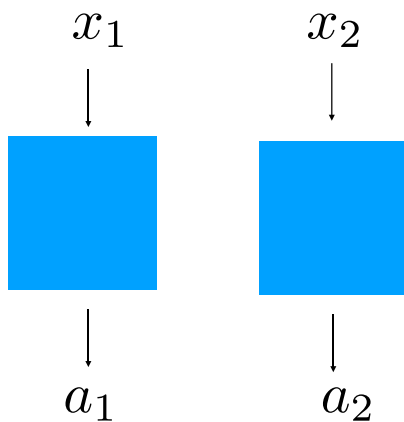


Fig. 5. Simplest scenario in which Bell nonlocality can be demonstrated is the famous CHSH experiment. There are two parties involved. Each party can perform two dichotomic measurements. Thus, it is a $(2, 2, 2)$ scenario. The experimental statistics corresponds to probabilities of the form $P(a_1, a_2 | x_1, x_2)$.

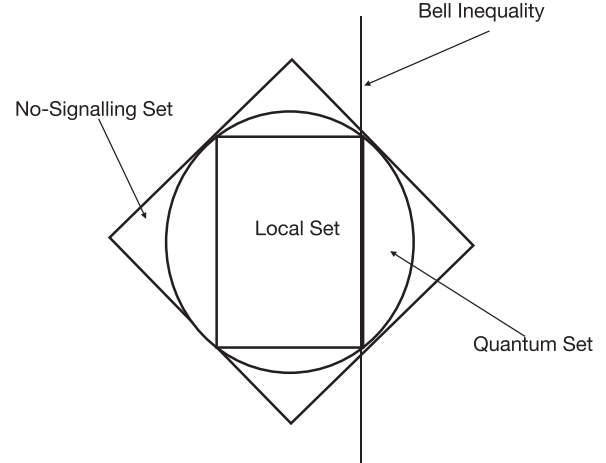


Fig. 6. Set of local behaviors (\mathcal{L}) forms a convex polytope (bounded polyhedron). The set of quantum behaviors \mathcal{Q} is a convex set. The set of behaviors compatible with the no-signaling condition (\mathcal{NS}) is again a convex polytope. It is important to note $\mathcal{L} \subseteq \mathcal{Q} \subseteq \mathcal{NS}$. The hyperplanes defining the boundary of the local set \mathcal{L} are Bell inequalities. It is important to reiterate that the violation of Bell inequalities means that a local realistic description of nature is impossible.

no-signaling condition. Intuitively speaking, it means that the choice of input of one party cannot be used for signaling among parties. Mathematically, it means that

$$\sum_{a_j} P(a_i, a_j | x_i, x_j) = P(a_i | x_i) \forall i, j \in \{1, \dots, N\} \text{ and } i \neq j. \quad (12)$$

The set of behaviors that satisfy the no-signaling condition are known as no-signaling behaviors. We will denote the aforementioned set by \mathcal{NS} . The no-signaling set also forms a polytope. Furthermore, $\mathcal{L} \subseteq \mathcal{Q} \subseteq \mathcal{NS}$ (see Fig. 6). The no-signaling behaviors, which do not lie in \mathcal{Q} , are also known as postquantum behaviors. In the $(2,2,2)$ scenario, i.e., CHSH Scenario, there is a unique Bell inequality, namely, CHSH inequality, up to relabeling of inputs and outputs. The CHSH inequality is given by

$$E_{0,0} + E_{0,1} + E_{1,0} + E_{1,1} \leq 2, \quad (13)$$

where $E_{x_1, x_2} = P(a_1 = a_2 | x_1 = x_2) - P(a_1 \neq a_2 | x_1 = x_2)$. All local hidden variable theories satisfy CHSH inequality. In quantum theory, suitably chosen measurement settings and state can lead to violation of CHSH inequality, and thus, the CHSH inequality can be used to witness the Bell nonlocal nature of quantum theory. Quantum behaviors achieve up to $2\sqrt{2}$, known as the Tsirelson bound. The upper bound for no-signaling behaviors (no-signaling bound) on the CHSH inequality is 4.

C. Contextuality

An intuitive feature of classical models is noncontextuality, which means that any measurement has a value independent of other compatible measurements it is performed together with. A set of compatible measurements is called context. It was shown by Kochen and Specker (and John Bell)¹⁰² that noncontextuality conflicts with quantum theory.

The contextual nature of quantum theory can be established via simple constructive proofs. In Fig. 7(a), one considers an array of operators on a two-qubit system in any quantum state. There are nine operators, and each of them has an eigenvalue of ± 1 . The three operators in each row or column commute, and so, it is easy to check that each operator is the product of the other two operators on a particular row or column, with a single exception that the third operator in the third column equals the minus of the product of the other two. Suppose that there exist preassigned values (-1 or $+1$) for the outcomes of the nine operators. Then, we can replace none of operators by the preassigned values. However, there is no consistent way to assign such values to the nine operators so that the product of the numbers in every row or column (except for the operators along the bold line) yield 1 (the product yields -1). Notice that each operator (node) appears in exactly two lines or context. Kochen and Specker provided the first proof of quantum contextuality with a complicated construct involving 117 operators on a three-dimensional space.¹⁰² Another example is the pentagram¹⁰³ in Fig. 7.

Bell nonlocality is regarded as a particular case of contextuality where the spacelike separation of parties involved creates “context.”^{104,105} The study of contextuality not only has led to insights into the foundations of quantum mechanics but also offers practical implications.^{106–120} There are several frameworks for contextuality including the sheaf-theoretic framework,¹²¹ graph and hypergraph framework,^{104,122} contextuality-by-default framework,^{123–125} and operational framework.¹²⁶ A scenario exhibits contextuality if it does not admit the noncontextual hidden variable (NCHV) model.

D. Quantum steering

Correlations produced by steering lie between the Bell nonlocal correlations and those generated from entangled states.^{127,128} A state that manifests Bell nonlocality for some suitably chosen measurement settings also exhibits steering.¹²⁹ Furthermore, a state that exhibits steering must be entangled. A state demonstrates steering if it does not admit the “local hidden state (LHS)” model.¹²⁸ We discuss this formally here.

Alice and Bob share some unknown quantum state ρ^{AB} . Alice can perform a set of POVM measurements $\{M_a^x\}_a$. The probability of her getting outcome a after choosing measurement x is given by

$$P(a|x) = \text{Tr}[(M_a^x \otimes I)\rho^{AB}] = \text{Tr}\{\text{Tr}_A[(M_a^x \otimes I)\rho^{AB}]\} = \text{Tr}[\rho_{a|x}^B], \quad (14)$$

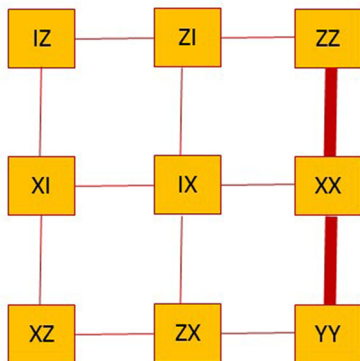
where $\rho_{a|x}^B = \text{Tr}_A[(M_a^x \otimes I)\rho^{AB}]$ is Bob’s residual state upon normalization. A set of operators $\{\rho_{a|x}^B\}_{a,x}$ acting on Bob’s space is called an assemblage if

$$\sum_a \rho_{a|x}^B = \sum_a \rho_{a|x'}^B \quad \forall x \neq x' \quad (15)$$

and

$$\sum_a \text{Tr}[\rho_{a|x}^B] = 1 \quad \forall x. \quad (16)$$

(A)



(B)

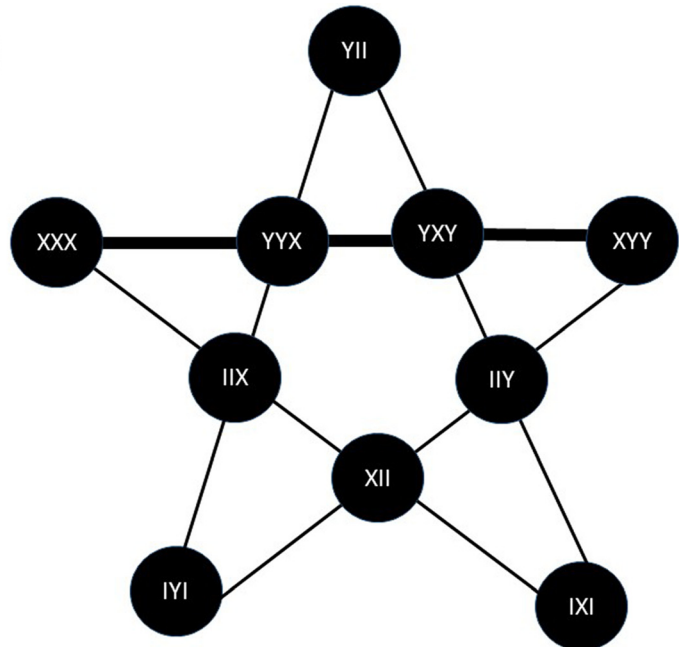


FIG. 7. Mermin-Peres square in (a) provides a proof of the Kochen–Specker theorem. Each line in the 3×3 grid describes mutually commuting operators. Since the square of the operator at each node is the identity matrix, its eigenvalues are ± 1 . The product of the operator along each line is the Identity, except for the bold one in which the product is minus the Identity matrix. There is no way to assign definite values ± 1 to each operator to get these product rules since each operator appears in exactly two lines (or contexts). (b) Mermin’s pentagram where we have four mutually commuting three-qubit operators. Along each line, the product of the operators is Identity except for the bold one, leading to a contradiction. Classically, it is impossible to assign the numbers ± 1 to each node such that the product of the numbers along a straight line is 1, except for the bold line, which is -1 .

Condition 15 is the analog of the no-signaling condition. An assemblage $\{\rho_{a|x}^B\}_{a,x}$ is said to admit the LHS model if there exists some hidden variable λ and some quantum state ρ_λ acting on Bob's space such that

$$\rho_{a|x}^B = \sum_{\lambda} P(\lambda) P(a|x, \lambda) \rho_{\lambda}. \quad (17)$$

A bipartite state ρ^{AB} is said to be steerable from Alice to Bob if there exist measurements for Alice such that the corresponding assemblage does not satisfy Eq. (17). Determining whether an assemblage admits the LHS model is a semidefinite program. The concept of steering is asymmetric by definition, i.e., even if Alice could steer Bob's state, Bob may not be able to steer Alice's state.

IV. NEURAL NETWORK AS ORACLE FOR BELL NONLOCALITY

The characterization of the local set for the convex scenario via Bell inequalities becomes intractable as the complexity of the underlying scenario grows (in terms of the number of parties, measurement settings, and outcomes). For networks where several independent sources are shared among many parties, the situation gets increasingly worse. The local set is remarkably nonconvex, and hence, proper analytical and numerical characterization, in general, is lacking. Applying the machine learning technique to tackle these issues was studied by Canabarro *et al.*⁵² and Kriváchy *et al.*⁵¹ In the work by Canabarro *et al.*, the detection and characterization of nonlocality are done through an ensemble of multilayer perceptrons blended with genetic algorithms (see Sec. IV A).

A. Machine learning nonlocal correlations

Given a behavior, deciding whether it is classical or nonclassical is an extremely challenging task since the underlying scenario grows in complexity very quickly. Canabarro *et al.*⁵² used supervised machine learning with an ensemble of neural networks to tackle the approximate version of the problem (i.e., with a small margin of error) via regression. They asked "How far is a given correlation from the local set." The input feature vector to the neural network is a random correlation vector. For a given behavior, the output (label) is the distance of the feature vector from the classical, i.e., local set. The nonlocality quantifier of a behavior q is the minimum trace distance, denoted by $NL(q)$.¹³⁰ For the two-party scenario, the nonlocality quantifier is given by

$$NL(q) \equiv \frac{1}{2|x||y|} \min_{p \in \mathcal{L}} \sum_{a,b,x,y} |q - p|, \quad (18)$$

where \mathcal{L} is the local set and $|x| = |y| = m$ is the input size for the parties. The training points are generated by sampling the set of non-signaling behaviors randomly and then calculating its distance from the local set via Eq. (18). Given a behavior q , the distance predicted by the neural network is never equal to the actual distance, i.e., there is always a small error ($\epsilon \neq 0$). Let us represent the learned hypothesis as $f: q \rightarrow \mathbb{R}$. The performance metric of the learned hypothesis f is given by

$$\mathcal{P}(f) \equiv \frac{1}{N} \sum_{i=1}^N |NL(q_i) - f(q_i)|. \quad (19)$$

In experiments such as entanglement swapping (Fig. 8), which comprises three separated parties sharing two independent sources of quantum states, the local set admits the following form, and the set is nonconvex,

$$P(a_1, a_2, a_3) = \sum_{\lambda_1, \lambda_2} P(\lambda_1) P(\lambda_2) P(a_1|x_1, \lambda_1) P(a_2|x_2, \lambda_2) \times P(a_3|x_3, \lambda_3). \quad (20)$$

Here, nonconvexity emerges from the independence of the sources, i.e., $P(\lambda_1, \lambda_2) = P(\lambda_1)P(\lambda_2)$.

In this case, the Bell inequalities are no longer linear. Characterizing the set of classical and quantum behaviors gets complicated for such scenarios.^{131,132}

Canabarro *et al.* trained the model for convex and nonconvex scenarios. They also trained the model to learn postquantum correlations. The techniques studied in the paper are valuable for understanding Bell nonlocality for large quantum networks, for example, those in quantum internet.

B. Oracle for networks

Given an observed probability distribution corresponding to scenarios where several independent sources are distributed over a network, deciding whether it is classical or nonclassical is an important question, both from practical and foundational viewpoints. The boundary separating the classical and nonclassical correlations is extremely nonconvex, and thus, a rigorous analysis is exceptionally challenging.

In Ref. 51, the authors encoded the causal structure into the topology of a neural network and numerically determined if the target distribution is "learnable." A behavior belongs to the local set if it is learnable. The authors harnessed the fact that both the information flow in feedforward neural networks and causal structures are determined by a directed acyclic graph. The topology of the neural network is chosen such that it respects the causality structure. The local set corresponding to even elementary causal networks such as triangle network is profoundly nonconvex, and thus, analytical characterization of the same is a notoriously tricky task. Using the neural network as an oracle, Kriváchy *et al.*⁵¹ converted the membership in a local set problem to a learnability problem. For a neural network with adequate model capacity, a target distribution can be approximated if it is local. The authors

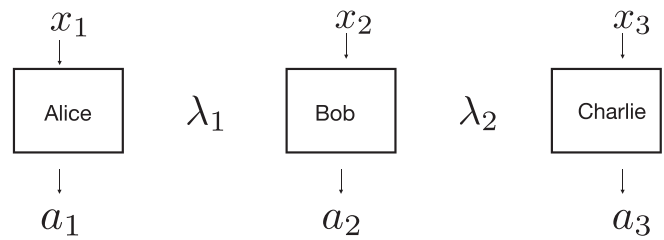


Fig. 8. There are three separated parties with two sources of hidden variables. The sources are independent $P(\lambda_1, \lambda_2) = P(\lambda_1)P(\lambda_2)$. The local set is no longer convex. In experiments such as entanglement swapping, such nonconvexity arises. Deciding whether a behavior is classical or quantum gets complicated for such scenarios.

examined the triangle network with quaternary outcomes as a proof-of-principle example. In such a scenario, there are three independent sources, say α , β , and γ . Each of the three parties receives input from two of the three sources and processes the inputs to provide outputs via fixed response functions. The outputs for Alice, Bob, and Charlie will be indicated by $a, b, c \in \{0, 1, 2, 3\}$. The scenario as discussed here can be characterized by the probability distribution $P(a, b, c)$ over the random variables a, b , and c . If the network is classical, then the distribution can be represented by a directed acyclic graph known as a Bayesian network.

Assuming the distribution $P(a, b, c)$ over the random variables a, b , and c to be classical, it is assumed without loss of generality that the sources send a random variable drawn uniformly from the interval 0 to 1. A classical distribution for such a case admits the following form:

$$P(a, b, c) = \int_0^1 dx dy dz P_A(a|\beta, \gamma) P_B(b|\gamma, \alpha) P_C(c|\alpha, \beta). \quad (21)$$

The neural network is constructed such that it can approximate the distribution of type Eq. (21). The inputs to the neural network are α , β , and γ drawn uniformly at random, and the outputs are the conditional probabilities, i.e., $P_A(a|\beta, \gamma)$, $P_B(b|\gamma, \alpha)$, and $P_C(c|\alpha, \beta)$ (see Fig. 9).

The cost function is chosen to be any measure of the distance between the target distribution P_t and the network's output P_M . The authors employed the techniques to a few other cases, such as the elegant distribution and a distribution proposed by Renou *et al.*¹³³ The application of the technique to the elegant distribution suggests that the distribution is indeed nonlocal as conjectured in Ref. 134. Furthermore, the distribution proposed by Renou *et al.* appears to

have nonlocal features for some parameter regime. For the sake of completeness, now we discuss the elegant distribution and the distribution proposed by Renou *et al.*

Elegant distribution—The distribution is generated by three parties performing entangled measurements on entangled systems. The three parties share singlets, i.e., $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. Every party performs entangled measurements on their two qubits. The eigenstates of the entangled measurements are given by

$$|\chi_j\rangle = \sqrt{\frac{3}{2}}|\alpha_j, -\alpha_j\rangle + i\frac{\sqrt{3}-1}{2}|\psi^-\rangle, \quad (22)$$

where $|\alpha_j\rangle$ are vectors symmetrically distributed on the Bloch sphere, i.e., point to the vertices of a tetrahedron, for $j \in \{1, 2, 3, 4\}$.

Distribution of Renou et al.—The distribution is generated by three parties sharing the entangled state $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and performing the same measurement on each of their two qubits. The measurements are characterized by a single parameter $\kappa \in [-\frac{1}{\sqrt{2}}, 1]$ with eigenstates $|01\rangle, |10\rangle, u|00\rangle + \sqrt{1-u^2}|11\rangle$, and $\sqrt{1-u^2}|00\rangle - u|11\rangle$.

V. MACHINE LEARNING FOR OPTIMIZING SETTINGS FOR BELL NONLOCALITY

Bell inequalities have become a standard tool to reveal the nonlocal structure of quantum mechanics. However, finding the best strategies to violate a given Bell inequality can be a difficult task, especially for many-body settings or even nonconvex scenarios. Specifically, the latter setting is challenging, as standard optimization tools are unable to be applied to this case. To violate a given Bell inequality, two interdependent tasks have to be addressed: Which measurements have to be performed to reveal the nonlocality? And which quantum states show the maximal violation? Recently, Deng has approached the latter task for convex settings with many-body Bell inequalities using restricted Boltzmann machines.⁵⁴ Bharti *et al.*⁵³ have approached both tasks in conjunction for both convex and nonconvex inequalities.

A. Detecting Bell nonlocality with many-body states

Several methods from machine learning have been adopted to tackle intricate quantum many-body problems.^{54,59,135,136} Deng⁵⁴ employed machine learning techniques to detect quantum nonlocality in many-body systems using the restricted Boltzmann machine (RBM) architecture. The key idea can be split into two parts:

- (1) After choosing appropriate measurement settings, Bell inequalities for convex scenarios can be expressed as an operator. This operator can be thought of as an Hamiltonian. The eigenstate with the maximal eigenvalue is the state that gives the maximum violation for the underlying Bell inequality. This state can be found by calculating the ground state of the negative of the Hamiltonian.
- (2) Finding the ground state of a quantum Hamiltonian is QMA-hard and thus in general difficult. However, using heuristic techniques involving the RBM architecture, the problem is recast into the task of finding the approximate answer in some cases.

Techniques such as the density matrix renormalization group (DMRG),¹³⁷ projected entangled pair states (PEPSs),¹³⁸ and multiscale entanglement renormalization ansatz¹³⁹ are traditionally used to find

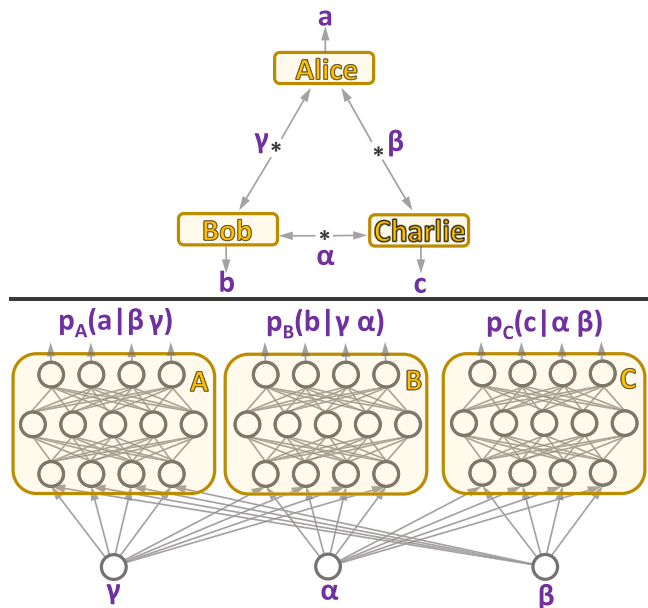


FIG. 9. Reprinted with permission from Kriváchy *et al.*, preprint arXiv:1907.10552 (2019). Copyright 2019 ■ (Ref. 51). (Top) The triangle network configuration. (Bottom) The topology of the neural network is selected such that it reproduces distributions compatible with the triangle configuration.

(or approximate) ground states of many-body Hamiltonians. However, these techniques only work reliably for optimization problems involving low-entanglement states. Moreover, the DMRG works well only for systems with short-range interactions in the one-dimensional case. As evident from Ref. 59, the RBM can represent quantum many-body problems beyond 1D and low-entanglement.

B. Playing Bell nonlocal games

Prediction of winning strategies for (classical) games and decision-making processes with RL has made significant progress in game theory in recent years. Motivated partly by these results, Bharti *et al.*⁵³ have looked at a game-theoretic formulation of Bell inequalities (known as Bell games) and applied machine learning techniques to it. To violate a Bell inequality, both the quantum state and the measurements performed on the quantum states have to be chosen in a specific manner. They transformed this problem into a decision-making process. This is achieved by choosing the parameters in a Bell game in a sequential manner, e.g., the angles of the measurement operators, the angles parameterizing the quantum states, or both. Using RL, these sequential actions are optimized for the best configuration corresponding to the optimal/near-optimal quantum violation (see Fig. 10). They trained the RL agent with a cost function that encourages high quantum violations via proximal policy optimization—a state-of-the-art RL algorithm that combines two neural networks. The approach succeeds for well-known convex Bell inequalities, but it can also solve Bell inequalities corresponding to nonconvex optimization problems, such as in larger quantum networks. So far, the field has struggled solving

these inequalities; thus, this approach offers a novel possibility toward finding optimal (or near-optimal) configurations.

Furthermore, they presented an approach to find settings corresponding to maximum quantum violation of Bell inequalities on near-term quantum computers. The quantum state is parameterized by a circuit consisting of single-qubit rotations and CNOT gates acting on neighboring qubits arranged in a linear fashion. Both the gates and the measurement angles are optimized using a variational hybrid classical-quantum algorithm, where the classical optimization is performed by RL. The RL agent learns by first randomly trying various measurement angles and quantum states. Over the course of the training, the agent improves itself by learning from experience and is capable of reaching the maximal quantum violation.

VI. MACHINE LEARNING-ASSISTED STATE CLASSIFICATION

A crucial problem in quantum information is identifying the degree of entanglement within a given quantum state. For Hilbert space dimensions up to 6, one can use the Peres–Horodecki criterion, also known as the positive partial transpose (PPT) criterion, to distinguish entangled and separable states. However, there is no generic observable or entanglement witness as it is in fact a Non-deterministic polynomial time hard (NP-hard) problem.⁸⁰ Thus, one must rely on heuristic approaches. This poses a fundamental question: Given partial or full information about the state, are there ways to classify whether it is entangled or not? Machine learning has offered a way to find answers to this question.

A. Classification with Bell inequalities

In Ref. 140, the authors blended Bell inequalities with a feed-forward neural network to use them as state classifiers. The goal is to classify states as entangled or separable. If a state violates a Bell inequality, it must be entangled. However, Bell inequalities cannot be used as a reliable tool for entanglement classification, i.e., even if a state is entangled, it might not violate an entanglement witness based on the Bell inequality. For example, the density matrix $\rho = p|\psi\rangle\langle\psi| + (1-p)\frac{I}{4}$ violates the CHSH inequality only for $p > \frac{1}{\sqrt{2}}$, but it is entangled for $p > \frac{1}{3}$.¹⁴¹ Moreover, given a Bell inequality, the measurement settings that witness the entanglement of a quantum state (if possible) depend on the quantum state. Prompted by these issues, the authors asked if they can transform Bell inequalities into a reliable separable-entangled state classifier. The coefficients of the terms in the CHSH inequalities are (1, 1, 1, -1). The local hidden variable bound on the inequality is 2 [see Eq. 13]. Assuming fixed measurement settings corresponding to CHSH inequality, the authors examined whether it is possible to get better performance in terms of entanglement classification compared with the values (1, 1, 1, -1, 2). The main challenge to answer such a question in a supervised learning setting is to get labeled data that are verified to be either separable or entangled.

To train the neural network, the correlation vector corresponding to the appropriate Bell inequality was chosen as input with the state being entangled or separable (1 versus 0) as the corresponding output. The correlation vector contains the expectation of the product of Alice's and Bob's measurement operators. The performance of the network improved as the model capacity was increased, which hints that the hypothesis that separates entangled states from separable ones

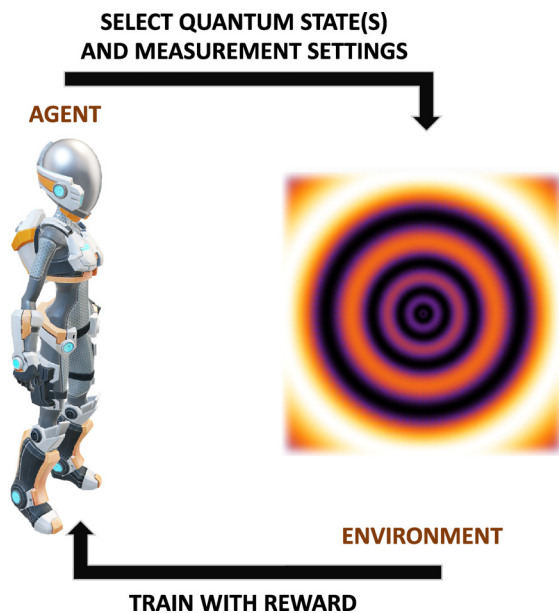


FIG. 10. Reprinted with permission from Bharti *et al.*, preprint arXiv:1912.10783 (2019). Copyright 2019 ■ (Ref. 53). Playing Bell games with AI: in Ref. 53, the authors trained the AI agent to play various Bell nonlocal games. The agent interacts with the quantum system by choosing quantum states and measurement angles and measures resulting violation of the Bell inequalities. Over repeated games, the agent trains himself using past results to realize maximal violation of Bell inequalities.

must be sufficiently “complex.” The authors also trained a neural network to distinguish biseparable and bound-entangled states.

B. Classification by representing quantum states with restricted Boltzmann machines

Harney *et al.*¹⁴² used reinforcement learning with RBMs to detect entangled states. RBMs have demonstrated being capable of learning complex quantum states. The authors modified RBMs such that they can only represent separable states. This is achieved by separating the RBM into K partitions that are only connected within themselves but not with the other partitions. Each partition represents a (possibly) entangled subsystem; however, it is not entangled with the other partitions. This choice enforces a specific K -separable ansatz of the wavefunction. This RBM is trained to represent a target state. If the training converges, it must be representable by the ansatz and thus be K -separable. However, if the training does not converge, the ansatz is insufficient, and the target state is either of a different K' -separable form or fully entangled.

C. Classification with tomographic data

Can tools from machine learning help to distinguish entangled and separable states given the full quantum state (e.g., obtained by quantum tomography) as an input? Two recent studies address this question.

In Ref. 143, Lu *et al.* detected entanglement by employing classic (i.e., nondeep learning) supervised learning. To simplify data generation of entangled and separable states, they approximated the set of separable states by a convex hull (CHA). States that lie outside the convex hull are assumed to be most likely entangled. For the decision process, they used ensemble training via bootstrap aggregating (bagging). Here, various supervised learning methods are trained on the data, and they form a committee together, which decides whether a given state is entangled or not. The algorithm is trained with the quantum state and information encoding the position relative to the convex hull as inputs. The authors showed that accuracy improves if bootstrapping of machine learning methods is combined with CHA.

In a different approach in Ref. 144, Goes *et al.* presented an automated machine learning approach to classify random states of two qutrits as SEP, entangled with positive partial transpose (PPTES) or entangled with negative partial transpose (NPT). For training, they elaborated a way to find enough samples to train on. The procedure is as follows: A random quantum state is sampled, and then, using the general robustness (GR) of entanglement and PPT criterion, it is classified to either SEP, PPTES, or NPT. The GR measures the closeness to the set of separable states. They compared various supervised learning methods to distinguish the states. The input features fed into the machine are the components of the quantum state vector and higher-order combinations thereof, whereas the labels are the type of entanglement. Besides, they also train to estimate the GR with regression techniques and use it to validate the classifiers.

VII. NEURAL NETWORKS AS “HIDDEN” VARIABLE MODELS FOR QUANTUM SYSTEMS

Understanding why deep neural networks work so well is an active area of research. The presence of the word hidden for hidden variables in quantum foundations and hidden neurons in deep

learning neurons may not be that accidental. Using conditional restricted Boltzmann machines (a variant of restricted Boltzmann machines), Steven Weinstein provided a completion of quantum theory in Ref. 145. The completion, however, does not contradict Bell’s theorem as the assumption of “statistical independence” is not respected. The statistical independence assumption demands that the complete description of the system before the measurement must be independent of the final measurement settings. The phenomenon where apparent nonlocality is observed by violating statistical independence assumption is known as “nonlocality without nonlocality.”¹⁴⁶

In a Bell-experiment corresponding to the CHSH scenario, the detector settings $\alpha \in \{a, a'\}$ and $\beta \in \{b, b'\}$ and the corresponding measurement outcomes $x_\alpha \in \{1, -1\}$ and $x_\beta \in \{1, -1\}$ for a single experimental trial can be represented as a four-dimensional vector $\{\alpha, \beta, x_\alpha, x_\beta\}$. Such a vector can be encoded in a binary vector $V = (v_1, v_2, v_3, v_4)$ where $v_i \in \{0, 1\}$. Here, $(+1)/(-1)$ has been mapped to $0/(+1)$. The four-dimensional binary vector V represents the value taken by four visible units of an RBM. The dependencies between the visible units are encoded using a sufficient number of hidden units $H = (h_1, h_2, \dots, h_j)$. With four hidden neurons, the authors could reproduce the statistics predicted by the EPR experiment with high accuracy. For example, say the vector $V = (0, 1, 1, 0)$ occurs in 3% of the trials, and then, after training, the machine would associate $P(V) \approx 0.03$. Quantum mechanics gives us only the conditional probabilities $P(x_\alpha, x_\beta | \alpha, \beta)$, and thus, learning joint probability using a RBM is resource-wasteful. The authors harnessed this observation by encoding the conditional statistics only in a conditional RBM (cRBM).

The difference between a cRBM and a RBM is that the units corresponding to the conditioning variables (detector settings here) are not dynamical variables. There are no probabilities assigned to conditioning variables, and thus, the only probabilities generated by cRBM are conditional probabilities. This provides a more compact representation compared to an RBM.

VIII. A FEW MORE APPLICATIONS

Work on quantum foundations has led to the birth of quantum computing and quantum information. Recently, the amalgamation of quantum theory and machine learning has led to a new area of research, namely, quantum machine learning.¹⁴⁷ For further reading on quantum machine learning, refer to Ref. 148 and references therein. Furthermore, for recent trends and exploratory works in quantum machine learning, we refer the reader to Ref. 149 and references therein.

Techniques from machine learning have been used to discover new quantum experiments.^{150,151} In Ref. 151, Krenn *et al.* provided a computer program that helps designing novel quantum experiments. The program was called Melvin. Melvin provided solutions that were quite counter-intuitive and different from something a human scientist ordinarily would come up with. Melvin’s solutions have led to many novel results.^{152–157} The ideas from Melvin have further provided machine generated proofs of the Kochen-Specker theorem.¹⁵⁸

IX. CONCLUSION AND FUTURE WORK

In this survey, we discussed the various applications of machine learning for problems in the foundations of quantum theory such as determination of the quantum bound for Bell inequalities, the classification of different behaviors in local/nonlocal sets, using hidden

neurons as hidden variables for completion of quantum theory, training AI for playing Bell nonlocal games, and ML-assisted state classification. Now, we discuss a few open questions at the interface. Some of these open questions have been mentioned in Refs. 51–54.

Witnessing Bell nonlocality in many-body systems is an active area of research.^{54,159} However, designing experimental-friendly many-body Bell inequalities is a difficult task. It would be interesting if machine learning could help design optimal Bell inequalities for scenarios involving many-body systems. In Ref. 54, the author used RBM based representation coupled with reinforcement learning to find near-optimal quantum values for various Bell inequalities corresponding to various convex Bell scenarios. It is well known that optimization becomes comparatively easier once the representation gets compact. It would be interesting if one can use other neural network based representations such as convolutional neural networks for finding optimal (or near-optimal) quantum values.

As mentioned in Ref. 52, it is an excellent idea to deploy techniques such as anomaly detection for the detection of nonclassical behaviors. This can be done by subjecting the machine to training with local behaviors only.

In many of the applications, e.g., classification of entangled states, the computer gives a guess, but we are not sure about the correctness. This is never said and cannot be overlooked, as it is a limitation of these applications. The intriguing question is to understand how the output of the computer can be employed to provide a certification of the result, for instance, an entanglement witness. One could use ideas from probabilistic machine learning in such cases.¹⁶⁰ The probabilistic framework, which explains how to represent and handle uncertainty about models and predictions, plays a fundamental role in scientific data analysis. Harnessing the tools from probabilistic machine learning such as Bayesian optimization and automatic model discovery could be conducive in understanding how to utilize machine output to provide a certification of the result.

In Ref. 53, the authors used reinforcement learning to train AI to play Bell nonlocal games and obtain optimal (or near-optimal) performance. The agent is offered a reward at the end of each epoch, which is equal to the expectation value of the Bell operator corresponding to the state and measurement settings chosen by the agent. Such a reward scheme is sparse, and hence, it might not be scalable. It would be interesting to come up with better reward schemes. Furthermore, in this approach, only a single agent tries to learn the optimization landscape and discovers near-optimal (or optimal) measurement settings and state. It would be exciting to extend the approach to multiagent setting where every spacelike separated party is considered a separate agent. It is worth mentioning that distributing the actions and observations of a single agent into a list of agents reduces the dimensionality of agent inputs and outputs. Furthermore, it also dramatically improves the amount of training data produced per step of the environment. Agents learn better if they tend to interact as compared to the case of solitary learning.

Bell inequalities separate the set of local behaviors from the set of nonlocal behaviors. The analogous boundary separating quantum from the postquantum set is known as quantum Bell inequality.^{161,162} Finding quantum Bell inequalities is an interesting and challenging problem. However, one can aim to obtain the approximate expression by supervised learning with the quantum Bell inequalities being the boundary separating the quantum set

from the postquantum set. Moreover, it is interesting to see if it is possible to guess physical principles by merely opening the neural-network black box.

Driven by the success of machine learning in Bell nonlocality, it is genuine to ask if the methods could be useful to solve problems in quantum steering and contextuality. Recently, ideas from the exclusivity graph approach to contextuality were used to investigate problems involving causal inference.¹⁶³ Ideas from quantum foundations could further assist in developing a deeper understanding of machine learning or in general artificial intelligence.

In artificial intelligence, one of the tests to distinguish between humans and machines is the famous “turing test (TT)” due to Alan Turing.^{62,164} The purpose of TT is to determine if a computer is linguistically distinguishable from a human. In TT, a human and a machine are sealed in different rooms. A human jury who does not know which room contains a human and which room not, asks questions to them, by email, for example. Based on the returned outcome, if the judge cannot do better than 50–50, then the machine in question is said to have passed TT. The task of distinguishing the humans from machine based on the statistics of the answers (say output a) given questions (say input x) is a statistical distinguishability test assuming the rooms plus its inhabitants as black boxes. In the black-box approach to quantum theory, experiments are regarded as a black box where the experimentalist introduces a measurement (input) and obtains the outcome of the measurement (output). One of the central goals of this approach is to deduce statements regarding the contents of the black box based on input-output statistics.⁶⁸ It would be nice to see if techniques from the black-box approach to quantum theory could be connected to TT.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of the Ministry of Education and the National Research Foundation, Singapore. The authors thank Valerio Scarani for valuable discussions.

DATA AVAILABILITY

Data sharing is not applicable to this survey as no new data were created or analyzed in this study.

REFERENCES

- ¹S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms* (Cambridge University, Cambridge, England, 2014).
- ²I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT, Cambridge, MA, 2016).
- ³M. AlQuraishi, *Bioinformatics* **35**, 4862 (2019).
- ⁴J. P. Bridge, S. B. Holden, and L. C. Paulson, *J. Autom. Reason.* **53**, 141 (2014).
- ⁵A. Lavecchia, *Drug Discovery Today* **20**, 318 (2015).
- ⁶L.-F. Arsenault, O. Anatole von Lilienfeld, and A. J. Millis, “Machine learning for many-body physics: efficient solution of dynamical mean-field theory,” preprint [arXiv:1506.08858](https://arxiv.org/abs/1506.08858) (2015).
- ⁷Y. Zhang and E.-A. Kim, *Phys. Rev. Lett.* **118**, 216401 (2017).
- ⁸J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
- ⁹E. P. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Nat. Phys.* **13**, 435 (2017).
- ¹⁰D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. B* **96**, 195145 (2017).
- ¹¹L. Wang, *Phys. Rev. B* **94**, 195105 (2016).
- ¹²P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, *Sci. Rep.* **7**, 1 (2017).

- ¹³K. Chng, J. Carrasquilla, R. G. Melko, and E. Khatami, *Phys. Rev. X* **7**, 031038 (2017).
- ¹⁴Y. Zhang, R. G. Melko, and E.-A. Kim, *Phys. Rev. B* **96**, 245119 (2017).
- ¹⁵S. J. Wetzel, *Phys. Rev. E* **96**, 022140 (2017).
- ¹⁶W. Hu, R. R. Singh, and R. T. Scalettar, *Phys. Rev. E* **95**, 062122 (2017).
- ¹⁷N. Yoshioka, Y. Akagi, and H. Katsura, *Phys. Rev. B* **97**, 205110 (2018).
- ¹⁸G. Torlai and R. G. Melko, *Phys. Rev. B* **94**, 165134 (2016).
- ¹⁹H.-Y. Huang, K. Bharti, and P. Rebentrost, "Near-term quantum algorithms for linear systems of equations," preprint [arXiv:1909.07344](https://arxiv.org/abs/1909.07344) (2019).
- ²⁰K.-I. Aoki and T. Kobayashi, *Mod. Phys. Lett. B* **30**, 1650401 (2016).
- ²¹Y.-Z. You, Z. Yang, and X.-L. Qi, *Phys. Rev. B* **97**, 045153 (2018).
- ²²M. Pasquato, "Detecting intermediate mass black holes in globular clusters with machine learning," preprint [arXiv:1606.08548](https://arxiv.org/abs/1606.08548) (2016).
- ²³Y. D. Hezaveh, L. P. Levasseur, and P. J. Marshall, *Nature* **548**, 555 (2017).
- ²⁴R. Biswas *et al.*, *Phys. Rev. D* **88**, 062003 (2013).
- ²⁵B. P. Abbott *et al.*, *Phys. Rev. Lett.* **116**, 061102 (2016).
- ²⁶S. V. Kalinin, B. G. Sumpter, and R. K. Archibald, *Nat. Mater.* **14**, 973 (2015).
- ²⁷S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, *Nat. Phys.* **12**, 469 (2016).
- ²⁸J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, *Phys. Rev. B* **95**, 041101 (2017).
- ²⁹L. Huang and L. Wang, *Phys. Rev. B* **95**, 035105 (2017).
- ³⁰G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nat. Phys.* **14**, 447 (2018).
- ³¹J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, *Phys. Rev. B* **97**, 085104 (2018).
- ³²Y. Huang and J. E. Moore, "Neural network representation of tensor network and chiral states," preprint [arXiv:1701.06246](https://arxiv.org/abs/1701.06246) (2017).
- ³³F. Schindler, N. Regnault, and T. Neupert, *Phys. Rev. B* **95**, 245134 (2017).
- ³⁴T. Haug, R. Dumke, L.-C. Kwek, C. Miniatura, and L. Amico, "Engineering quantum current states with machine learning," preprint [arXiv:1911.09578](https://arxiv.org/abs/1911.09578) (2019).
- ³⁵Z. Cai and J. Liu, *Phys. Rev. B* **97**, 035116 (2018).
- ³⁶P. Broecker, F. F. Assaad, and S. Trebst, "Quantum phase recognition via unsupervised machine learning," preprint [arXiv:1707.00663](https://arxiv.org/abs/1707.00663) (2017).
- ³⁷Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, *Phys. Rev. B* **96**, 205152 (2017).
- ³⁸J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, *Nature* **549**, 195 (2017).
- ³⁹T. Haug, W.-K. Mok, J.-B. You, W. Zhang, C. E. Png, and L.-C. Kwek, "Classifying global state preparation via deep reinforcement learning," preprint [arXiv:2005.12759](https://arxiv.org/abs/2005.12759) (2020).
- ⁴⁰G. Torlai and R. G. Melko, *Phys. Rev. Lett.* **119**, 030501 (2017).
- ⁴¹M. Bukov, *Phys. Rev. B* **98**, 224305 (2018).
- ⁴²M. Bukov, A. G. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, *Phys. Rev. X* **8**, 031086 (2018).
- ⁴³K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, *Phys. Rev. D* **98**, 046019 (2018).
- ⁴⁴J. Carifio, J. Halverson, D. Krioukov, and B. D. Nelson, *J. High Energy Phys.* **2017**, 157.
- ⁴⁵H. W. Lin, M. Tegmark, and D. Rolnick, *J. Stat. Phys.* **168**, 1223 (2017).
- ⁴⁶A. Cichocki, "Tensor networks for big data analytics and large-scale optimization problems," preprint [arXiv:1407.3124](https://arxiv.org/abs/1407.3124) (2014).
- ⁴⁷G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Rev. Mod. Phys.* **91**, 045002 (2019).
- ⁴⁸V. Dunjko, J. M. Taylor, and H. J. Briegel, *Phys. Rev. Lett.* **117**, 130501 (2016).
- ⁴⁹M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, *Quantum Sci. Technol.* **4**, 043001 (2019).
- ⁵⁰L. Hardy and R. Spekkens, "Why physics needs quantum foundations," preprint [arXiv:1003.5008](https://arxiv.org/abs/1003.5008) (2010).
- ⁵¹T. Kriváček, Y. Cai, D. Cavalcanti, A. Tavakoli, N. Gisin, and N. Brunner, "A neural network oracle for quantum nonlocality problems in networks," preprint [arXiv:1907.10552](https://arxiv.org/abs/1907.10552) (2019).
- ⁵²A. Canabarro, S. Brito, and R. Chaves, *Phys. Rev. Lett.* **122**, 200401 (2019).
- ⁵³K. Bharti, T. Haug, V. Vedral, and L.-C. Kwek, "How to teach ai to play bell non-local games: Reinforcement learning," preprint [arXiv:1912.10783](https://arxiv.org/abs/1912.10783) (2019).
- ⁵⁴D.-L. Deng, *Phys. Rev. Lett.* **120**, 240402 (2018).
- ⁵⁵R. Iten, T. Metger, H. Wilming, L. Del Rio, and R. Renner, *Phys. Rev. Lett.* **124**, 010508 (2020).
- ⁵⁶T. M. Mitchell *et al.*, *Mach. Learn.* **45**(37), 870 (1997).
- ⁵⁷D. Silver *et al.*, *Nature* **550**, 354 (2017).
- ⁵⁸G. E. Hinton, "A practical guide to training restricted Boltzmann machines," in *Neural Networks: Tricks of the Trade* (Springer, Berlin, Germany, 2012), pp. 599–619.
- ⁵⁹G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- ⁶⁰J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, *AI Mag.* **27**, 1212 (2006).
- ⁶¹A. M. Turing, "Computing machinery and intelligence," in *Parsing the Turing Test* (Springer, Berlin, Germany, 2009), p. 23.
- ⁶²S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- ⁶³C. A. Fuchs and R. Schack, *Found. Phys.* **41**, 345 (2011).
- ⁶⁴H. Everett III, *Rev. Mod. Phys.* **29**, 454 (1957).
- ⁶⁵B. S. DeWitt and N. Graham, *The Many Worlds Interpretation of Quantum Mechanics* (Princeton University, Princeton, NJ, 2015).
- ⁶⁶R. Omnes, *Rev. Mod. Phys.* **64**, 339 (1992).
- ⁶⁷J. Barrett, *Phys. Rev. A* **75**, 032304 (2007).
- ⁶⁸A. Acín and M. Navascués, "Black box quantum mechanics," in *Quantum [Un] Speakables II* (Springer, Berlin, Germany, 2017), pp. 307–319.
- ⁶⁹W. Van Dam, *Nat. Comput.* **12**, 9 (2013).
- ⁷⁰G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, *Phys. Rev. Lett.* **96**, 250401 (2006).
- ⁷¹M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, *Nature* **461**, 1101 (2009).
- ⁷²M. Navascués and H. Wunderlich, *Proc. R. Soc. A* **466**, 881 (2010).
- ⁷³N. Linden, S. Popescu, A. J. Short, and A. Winter, *Phys. Rev. Lett.* **99**, 180502 (2007).
- ⁷⁴B. Yan, *Phys. Rev. Lett.* **110**, 260406 (2013).
- ⁷⁵T. Fritz, A. Belén Sainz, R. Augusiak, J. Bohr Brask, R. Chaves, A. Leverrier, and A. Acín, *Nat. Commun.* **4**, 1 (2013).
- ⁷⁶K. J. McQueen, *Stud. Hist. Philos. Sci., Part B* **49**, 10 (2015).
- ⁷⁷R. D. Sorkin, *Mod. Phys. Lett. A* **9**, 3119 (1994).
- ⁷⁸O. Oreshkov, F. Costa, and C. Brukner, *Nat. Commun.* **3**, 1 (2012).
- ⁷⁹J.-M. Raimond, M. Brune, and S. Haroche, *Rev. Mod. Phys.* **73**, 565 (2001).
- ⁸⁰R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- ⁸¹W. K. Wootters, *Quantum Inf. Comput.* **1**, 27 (2001).
- ⁸²M. Horodecki, P. Horodecki, and R. Horodecki, *Phys. Lett. A* **223**, 1 (1996).
- ⁸³B. M. Terhal, *Theor. Comput. Sci.* **287**, 313 (2002).
- ⁸⁴P. Krammer, H. Kampermann, D. Bruß, R. A. Bertlmann, L. C. Kwek, and C. Macchiavello, *Phys. Rev. Lett.* **103**, 100502 (2009).
- ⁸⁵A. K. Ekert, C. M. Alves, D. K. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, *Phys. Rev. Lett.* **88**, 217901 (2002).
- ⁸⁶J. Pearl, *Causality* (Cambridge University, Cambridge, 2000).
- ⁸⁷S. J. Freedman and J. F. Clauser, *Phys. Rev. Lett.* **28**, 938 (1972).
- ⁸⁸C. A. Kocher and E. D. Commins, *Phys. Rev. Lett.* **18**, 575 (1967).
- ⁸⁹A. Aspect, *J. Phys. Coll.* **42**, C2-63 (1981).
- ⁹⁰A. Aspect, J. Dalibard, and G. Roger, *Phys. Rev. Lett.* **49**, 1804 (1982).
- ⁹¹C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* **70**, 1895 (1993).
- ⁹²A. K. Ekert, *Phys. Rev. Lett.* **67**, 661 (1991).
- ⁹³C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, *Phys. Rev. Lett.* **76**, 722 (1996).
- ⁹⁴C. H. Bennett and S. J. Wiesner, *Phys. Rev. Lett.* **69**, 2881 (1992).
- ⁹⁵D. Deutsch, *Proc. R. Soc. London, A* **400**, 97 (1818–1985).
- ⁹⁶R. P. Feynman, *Feynman Lectures on Computation* (CRC, Boca Raton, FL, 2018).
- ⁹⁷J. S. Bell, *Phys. Phys. Fiz.* **1**, 195 (1964).
- ⁹⁸N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Rev. Mod. Phys.* **86**, 419 (2014).

- ⁹⁹S. Pironio *et al.*, *Nature* **464**, 1021 (2010).
- ¹⁰⁰D. Mayers and A. Yao, *Quantum Inf. Comput.* **4**, 273 (2004).
- ¹⁰¹J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
- ¹⁰²S. Kochen and E. P. Specker, *J. Math. Mech.* **17**, 59 (1967).
- ¹⁰³N. David Mermin, *Rev. Mod. Phys.* **65**, 803 (1993).
- ¹⁰⁴A. Cabello, S. Severini, and A. Winter, *Phys. Rev. Lett.* **112**, 040401 (2014).
- ¹⁰⁵B. Amaral and M. T. Cunha, *On Graph Approaches to Contextuality and Their Role in Quantum Theory* (Springer, Cham, 2018).
- ¹⁰⁶A. Cabello, V. D'Ambrosio, E. Nagali, and F. Sciarrino, *Phys. Rev. A* **84**, 030302 (2011).
- ¹⁰⁷K. Bharti, M. Ray, and L.-C. Kwek, *Entropy* **21**, 134 (2019).
- ¹⁰⁸R. Raussendorf, *Phys. Rev. A* **88**, 022322 (2013).
- ¹⁰⁹K. Bharti, M. Ray, A. Varvitsiotis, N. A. Warsi, A. Cabello, and L.-C. Kwek, *Phys. Rev. Lett.* **122**, 250403 (2019).
- ¹¹⁰S. Mansfield and E. Kashefi, *Phys. Rev. Lett.* **121**, 230401 (2018).
- ¹¹¹D. Saha, P. Horodecki, and M. Pawłowski, *New J. Phys.* **21**, 093057 (2019).
- ¹¹²K. Bharti, M. Ray, V. Antonios, A. Cabello, and L.-C. Kwek, preprint [arXiv:1911.09448](https://arxiv.org/abs/1911.09448) (2019).
- ¹¹³N. Delfosse, P. Allard Guerin, J. Bian, and R. Raussendorf, *Phys. Rev. X* **5**, 021003 (2015).
- ¹¹⁴J. Singh *et al.*, *Phys. Rev. A* **95**, 062333 (2017).
- ¹¹⁵H. Pashayan, J. J. Wallman, and S. D. Bartlett, *Phys. Rev. Lett.* **115**, 070501 (2015).
- ¹¹⁶K. Bharti, A. S. Arora, L. C. Kwek, and J. Roland, "A simple proof of uniqueness of the KCBS inequality," preprint [arXiv:1811.05294](https://arxiv.org/abs/1811.05294) (2018).
- ¹¹⁷J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf, *Phys. Rev. Lett.* **119**, 120505 (2017).
- ¹¹⁸L. Catani and D. E. Browne, *Phys. Rev. A* **98**, 052108 (2018).
- ¹¹⁹A. Singh Arora *et al.*, *Phys. Lett. A* **383**, 833 (2019).
- ¹²⁰M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Nature* **510**, 351 (2014).
- ¹²¹S. Abramsky and A. Brandenburger, *New J. Phys.* **13**(11), 113036 (2011).
- ¹²²A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, *Commun. Math. Phys.* **334**, 533 (2015).
- ¹²³E. N. Dzhafarov, V. H. Cervantes, and J. V. Kujala, *Philos. Trans. R. Soc., A* **375**, 20160389 (2017).
- ¹²⁴E. N. Dzhafarov, *Philos. Trans. R. Soc., A* **377**, 20190144 (2019).
- ¹²⁵J. V. Kujala and E. N. Dzhafarov, *Philos. Trans. R. Soc., A* **377**, 20190149 (2019).
- ¹²⁶R. W. Spekkens, *Phys. Rev. A* **71**, 052108 (2005).
- ¹²⁷M. T. Quintino, T. Vértesi, D. Cavalcanti, R. Augusiak, M. Demianowicz, A. Acín, and N. Brunner, *Phys. Rev. A* **92**, 032107 (2015).
- ¹²⁸H. M. Wiseman, S. J. Jones, and A. C. Doherty, *Phys. Rev. Lett.* **98**, 140402 (2007).
- ¹²⁹E. Schrödinger, "Discussion of probability relations between separated systems," in *Mathematical Proceedings of the Cambridge Philosophical Society* (Cambridge University, Cambridge, England, 1935), Vol. 31, pp. 555–563.
- ¹³⁰S. G. de Aguiar Brito, B. Amaral, and R. Chaves, *Phys. Rev. A* **97**, 022111 (2018).
- ¹³¹A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, *Phys. Rev. A* **90**, 062109 (2014).
- ¹³²C. Branciard, N. Gisin, and S. Pironio, *Phys. Rev. Lett.* **104**, 170401 (2010).
- ¹³³M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, *Phys. Rev. Lett.* **123**, 140401 (2019).
- ¹³⁴N. Gisin, *Entropy* **21**, 325 (2019).
- ¹³⁵H. Saito and M. Kato, *J. Phys. Soc. Jpn.* **87**, 014001 (2018).
- ¹³⁶X. Gao and L.-M. Duan, *Nat. Commun.* **8**, 1 (2017).
- ¹³⁷U. Schollwöck, *Rev. Mod. Phys.* **77**, 259 (2005).
- ¹³⁸F. Verstraete, V. Murg, and J. Ignacio Cirac, *Adv. Phys.* **57**, 143 (2008).
- ¹³⁹G. Vidal, *Phys. Rev. Lett.* **101**, 110501 (2008).
- ¹⁴⁰Y.-C. Ma and M.-H. Yung, *npj Quantum Inf.* **4**, 1 (2018).
- ¹⁴¹R. F. Werner, *Phys. Rev. A* **40**, 4277 (1989).
- ¹⁴²C. Harney, S. Pirandola, A. Ferraro, and M. Paternostro, "Entanglement classification via neural network quantum states," *New J. Phys.* **22**(4), 045001 (2020).
- ¹⁴³S. Lu *et al.*, *Phys. Rev. A* **98**, 012315 (2018).
- ¹⁴⁴C. B. Goes, A. Canabarro, E. I. Duzzioni, and T. O. Maciel, "Automated machine learning can classify bound entangled states with tomograms," preprint [arXiv:2001.08118](https://arxiv.org/abs/2001.08118) (2020).
- ¹⁴⁵S. Weinstein, "Neural networks as 'hidden' variable models for quantum systems," preprint [arXiv:1807.03910](https://arxiv.org/abs/1807.03910) (2018).
- ¹⁴⁶S. Weinstein, *Found. Phys.* **39**, 921 (2009).
- ¹⁴⁷M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers* (Springer, Berlin, Germany, 2018), Vol. 17.
- ¹⁴⁸M. Schuld, I. Sinayskiy, and F. Petruccione, *Contemp. Phys.* **56**, 172 (2015).
- ¹⁴⁹V. Dunjko and P. Wittek, *Quantum Views* **4**, 32 (2020).
- ¹⁵⁰A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, *Proc. Natl. Acad. Sci.* **115**, 1221 (2018).
- ¹⁵¹M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, *Phys. Rev. Lett.* **116**, 090405 (2016).
- ¹⁵²X. Gu, M. Krenn, M. Erhard, and A. Zeilinger, *Phys. Rev. Lett.* **120**, 103601 (2018).
- ¹⁵³F. Wang, M. Erhard, A. Babazadeh, M. Malik, M. Krenn, and A. Zeilinger, *Optica* **4**, 1462 (2017).
- ¹⁵⁴A. Babazadeh, M. Erhard, F. Wang, M. Malik, R. Nouroozi, M. Krenn, and A. Zeilinger, *Phys. Rev. Lett.* **119**, 180510 (2017).
- ¹⁵⁵F. Schleeder, M. Krenn, R. Fickler, M. Malik, and A. Zeilinger, *New J. Phys.* **18**, 043019 (2016).
- ¹⁵⁶M. Erhard, M. Malik, M. Krenn, and A. Zeilinger, *Nat. Photonics* **12**(12), 759 (2018).
- ¹⁵⁷M. Malik, M. Erhard, M. Huber, M. Krenn, R. Fickler, and A. Zeilinger, *Nat. Photonics* **10**(4), 248 (2016).
- ¹⁵⁸M. Pavičić, M. Waegell, N. D. Megill, and P. K. Aravind, *Sci. Rep.* **9**(1), 1 (2019).
- ¹⁵⁹J. Tura, R. Augusiak, A. B. Sainz, T. Vértesi, M. Lewenstein, and A. Acín, *Science* **344**, 1256 (2014).
- ¹⁶⁰Z. Ghahramani, *Nature* **521**, 452 (2015).
- ¹⁶¹L. Masanes, "Extremal quantum correlations for n parties with two dichotomic observables per site," preprint [arXiv:quant-ph/0512100](https://arxiv.org/abs/quant-ph/0512100) (2005).
- ¹⁶²T. Le Phuc, "Computing quantum bell inequalities," preprint [arXiv:1909.05472](https://arxiv.org/abs/1909.05472) (2019).
- ¹⁶³D. Poderini, R. Chaves, I. Agresti, G. Carvacho, and F. Sciarrino, "Exclusivity graph approach to instrumental inequalities," preprint [arXiv:1909.09120](https://arxiv.org/abs/1909.09120) (2019).
- ¹⁶⁴A. Pinar Saygin, I. Cicekli, and V. Akman, *Minds Machines* **10**, 463 (2000).