Ktokolwiek widział, ktokolwiek wie! Ukradziono Nagrodę Nobla z fizyki! Beware! The Nobel Prize in physics got stolen!

Anna Dawid*

⟨aQa^L⟩ Applied Quantum Algorithms – Lorentz Insitute for Theoretical Physics
& Leiden Institute of Advanced Computer Science, Uniwersytet w Lejdzie, Holandia

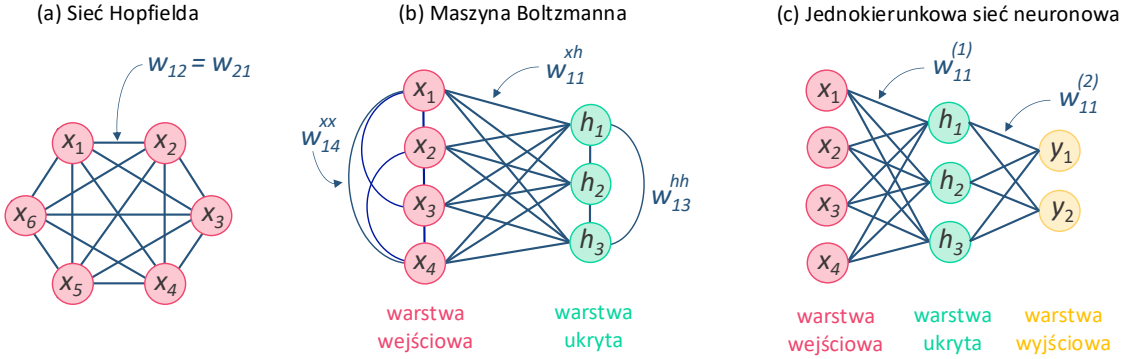
Abstrakt. Nagroda Nobla z fizyki w 2024 została przyznana Johnowi J. Hopfieldowi z Uniwersytetu w Princeton i Geoffrey'owi E. Hintonowi z Uniwersytetu w Toronto za *fundamentalne odkrycia i wynalazki umożliwiające uczenie maszynowe przy użyciu sztucznych sieci neuronowych*. Choć wywołała kontrowersje wśród naukowców, to wpisuje się ona w trend nagradzania twórców nowych przyrządów do badania świata, a takim staje się ostatnio uczenie maszynowe. Opisuję w tym artykule dokonania noblistów, w szczególności sieć Hopfielda i maszynę Boltzmanna i wyjaśniam, jak różnią się od współczesnego paradygmatu uczenia maszynowego. Zwracam też uwagę na ograniczenia sieci neuronowych, a także ekscytujący dwukierunkowy wpływ, jaki wciąż mają na siebie nawzajem uczenie maszynowe i fizyka.
Słowa kluczowe: 2024 Nagroda Nobla z fizyki w 2024, uczenie maszynowe, sieci neuronowe, sieć Hopfielda, maszyna Boltzmanna

Abstract. The 2024 Nobel Prize in Physics was awarded to John J. Hopfield of Princeton University and Geoffrey E. Hinton of the University of Toronto for *fundamental discoveries and inventions that enable machine learning using artificial neural networks*. Although controversial among scientists, the award is part of a trend of rewarding creators of new devices for studying the world, and machine learning has recently become such a device. In this article, I describe the achievements of the Nobel Prize winners, in particular the Hopfield network and the Boltzmann machine, and explain how they differ from the modern paradigm of machine learning. I also describe the limitations of current neural networks, as well as the exciting bidirectional influence that machine learning and physics continue to have on each other.
Keywords: Nobel Prize in Physics, machine learning, neural networks, Hopfield network, Boltzmann machine

„Ukradziona" Nagroda Nobla. Dawno Nagroda Nobla z fizyki nie wywołała takiego poruszenia jak ta ogłoszona 8 października 2024. Wówczas poznaliśmy werdykt Królewskiej Szwedzkiej Akademii Nauk, która przyznała tę najbardziej prestiżową z nagród Johnowi J. Hopfieldowi z Uniwersytetu w Princeton i Geoffrey'owi E. Hintonowi z Uniwersytetu w Toronto *za fundamentalne odkrycia i wynalazki umożliwiające uczenie maszynowe przy użyciu sztucznych sieci neuronowych*. Nawiasem mówiąc, dzień później połowa Nagrody Nobla z chemii przypadła Demisowi Hassabisowi i Johnowi Jumperowi z firmy Google DeepMind za przewidywania struktury białek, umacniając dominację sieci neuronowych w nauce. Jak to się stało, że informatyka „ukradła" Nagrodę Nobla fizyce?

Nowe narzędzie do badania świata. Uczenie maszynowe to nowe podejście do starych i nowych problemów, będące nowym spojrzeniem na posiadane informacje, potencjalnie pozwalającym uniknąć ludzkich uprzedzeń. Gdy spojrzymy na nie jak na nowe narzędzie do badania naszej rzeczywistości, natychmiast zobaczymy, że Nobel 2024 wpisuje się w trend nagradzania budowania nowych przyrządów pomiarowych. W 1907 Nagroda Nobla z fizyki została przyznana za konstrukcję interferometru, w 1924 nagrodzono wysiłki prowadzące do spektroskopii rentgenowskiej, a w roku 1986 nagrodzone zostały projekty mikroskopu elektronowego i skaningowego mikroskopu tunelowego, nie wspominając już o wynalezieniu tranzystora (Nobel 1956) i układu scalonego (Nobel 2000). Jeden z noblistów z fizyki, Steven Chu (nagrodzony w 1997 za rozwój metod chłodzenia atomów do temperatur bliskich zera bezwzględnego) powiedział nawet, że najpewniejszy sposób na Nobla to zrobienie nowego przyrządu pomiarowego. Porównał to do zaglądania po raz pierwszy pod nowy kamień – z dużym prawdopodobieństwem znajdziecie tam coś ciekawego!

*ORCID: 0000-0001-9498-1732



Rys. 1. Przykłady (a) sieci Hopfielda, (b) maszyny Boltzmanna i (c) prostej sieci neuronowej

Modele Hopfielda i Boltzmanna a sieci neuronowe. Tegoroczni laureaci Nagrody Nobla z fizyki postawili kamienie węgielne pod fundament współczesnego uczenia maszynowego. Co więcej, zrobili to metodami wywodzącymi się z fizyki statystycznej. Zaproponowali i rozwinęli sieci, inspirowane modelami spinowymi w fizyce, np. modelem Isinga, których *uczenie* polegało na minimalizacji energii. To podejście oparte na energii różni się od współcześnie używanych najpopularniejszych sieci neuronowych, które trenuje się przez minimalizację błędu, jaki popełniają w danym zadaniu. Modele noblistów przygotowały jednak grunt pod współczesne sieci (ponadto Hinton zaproponował wiele skutecznych heurystyk wykorzystywanych w treningu sieci po dziś dzień). Modele bazujące na energii są również stosowane w fizyce kwantowej jako reprezentacja stanów kwantowych. Ponadto, niektórzy czołowi specjaliści w uczeniu maszynowym jak Yann LeCun postulują powrót do korzeni i porzucenie współczesnego paradygmatu na rzecz modeli opartych na energii [1]. Prace noblistów są więc dalej istotne w szybko zmieniającym się krajobrazie uczenia maszynowego; także w kolejnych paragrafach spojrzymy na ich dokonania nieco dokładniej.

Sieć Hopfielda. Fizyka statystyczna i modele spinowe wkroczyły do uczenia maszynowego po raz pierwszy jako sieci Hopfielda. Sieci Hopfielda to w pełni połączone sieci rekurencyjne zaproponowane przez Williama A. Little'a w 1974 [2] i rozwinięte przez Hopfielda w 1982 [3]. Schemat takiej sieci przedstawiono na rys. 1(a). Każdy węzeł tej sieci, zwany też neuronem, niczym spin 1/2, przyjmuje binarne wartości, $x_i \in \{-1, 1\}$, zwane aktywacjami. Węzły tej sieci są połączone w sposób symetryczny, $w_{ij} = w_{ji}$, a wagi tych połączeń są rzeczywiste. Danej konfiguracji (wartości aktywacji i wag) sieci Hopfielda można przypisać energię

$$F_{\boldsymbol{w}}(\boldsymbol{x}) = -\sum_{ij} x_i w_{ij} x_j \,. \tag{1}$$

Little zauważył, że taka sieć może przechowywać (zapamiętywać) wzorce. Jej trening polega na utrwaleniu m pożądanych wzorców binarnych o wymiarze d, $\boldsymbol{X}^{\text{trening}} = \{\boldsymbol{x}^{\text{trening},\,1}, \ldots, \boldsymbol{x}^{\text{trening},\,m}\}$, gdzie $\boldsymbol{x}^{\text{trening},\,i} = \{x_1^{\text{trening},\,i}, \ldots, x_d^{\text{trening},\,i}\}$, przez uczynienie ich minimami energetycznymi sieci. Innymi słowy, sieć zmienia wagi swoich połączeń w taki sposób, żeby zminimalizować energie odpowiadające wzorcom treningowym poprzez zasadę uczenia hebbiańskiego: $w_{ij} \leftarrow w_{ij} + y_i^{\text{trening}} y_j^{\text{trening}}$. Uczenie hebbiańskie, zaobserwowane w biologicznych sieciach neuronowych, polega na wzmacnianiu połączeń między tymi neuronami, które są jednocześnie aktywne. Końcowe wagi wytrenowanej sieci oznaczamy $\check{\boldsymbol{w}}$. Na etapie przewidywania, sieć jest w stanie wskazać, do którego wzorca dana wejściowa jest najbardziej podobna, poprzez sprawdzenie energii odpowiadającej danej wejściowej i znalezienie najbliższego minimum energetycznego. Bardziej precyzyjnie, sieć odnajduje wzorzec poprzez wielokrotną aktualizację stanów neuronów w zależności od wartości wszystkich innych neuronów: $x_i \longleftarrow \text{sign}(\sum_j \check{w}_{ij} x_j)$. W ramach tej procedury $\boldsymbol{x} = \{x_i\}$ zbiegają do lokalnego minimum funkcji energii i odtwarzają odpowiadający zapamiętany wzorzec. Hopfield pokazał, że taka sieć może działać w praktyce oraz zbadał jej pojemność (pamięć) stosując metodę pola średniego dla szkła spinowego [3]. Stosowanie w praktyce sieci Hopfielda ograniczone jest jej niewielką pojemnością, a także zdarzającym się czasem przypadkowym wytwarzaniem w trakcie treningu tzw. minimów (wzorców) pozornych, tj. przypadkowych minimów energii nieukształtowanych przez dane treningowe, które można znaleźć podczas predykcji.

Maszyna Boltzmanna. Rok po pracy Hopfielda, w 1983, Hinton i Terrence Sejnowski zaproponowali rozszerzenie sieci Hopfielda, zwane maszyną Boltzmanna [4], wprowadzając do modelu neurony zwane *ukrytymi* (ang. *hidden*), jak przedstawiono na rys. 1(b). To rozszerzenie stało się ważne dla całej dziedziny uczenia maszynowego, ponieważ było pierwszym wprowadzeniem warstw neuronów ukrytych, absolutnie kluczowych dla sukcesu współczesnych sieci neuronowych. Funkcja energii maszyny Boltzmanna jest nieco bardziej skomplikowana niż sieci Hopfielda, gdyż jest rozkładem brzego-

wym uwzględniających różne wartości neuronów ukrytych

$$E_{\boldsymbol{w}}(\boldsymbol{x},\boldsymbol{z}) = -\sum_{ij} x_i w_{ij}^{xx} x_j - \sum_{ij} z_i w_{ij}^{zz} z_j - \sum_{ij} x_i w_{ij}^{xz} z_j\,, \tag{2}$$

$$F_{\boldsymbol{w}}(\boldsymbol{x}) = -\log \sum_{\boldsymbol{z}} \exp\left[-E_{\boldsymbol{w}}(\boldsymbol{x},\boldsymbol{z})\right]. \tag{3}$$

Poza obecnością węzłów ukrytych, maszyna Boltzmanna różni się też od sieci Hopfielda sposobem trenowania. Dalej minimalizowana jest energia dla danych treningowych, ale dodatkowo minimalizowana jest energia próbek „kontrastowych", generowanych z maszyny Boltzmanna za pomocą próbkowania Monte Carlo łańcuchami Markowa. W efekcie, maszyna Boltzmanna nie zapamiętuje konkretnych wzorców jak sieć Hopfielda, ale stara się odtworzyć rozkład prawdopodobieństwa, z którego pochodzą wzorce treningowe. Zmniejsza to ryzyko wytworzenia minimów pozornych, ale generuje duże koszty obliczeniowe, co spowodowało, że współcześnie są rzadziej stosowane.

Głębokie sieci neuronowe. Współczesne sieci neuronowe są często jednokierunkowe (z elementami rekurencyjnymi) i nie bazują na minimalizacji energii. Prosty przykład współczesnego paradygmatu obrazuje rys. 1(c). Taka jednokierunkowa sieć ma warstwę wyjściową, w której podaje swoje odpowiedzi. Mówimy, że jest probabilistyczna, bo jej warstwa wyjściowa zawiera wszystkie możliwe odpowiedzi i w związku z tym normalizujemy odpowiadające im wartości y_i, by odpowiadały prawdopodobieństwu danej odpowiedzi. Jej trening odbywa się poprzez minimalizację błędu, czyli zmienianiu parametrów sieci w taki sposób, żeby sieć w warstwie wyjściowej przypisywała jak największe prawdopodobieństwo oczekiwanej odpowiedzi. Rozpatrzmy przykład, aby unaocznić różnicę w działaniu modeli bazujących na minimalizacji energii i sieci opartych na minimalizacji błędu. Standardowym zadaniem jest rozpoznawanie, czy na obrazku znajduje się kot czy pies. W warstwie wejściowej sieci jednokierunkowej umieszczamy wówczas obrazek i zmieniamy wagi sieci tak, by w warstwie wyjściowej pierwszy neuron miał większą (mniejszą) wartość niż drugi, jeśli na obrazku jest kot (pies). W przypadku modeli bazujących na energii, takich jak sieć Hopfielda, jeśli na wejściu umieścimy obraz kota, to w wyniku aktualizacji pikseli, sieć sprowadziłaby obrazek do zapamiętanego wzorca kota i w taki sposób odpowiedziała na pytanie. Tutaj warto wspomnieć, że Hinton miał również wielki wkład w rozwój głębokiego uczenia maszynowego, w szczególności zaproponował działający sposób treningu głębokich sieci neuronowych (nawiasem mówiąc, użył do tego maszyn Boltzmanna!), czym poszerzył w latach 2000. drogę do nowoczesnego uczenia maszynowego.

Uczenie maszynowe dla fizyki. Mimo że rzadziej stosowane w problemach inżynieryjnych, modele-dzieci sieci Hopfielda i maszyny Boltzmanna z sukcesami służą fizyce kwantowej [5]. W tym momencie są one najbardziej obiecującym podejściem do szukania stanów podstawowych największych i najtrudniejszych układów kwantowych (dwu- i trójwymiarowych o wysokim poziomie korelacji między cząstkami kwantowymi) [6]. Bardziej ogólnie, sieci neuronowe celują w szukaniu wzorców w danych, więc stosowane są w detekcji faz materii i do wykrywania nowych ciekawych zachowań układów, ale także w CERN, gdzie szukają anomalii (w których może kryć się np. nowa cząstka elementarna), a także w astrofizyce, gdzie mają pomóc zrozumieć podstawowe prawa rządzące formacją gwiazd i galaktyk [7].

Fizyka dla uczenia maszynowego. Interakcja fizyki i uczenia maszynowego działa w obie strony, w szczególności fizyka nie tylko położyła kamienie węgielne pod uczenie maszynowe, ale i dalej je współcześnie rozwija. Fizycy statystyczni, np. tacy jak Lenka Zdeborová [8], próbują rozwiązać fundamentalne zagadki nowoczesnych sieci neuronowych, takie jak ich zdolność do uogólniania i unikania przeuczenia polegającego na głupim zapamiętywaniu każdego przykładu, przy ogromnym przeparametryzowaniu, które teoretycznie umożliwia takie przeuczenie. Fizycy kwantowi jak Maria Schuld szukają sposobów, żeby połączyć modele uczenia maszynowego z kwantowym przetwarzaniem informacji [9]. Poza tym problemy z różnych gałęzi fizyki są źródłem ciekawych, dobrze poznanych danych, które pomagają wyłapywać ograniczenia sieci neuronowych i lepiej zrozumieć ich działanie [10].

Nobel za wcześnie? Uczenie maszynowe bez wątpienia jest nowym obiecującym narzędziem, które, podążając za analogią Stevena Chu, pozwala zajrzeć pod nowe kamienie. Ale czy faktycznie zobaczyliśmy coś ciekawego pod jednym z tych kamieni? Wydaje się, że do tej pory Nagrody Nobla za nowe przyrządy pomiarowe przyznawane były już po tym, gdy nagrodzone przyrządy udowodniły swoją wartość dla nauki. Michelson otrzymał Nobla już po wykonanym w 1887 roku doświadczeniu Michelsona-Morleya dowodzącym, że prędkość światła w układzie źródła nie zależy od ruchu Ziemi. Siegbahn dostał Nobla za spektroskopię rentgenowską, ale już po tym jak dzięki niej poprawił układ pierwiastków i zrozumiał lepiej powłokę elektronową. Przy Noblu za projekty mikroskopu elektronowego i skaningowego mikroskopu tunelowego wiadomo było, gdzie takich odkryć szukać.

Sieci neuronowe to nie koniec fizyki. Choć narzędzie to jest bardzo obiecujące, to jednak daleko jest jeszcze do faktycznego rozwiązywania problemów naukowych za jego pomocą, mimo twierdzeń niektórych twórców (w tym Sama Altmana – dyrektora generalnego OpenAI,

który w wywiadzie z NBC News na Aspen Ideas Festival zadeklarował, że w przyszłości będziemy mogli wydawać polecenie komputerowi *hey, computer, discover all of physics* i on to zrobi). Kiedy ktoś wypowiada się ekstatycznie o nowym sprzedawanym narzędziu, warto sprawdzić, czy słowa tej osoby mogą mieć drugie dno, a w szczególności czy jej zarobki mogą zależeć od opinii społeczeństwa o tworzonych przez nią narzędziach. Póki co, w najlepszym razie, sieci neuronowe dają nam odpowiedzi na różne pytania, nie informując jednak o ścieżce rozumowania i w związku z tym, mają ograniczone możliwości odkrywania przed nami mechanizmów rządzących rzeczywistością. To jest coś, co najbardziej zaskakuje mnie w werdykcie komisji noblowskiej: ta nagroda wydaje się przedwczesna. Choć sieci neuronowe pomagają nam w pracy naukowej, to brakuje wciąż przykładu, w którym dzięki sieciom dowiedzieliśmy się czegoś przełomowego i nowego o świecie. Dobrym przykładem tego ograniczenia jest AlphaFold [11], czyli największy dotychczasowy sukces uczenia maszynowego w nauce. AlphaFold pobiło wszelkie inne numeryczne podejścia do przewidywania struktury trójwymiarowej białek, znając tylko kolejność aminokwasów w białku. Jest fantastycznym, choć nieidealnym narzędziem do predykcji, które przyspieszyło pracę biologów podając dobre punkty wyjściowe do dalszych badań, ale **nie** rozwiązało problemu składania białek, a w szczególności niewiele powiedziało nam o mechanizmach rządzących składaniem białek [12]. Problemem AlphaFold, jak i innych sieci neuronowych, jest brak ich *interpretowalności*. Pracujemy nad tym i np. w celu detekcji przejść fazowych budujemy specjalne sieci, które zmuszone są mówić naszym językiem i dzięki temu rozumiemy, co dokładnie robią [13]. Wymaga to jednak jakiegoś stopnia zrozumienia atakowanego problemu, więc wielkim wyzwaniem jest projektowanie interpretowalnych sieci dla problemów, o których niewiele wiemy.

ChatGPT i jego krewniacy też nie zwiastują końca fizyki. Najbardziej imponujące współcześnie modele uczenia maszynowego, które spowodowały ostatni zryw zainteresowania tymi metodami to niewątpliwie wielkie modele językowe, takie jak ChatGPT, Claude, czy Gemini. Są one modelami probablistycznymi, tak trenowanymi by przewidywały jak najbardziej prawdopodobne kontynuacje zdań. Choć wykazują imponujące zachowania emergentne (np. duże modele potrafią podążać za instrukcjami, choć nie były konkretnie tego uczone), to ostatecznie ich interakcja ze światem zewnętrznym odbywa się, w dużej mierze, za pośrednictwem tekstu. Wyobraźcie sobie, że nigdy niczego nie dotknęliście, nie zobaczyliście, nigdy niczym nie rzuciliście. Możecie tylko czytać o tym co inni robią, a waszym jedynym celem jest zgadywać, jakie słowa pojawią się za chwilę. Co najwyżej oglądacie czasem zdjęcia lub wideo, ale niewiele pomaga to w zrozumieniu przyczynowości. Bardzo ciężko na tej podstawie nauczyć się zasad działania rzeczywistości czy ciekawości i zadawania niezadanych jeszcze pytań. Łatwo za to „halucynować" [14] i dochodzić przez to do błędnych (w wersji pesymistycznej) lub zaskakujących (w wersji optymistycznej) wniosków na temat świata. Te ograniczenia objawiają się szczególnie w zadaniach wymagających rozumowania i planowania, czego wielkie modele językowe wciąż nie opanowały [15]. Są to ograniczenia, z których koniecznie trzeba zdawać sobie sprawę, gdy chce się korzystać z tych narzędzi. Zawsze sprawdzajcie fakty podawane przez modele językowe, zanim się na nie powołacie – wystarczy szybkie wyszukiwanie internetowe.

Ekscytująca przyszłość. Wielkie modele językowe będą świetnymi asystentami osobistymi i będą tylko lepiej i lepiej wykonywać zadania związane z tekstem. Naukowcy dalej będą stawiać krytyczne hipotezy i testować je wykonując eksperymenty oraz analizując wyniki. Będziemy mieć za to teraz inspirującą pomoc w postaci sztucznych sieci neuronowych, które usprawnią różne elementy tego procesu: poprawią błędy językowe, streszczą artykuły naukowe, zautomatyzują powtarzalne elementy pracy doświadczalnej [16], czy nawet będą partnerem do dyskusji na temat dowodów matematycznych czy nowych kierunków badań [17]. Gdy uczynimy je interpretowalnymi, może pokażą nam też wzorce, które do tej pory przegapialiśmy w jakichś danych naukowych? Może wytłumaczą rozwiązanie problemu, które do tej pory nam się wymykało? Takie odkrycie byłoby przełomowe i z pewnością warte Nobla!

Literatura

[1] A. Dawid and Y. LeCun, Introduction to latent variable energy-based models: a path toward autonomous machine intelligence, J. Stat. Mech. 2024, 104011 (2024).

[2] W. Little, The existence of persistent states in the brain, Mathematical Biosciences 19, 101 (1974).

[3] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. U.S.A. 79, 2554 (1982).

[4] G. E. Hinton and T. J. Sejnowski, Optimal perceptual inference, in Proc. IEEE Comput. Vis. Pattern Recognit. (1983).

[5] A. Dawid et al., Modern applications of machine learning in quantum sciences (2023), arXiv:2204.04198 [quant-ph].

[6] J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W. M. C. Foulkes, D. Pfau, G. Carleo, and F. Noé, Ab initio quantum chemistry with neural-network wavefunctions, Nat. Rev. Chem. 7, 692–709 (2023).

[7] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. 91, 045002 (2019).

[8] J. Pavlus, The computer scientist who builds big pictures from small details (2024), Quanta Magazine.

[9] M. Schuld and F. Petruccione, Machine Learning with Quantum Computers (Springer International Publishing, 2021).

[10] S. Thais, Physics and the empirical gap of trustworthy AI, Nat. Rev. Phys. 6, 640–641 (2024).

[11] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596, 583 (2021).

[12] Y. Saplakoglu, How AI revolutionized protein science, but didn't end it (2024), Quanta Magazine.

[13] K. Cybiński, J. Enouen, A. Georges, and A. Dawid, Speak so a physicist can understand you! TetrisCNN for detecting phase transitions and order parameters (2024), arXiv:2411.02237 [quant-ph].

[14] S. Farquhar, J. Kossen, L. Kuhn, and Y. Gal, Detecting hallucinations in large language models using semantic entropy, Nature 630, 625–630 (2024).

[15] R. Patil, Can LLMs reason and plan? Exploring Blockworld, Mystery Blockworld (2024), Medium.

[16] J. P. Zwolak, J. M. Taylor, R. W. Andrews, J. Benson, G. W. Bryant, D. Buterakos, A. Chatterjee, S. Das Sarma, M. A. Eriksson, E. Greplová, M. J. Gullans, F. Hader, T. J. Kovach, P. S. Mundada, M. Ramsey, T. Rasmussen, B. Severin, A. Sigillito, B. Undseth, and B. Weber, Data needs and challenges for quantum dot devices automation, npj Quantum Inf. 10, 105 (2024).

[17] M. Krenn and A. Zeilinger, Predicting research trends with semantic and neural networks with an application in quantum physics, PNAS 117, 1910–1916 (2020).