doi:10.2533/chimia.2023.56 Chimia 77 (2023) 56–61

The Societal and Scientific Importance of Inclusivity, Diversity, and Equity in Machine Learning for Chemistry

Daniel Probst*

Abstract: While the introduction of practical deep learning has driven progress across scientific fields, recent research highlighted that the requirement of deep learning for ever-increasing computational resources and data has potential negative impacts on the scientific community and society as a whole. An ever-growing need for more computational resources may exacerbate the concentration of funding, the exclusiveness of research, and thus the inequality between countries, sectors, and institutions. Here, I introduce recent concerns and considerations of the machine learning research community that could affect chemistry and present potential solutions, including more detailed assessments of model performance, increased adherence to open science and open data practices, an increase in multinational and multi-institutional collaboration, and a focus on thematic and cultural diversity.

Keywords: Drug discovery · Machine learning · Organic chemistry

Daniel Probst is a computer scientist, bioinformatician, and cheminformatician born in Switzerland. He obtained his PhD in Chemistry and Molecular Sciences at the University of Bern under the supervision of Prof. Jean-Louis Reymond in 2020. After a two-year stay at IBM Research Europe as a research staff member in the group of Dr Teodoro Laino, he joined the group of Prof. Pierre Vandergheynst at EPFL. His current research focuses on the intersection of biology, chemistry, and computer science.

1. Introduction

Advances in computer hardware, specifically the introduction of graphics cards that support programmable shaders in the early 2000s, made deep neural networks computationally tractable and sparked a revolution in machine learning. This also affected natural sciences, including chemistry, where the fraction of machine learning-related articles in the literature has more than doubled in the past six years (Fig. 1). The past decade has seen machine learning-driven breakthroughs in multiple fields, including biology, natural language processing, and computer vision and image processing.[1–3] In all cases, a prerequisite for achieving these breakthroughs was the availability of large data sets on which to train the deep neural networks – a condition also met in chemistry, where large molecule and reaction databases exist.[4–7] However, these breakthroughs came at high computational costs, as complex deep learning models take up to hundreds of graphical processing unit (GPU) years to train (Table 1). For example, the electricity cost for a single training run, which does not include data processing or experimentation, of the language model MT-NLG has been estimated to be approximately USD 1 million.[8] This high computational cost has also raised concerns regarding the carbon footprint of such models, as training runs can consume energy in the GWh range. However, cost-benefit analyses of large models generally remain estimations due to missing details of the development process and the possibility of reusing or fine-tuning the models for various tasks, where they can reduce further training or computation time, thereby offsetting training costs and potential carbon emissions. The enormous computational costs during training compared to previous machine-learning methods are driven by an intrinsic property of deep-learning models: they scale relatively well compared to other approaches. However, this comes at a price, as the performance of deep neural networks scales linearly at an exponential cost in both data and computation,[9,10] resulting in diminishing returns.[11] This also applies to new neural network architectures, such as the transformer, which has found broad application in natural language processing and chemistry. While this new architecture expanded the abilities of deep neural networks, the likely intrinsic limitations remain, as large language models (LLMs), such as MT-NLG, have shown.

MT-NLG, trained on thousands of GPUs for three months by Microsoft and Nvidia, is a transformer English language model with 530 billion parameters.[13] Compared to a previous large language model, GPT-3, MT-NLG consumed four times as much energy (Table 1). However, the increase in accuracies across several language benchmarks was somewhat limited: Below 1% in LAMBADA, a word prediction benchmark, 1–3% in the Winogrande, HellaSWAG, and PiQA reasoning benchmarks, 2.5% in the RACE-h reading comprehension benchmark, and 3% in the WiC word sense disambiguation benchmark. It did fare better in the BoolQ reading comprehension benchmark (7–18% increase) and specific subtasks of other benchmarks. However, with the model being three times as big compared to GPT-3, it is an example of diminishing returns when pushing the limits of current deep-learning approaches.

Similar tendencies, although on a smaller scale, have started appearing in chemistry machine-learning research. Earlier this year (2022), Nvidia released the MegaMolBART model for *de novo* molecular generation and reaction prediction (Table 1). While training using 32 GPUs for one day seems very reasonable compared to the resources used by large language models, the training may require months in many academic settings due to limited resources. At the time of writing, the GPU cluster at EPFL, a comparatively well-funded public research institution, shared

*Correspondence: Dr. D. Probst, E-mail: daniel.probst@epfl.ch
School of Engineering, EPFL, Lausanne, CH-1015 Lausanne

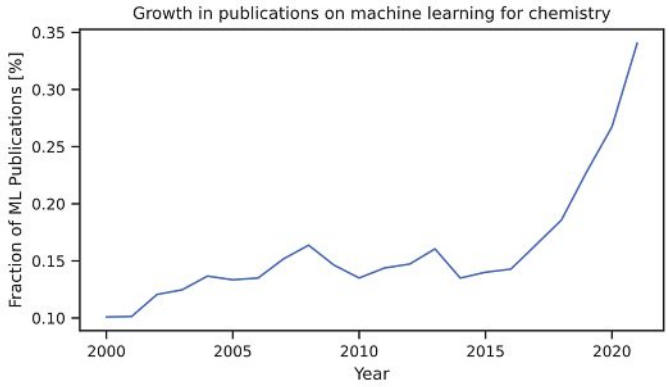

Fig. 1. Over the past six years, the fraction of machine learning publications in all chemistry publications has more than doubled. Adapted from previous publication.[8]

on the direction of an entire research field. One major concern is a narrowing of machine-learning research, as scarce resources are increasingly allocated to a specific research direction, in this case, deep learning, before other venues have been sufficiently explored. Furthermore, research into the thematic diversity of machine learning has shown that the private sector may have a narrower focus than academia and that, within academia, elite universities that often collaborate with private organisations have a narrower focus than other institutions.[23]

The subsequent sections will, based on the key findings in the exploratory study '*Social and Environmental Impact of Recent Developments in Machine Learning on Biology and Chemistry Research*',[8] make suggestions for how to tackle the described challenges facing machine learning in chemistry.

Table 1. Selected deep learning models from different domains. Recent developments in machine learning have driven the implementation of ever-larger neural networks. This trend is mainly driven by large technology companies with access to vast computational resources. The examples in this table show the hardware and time used to train the models, their use, as well as the estimated energy consumed by training them. It is important to note that the energy consumption represents only the final training and not previous experimentation where commonly different variants of the model are trained and evaluated.

	Use	Hardware	Time	Energy
GPT-3	Large language model (LLM)	310 V100s (estimated)[12]	90 days[12]	1.887 GWh[12]
MT-NLG	Large language model (LLM)	NVIDIA Selene supercomputer with 560 DGX A100 nodes (4,480 GPUs)[13]	90 days[14]	7.862 GWh[8]
AlphaFold2	Protein structure prediction	128 Google TPUv3[15]	11 days[15]	15 MWh[16]
FastFold	Protein structure prediction	512 A100 GPUs[15]	2.8 days[15]	28 MWh[17]
RoseTTAFold	Protein structure prediction	8 V100 GPUs (Assumed to be 1 DGX-1 node)	28 days[18]	0.672 MWh
MegaMolBART	*De novo* molecular generation, reaction prediction	4 DGX-1 nodes (32 V100 GPUs)	1 day[19]	0.336 MWh[20]

by hundreds of researchers provides access to 136 of the same GPU models used to train MegaMolBART. Such unequal access to computational resources has recently been raised as a concern within the machine-learning community.[21] An analysis of papers presented at prestigious computer science conferences shows that large technology companies, through their own publications and collaborations with elite universities, may have started to crowd out mid-tier and lower-tier universities from these conferences.[21] This unequal access to resources may extend to labour and expertise, as a recent brain-drain from academia to industry, specifically the 'big tech' industry, has been documented.[22]

In a recent study,[8] I analysed whether such potential socioeconomic and scientific impacts are also present in biology and chemistry machine-learning research, as deep learning has found broad application within both fields. By processing 16,301 chemistry works filtered for machine-learning topics, I was able to reproduce multiple findings that raised concern within the machine-learning community, including a potential trend towards increasing global inequality in machine-learning research for chemistry, a potential brain-drain from academia to industry, and significant citation inequality between publications with and without industrial affiliations. In addition, I found an increase in the use of computational resources per publication during the past three years that outpaced that of previous years. However, the statistical power of this last important metric was lacking since very few (0.4%) of the articles included both the hardware used and the time spent to train the models.

As has been increasingly discussed within the machine-learning community, such developments may have a substantial impact

2. Evaluation of Models beyond Accuracy

Given the nature of deep-learning algorithms, creating increasingly large models or training networks on ever bigger data sets will ultimately lead to ever smaller improvements over previous versions.[11] This means that accuracy as a metric is primarily of true interest when evaluating novel neural network architectures or newly introduced data sets, but after that tends to merely become a logarithmic function of the model cost, leading to a linear increase in accuracy at an exponential cost. In their seminal paper Green AI, Schwartz *et al.* defined the estimated cost of achieving a given result R as linearly proportional to the product of the model size (number of parameters) E, the size of the training data D, and the number of hyperparameter experiments (*e.g.* searching for the optimal number of hidden layers or neurons per layer) H (Eqn. 1).[24]

$$Cost(R) \propto E \cdot D \cdot H \quad (1)$$

In practice, this means that a model's reported accuracy is based not only on the conceptual approach but also on an exponentially growing, ultimately financial, investment. The vast differences in funding between sectors, institutions, and individual research groups thus make it necessary to evaluate reported results not only on the reported accuracy of a model but also on the associated cost.[24] Failing to do so and to assess only a model's accuracy may lead to the unfortunate circumstance in which a superior

approach is discarded because it was seemingly outperformed by a conceptually inferior approach where sufficient resources were available to increase the accuracy. This situation can significantly hamper scientific progress and, in a worst-case scenario, could drive research in a suboptimal direction. In that sense, reporting the accuracy of a model without the associated costs can be equated to reporting an observed value of a statistic without a value for significance, such as the *p*-value. Therefore, researchers and publishers should implement measures to ensure that all metrics necessary to evaluate a model independently from the invested computational resources are reported. This includes, in addition to the final model's accuracy, the training time and the hardware used for training, as exemplified in Table 1. Furthermore, the same metrics (accuracy, training time, and hardware used) should be reported for the hyperparameter experiments, which will enable the evaluation of to what degree a model has been optimised and whether there remains room for improvement. Finally, the time and hardware requirements at inference (when using the model to make predictions) should be included, as some models, such as AlphaFold, require substantial resources at this stage.

The complete suggested metrics and additional information to be reported are summarised in Table 2. Reporting these metrics, in addition to existing good research practices such as open data and open source, will not only contribute towards levelling the playing field in applied machine-learning research for chemistry but lower the risk of dismissing potentially better models due to a lack of resources invested during training and optimisation for the initial publication. Reporting these measures can also implicitly cause a focus and eventual increase in efficiency.[24] This has the potential to lower the significant carbon footprint of large models,[25] make models and neural network architectures more accessible, increase the feasibility of fully optimised models, and ease the choice of models for a given task that may include small training sets. Finally, reporting comprehensive metrics and publishing the code and data will increase the reproducibility of the research and enable better peer review, as the retraining of a model for review purposes is prohibitively costly.

A note on efficiency in machine learning: From a purely scientific perspective, it is also important to remember, especially with the recent tendency to refer to machine-learning models as artificial intelligence, that human intelligence generally requires to train for about 16 years to be able to easily solve the tasks of large language models among many others while consuming a total of 2.8 MWh[541] – a small fraction of the energy used to train a large machine learning model. A comparative lack of efficiency and the diminishing returns of deep neural networks are important to consider when realistically evaluating the potential of machine-learning models, even in tasks where they potentially considerably outperform humans.

3. Inclusive Machine-learning Research

It has been shown that big technology companies such as Google and Facebook, often in collaboration with elite universities, may crowd out mid-tier and low-tier universities out of respectable machine learning conferences.[21] In addition, following a general trend in science, I previously identified considerable quotation inequalities between 'big tech' and academic institutions and a possible emerging increase in global publication inequality based on research funding in machine learning for chemistry.[8,26,27] Potential causes for these developments are inequities in access to computational resources and private research increasingly attracting talent from machine-learning fields in academia.[22,24] Furthermore, unequal funding between academic institutions and countries is well documented.[8,26] These conditions indicate a potential increase in the exclusivity of machine-learning research, including applied machine-learning research in chemistry.

Table 2. When publishing a new model, the training times and hardware used for both the hyperparameter experiments as well as the final training should be reported in addition to the accuracy or a similar performance metric. The hardware and associated time at inference should also be reported. Finally, the data should also be made accessible to enable reproducibility.

	Metrics/Data/Code
Data	• Data download instructions • Training, validation, and testing data • Data splits and splitting method • Summary statistics of data set and splits
Code	• Data preprocessing scripts • Training and evaluation scripts
Hyperparameter experiments	• Hyperparameter values ◦ Accuracy ◦ Training time ◦ Hardware used
Training	• Accuracy • Training time • Hardware used
Inference	• Inference time • Hardware used

As the recent training of the open BLOOM language model has shown, extensive public and private funding, including the donation of computational resources by smaller technology companies, and an international research effort are required to train a model commonly trained by big technology companies.[28] Similar challenges are also faced globally by less-funded institutions or groups who train models comparable in complexity to those published by their well-funded counterparts, generally found in Europe, China, and Northern America. Indeed, as there is potential for machine learning to be a driver for increasing global socioeconomic inequality, similar developments should be discussed among the international scientific community.[29,30] In other fields, this concentration of machine-learning research at elite institutions in the Global North has already led to negative societal effects. These negative effects prominently include racially biased machine-learning models used in healthcare and public health.[31–33] Such problematic biases are commonly attributed to data sets; however, they also include other statistical and computational biases as well as human and systemic biases.[34–36] Similar biases could become increasingly problematic in medicinal chemistry and drug development, given the vast diversity of ADME genes in Sub-Saharan Africa as an example.[37] However, promoting awareness of biases and diversifying clinical and genomic data may not be enough to solve the challenge of inequality in machine learning. Keeping with the example of Sub-Saharan Africa and medicine, there has been a deluge of interesting and important deep-learning approaches for analysing mammography data in the past five years.[38,39] However, in 2019, a study showed that, for socioeconomic reasons, mammographies have little positive effect in Sub-Saharan Africa.[40] This example shows that machine-learning models developed in high-income countries may have no applications in low- or middle-income countries, suggesting that merely allowing access to, *e.g.* pretrained models may not solve problems for different local contexts.[41] Instead, local researchers must be provided with the tools to develop local solutions to local issues.[42] This echoes the still unsolved problem of drug development for neglected tropical diseases, where local bottom-up approaches hold great promise.[43]

Given the span of available financial resources, infrastructure, and know-how, practising inclusive machine-learning research should be of interest to most institutions and individual groups, as

there is always a bigger fish. While discussing potential solutions to funding concentration in research or global inequality is far beyond the scope of this article, there are mitigating measures. The following subsections introduce simple measures that have the potential to counter the increasing exclusivity in machine learning for chemistry.

3.1 *Publishing Detailed Training and Evaluation Metrics beyond Final Values*

A deep neural network is trained iteratively by passing batches of data through the network and updating the model's parameters to find a local, or optimally global, minimum of a loss function. In the example of a simple linear regression, the loss function calculates the mean squared error between the actual value and the value predicted by the model. In a deep neural network, the gradient of the loss function is calculated, and the model parameters are updated to descend along this gradient towards a lower value. A lower loss then leads to higher accuracy of the model. Plotting the loss function values during model training generally results in a curve reminiscent of a power law function, which means there is an initial steep drop in loss (increase in accuracy) followed by a long tail (Fig. 2).[44]

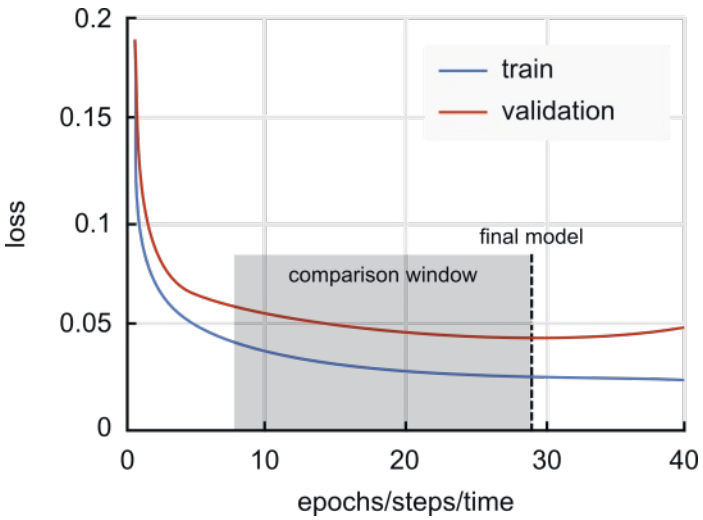

Fig. 2. An often observed loss curve. Making such complete metrics available would allow for continuous comparison with a new model once state-of-the-art performance is reached or the researcher chooses to discard the training run before the final model has been trained, helping to reduce computational resource usage and cost.

Publishing these entire training and evaluation metrics instead of just the final values would allow researchers with limited computational resources to evaluate their models against the state-of-the-art without necessarily running the training for weeks or months. Based on this early evaluation, the model can then be published as-is, with potentially lower accuracy but higher efficacy, or scarce computational resources can be invested more selectively.

3.2 *Training Models on Smaller Data Sets*

A common way to evaluate a deep neural network architecture is through data ablation studies. One flavour of such a study is a step-wise reduction in training set size to explore the influence on the model performance. As the training set size significantly influences the computational cost of training a model, publishing the results of such a data ablation study allows researchers with limited access to computational resources to train and compare their models on a smaller training set. Training on a smaller set would require fewer computational resources.

As the example in Table 3 shows, 10% of the training data can be sufficient to draw a conclusion about the final performance of a model. In addition, such a study can be informative when choosing an existing deep neural network architecture given limited training data.

Table 3. R^2 metrics from a data ablation study adapted from '*Reaction classification and yield prediction using the differential reaction fingerprint DRFP*'.[45]

% of Training Set	RF	BERT	XGBOOST
70%	0.92	0.95	0.95
50%	0.9	0.92	0.93
30%	0.85	0.88	0.89
20%	0.81	0.86	0.87
10%	0.77	0.79	0.81
5%	0.68	0.61	0.73
2.5%	0.59	0.45	0.62

3.3 *Collaborations and Sharing of Resources*

Unlike laboratory equipment and the associated experiments, machine-learning experiments don't require physical access to equipment, can be distributed globally, and can be paused and resumed at any given time. However, geographic and economic distance remain significant negative influences on scientific collaboration in the machine-learning domain.[46] This shows ample room for increasing equitable collaborations between computational chemists, cheminformaticians and chemists across high- and low- and middle-income countries.[47] Beyond direct collaboration, the sharing or donating of computational resources has proven fruitful in computational biology.[48] On a larger scale, volunteer projects such as Folding@Home, which has broken the exaflop barrier during the early stages of the COVID-19 pandemic,[49] have shown themselves to be highly successful and may be adapted to provide computational resources to disadvantaged researchers in machine learning for chemistry.[50]

4. Thematic Diversity

With the advent of a promising new method, in this case, deep neural networks, it is common that resources are shifted towards exploring and further developing this method. This is currently no different in the machine-learning research community. However, as this fast adaption is driven by big technology corporations and weaknesses such as a high environmental cost, the potential propagation of biases, and a lack of robustness of the models have largely been ignored, concerns about a potential premature narrowing of machine-learning research have been raised.[23] Indeed, recent research has found that thematic diversity in machine learning may have stagnated and that industry research has a potentially much narrower focus than academic research.[51]

However, in chemistry, the analysis in '*Social and environmental impact of recent developments in machine learning on biology and chemistry research*' has shown that during the recent rise of deep learning methods, other methods have not been abandoned. While most of the increase in scientific output can be attributed to deep neural networks and random forest/boosting, the use of all observed categories of machine-learning methods has been growing between 2020 and 2021 (Fig. 3).[8,52] These results suggest that machine learning in chemistry is still methodologically diverse, and contribution to the field is not limited by access to extensive computational resources required by deep learning. However, as deep learning has outpaced other methods in terms of citations,

and papers with industry affiliations gather significantly more citations than academic-only publications, the community should remain vigilant to avoid a narrowing of research in the future.[8,23]

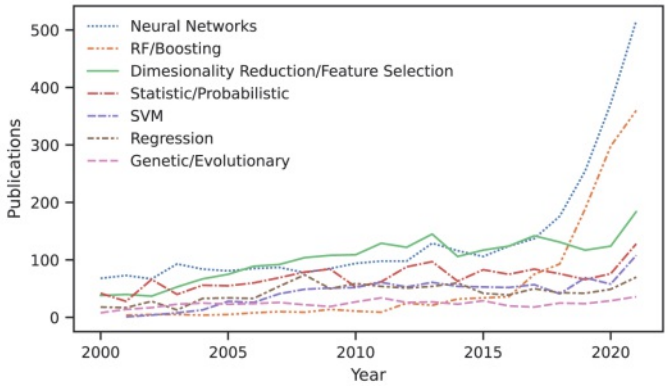

Fig. 3. Number of publications in machine learning for chemistry by use of method. Neural Networks, together with Random Forest and Boosting, have been the main contributors to the recent increase in publications. Adapted from previous publication, based on OpenAlex data.[8,52]

5. Conclusions

In this article, I have introduced current concerns and considerations from the machine-learning community on how technological changes over the past decade have influenced the international scientific effort. In chemistry, where research on applied machine learning has seen significant growth and increasingly diverse applications during the past six years, these concerns and considerations still need to be discussed. Based on a previous bibliometric study on machine-learning publications, I identified the potential for increased exclusivity in machine learning for chemistry driven by an ever-growing need for computational resources. As this development may have a considerable negative impact on the future of the field, I selected three areas that have the potential to counter these trends with relatively small efforts by the research community and by publishers.

Evaluating machine-learning models beyond their accuracy will potentially contribute towards levelling the playing field between researchers with access to vast computational resources and those without. This will also avoid circumstances where a conceptually superior method is passed over due to a lack of resources for training. Furthermore, this will advance research on data-efficient models, which show good performance in settings with small training sets, of which many examples exist in chemistry. Finally, more efficient models will lower the machine-learning field's carbon footprint, which is currently being addressed by the broader machine-learning community. This effort is being complicated by a lack of standardised evaluation techniques.[53] However, as standardised evaluation techniques are being developed, the need for an evaluation beyond accuracy will have to be addressed.

Making machine learning in chemistry more inclusive is required in a field with multiple widening gaps in terms of access to computational resources. Academia in high-income countries is increasingly compelled to collaborate with the private sector to gain access to computational resources that enable them to compete with big technology companies, while elite universities increasingly collaborate with these big corporations. Meanwhile, researchers in low- and middle-income countries are at risk of being excluded entirely. Publishing detailed metrics of model training and the performance of models on smaller data sets could be an effective way of enabling research on these models, and potentially superior alternatives, in settings where computational resources are scarce.

Sustaining thematic diversity in machine learning for chemistry is essential, as it counters a potentially premature narrowing of research that may lead the field, or parts of the field, down the wrong path. In addition, the increasing necessity to invest in deep learning-specialised hardware to compete with state-of-the-art research poses the danger of lock-in or vendor lock-in due to the current quasi-monopoly by Nvidia. While machine-learning research in chemistry is currently thematically diverse, recent significant citation inequalities between machine learning methods and sectors should be monitored in the future.

Acknowledgements

I would like to thank Prof. Pierre Vandergheynst for hosting me as a postdoctoral researcher in his group at EPFL.

Received: November 15, 2022

[1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, *Arxiv* **2020**, https://doi.org/10.48550/arxiv.2005.14165.

[2] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583, https://doi.org/10.1038/s41586-021-03819-2.

[3] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, I. Sutskever, *Arxiv* **2021**, https://doi.org/10.48550/arxiv.2102.12092.

[4] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res*. **2012**, *40*, D1100, https://doi.org/10.1093/nar/gkr777.

[5] T. Sterling, J. J. Irwin, *J. Chem. Inf. Model*. **2015**, *55*, 2324, https://doi.org/10.1021/acs.jcim.5b00559.

[6] Chemical reactions from US patents (1976-Sep **2016**), https://doi.org/10.6084/m9.figshare.5104873.v1.

[7] P. Bansal, A. Morgat, K. B. Axelsen, V. Muthukrishnan, E. Coudert, L. Aimo, N. Hyka-Nouspikel, E. Gasteiger, A. Kerhornou, T. B. Neto, M. Pozzato, M.-C. Blatter, A. Ignatchenko, N. Redaschi, A. Bridge, *Nucleic Acids Res*. **2021**, *50*, D693, https://doi.org/10.1093/nar/gkab1016.

[8] D. Probst, *Arxiv* **2022**, 2210.00356, https://doi.org/10.48550/arxiv.2210.00356

[9] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, K. Murphy, *2017 Ieee Conf. Comput. Vis. Pattern Recognit. Cvpr* **2017**, 3296, https://doi.org/10.1109/cvpr.2017.351.

[10] C. Sun, A. Shrivastava, S. Singh, A. Gupta, *2017 Ieee Int. Conf. Comput. Vis. Iccv* **2017**, 843, https://doi.org/10.1109/iccv.2017.97.

[11] N. C. Thompson, K. Greenewald, K. Lee, G. F. Manso, *Ieee Spectrum* **2021**, *58*, 50, https://doi.org/10.1109/mspec.2021.9563954.

[12] L. F. W. Anthony, B. Kanding, R. Selvan, *Arxiv* **2020**, https://doi.org/10.48550/arxiv.2007.03051.

[13] S. Smith, M. Patwary, B. Norick, P. LeGresley, S. Rajbhandari, J. Casper, Z. Liu, S. Prabhumoye, G. Zerveas, V. Korthikanti, E. Zhang, R. Child, R. Y. Aminabadi, J. Bernauer, X. Song, M. Shoeybi, Y. He, M. Houston, S. Tiwary, B. Catanzaro, *Arxiv* **2022**, https://doi.org/10.48550/arxiv.2201.11990.

[14] https://www.deepspeed.ai/2021/12/09/deepspeed-moe-nlg.html

[15] S. Cheng, R. Wu, Z. Yu, B. Li, X. Zhang, J. Peng, Y. You, *Arxiv* **2022**, https://doi.org/10.48550/arxiv.2203.00854.

[16] https://www.nextplatform.com/2018/05/10/tearing-apart-googles-tpu-3-0-ai-coprocessor/

[17] https://images.nvidia.com/aem-dam/Solutions/Data-Center/nvidia-dgx-a100-80gb-datasheet.pdf

[18] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, *Science* **2021**, *373*, 871, https://doi.org/10.1126/science.abj8754.

[19] https://catalog.ngc.nvidia.com/orgs/nvidia/teams/clara/models/megamolbart

[20] https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/dgx-1/dgx-1-rhel-datasheet-nvidia-us-808336-r3-web.pdf

[21] N. Ahmed, M. Wahed, *Arxiv* **2020**, https://doi.org/10.48550/arxiv.2010.15581.
[22] R. Jurowetzki, D. Hain, J. Mateos-Garcia, K. Stathoulopoulos, *Arxiv* **2020**, 2009.10385, https://doi.org/10.48550/arxiv.2009.10385.
[23] J. Klinger, J. C. Mateos-Garcia, K. Stathoulopoulos, *Ssrn Electron J*. **2020**, https://doi.org/10.2139/ssrn.3698698.
[24] R. Schwartz, J. Dodge, N. A. Smith, O. Etzioni, *Commun. Acm*. **2020**, *63,* 54, https://doi.org/10.1145/3381831.
[25] C.-J. Wu, R. Raghavendra, U. Gupta, B. Acun, N. Ardalani, K. Maeng, G. Chang, F. A. Behram, J. Huang, C. Bai, M. Gschwind, A. Gupta, M. Ott, A. Melnikov, S. Candido, D. Brooks, G. Chauhan, B. Lee, H.-H. S. Lee, B. Akyildiz, M. Balandat, J. Spisak, R. Jain, M. Rabbat, K. Hazelwood, *Arxiv* **2021**, https://doi.org/10.48550/arxiv.2111.00364.
[26] M. W. Nielsen, J. P. Andersen, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2012208118, https://doi.org/10.1073/pnas.2012208118.
[27] C. J. Gomez, A. C. Herman, P. Parigi, *Nat. Hum. Behav*. **2022**, *6*, 919, https://doi.org/10.1038/s41562-022-01351-5.
[28] T. L. Scao, A. Fan, C. Akiki, E. Pavlick, S. Ilić, D. Hesslow, R. Castagné, A. S. Luccioni, F. Yvon, M. Gallé, J. Tow, A. M. Rush, S. Biderman, A. Webson, P. S. Ammanamanchi, T. Wang, B. Sagot, N. Muennighoff, A. V. del Moral, O. Ruwase, R. Bawden, S. Bekman, A. McMillan-Major, I. Beltagy, H. Nguyen, L. Saulnier, S. Tan, P. O. Suarez, V. Sanh, H. Laurençon, Y. Jernite, J. Launay, M. Mitchell, C. Raffel, A. Gokaslan, A. Simhi, A. Soroa, A. F. Aji, A. Alfassy, A. Rogers, A. K. Nitzav, C. Xu, C. Mou, C. Emezue, C. Klamm, C. Leong, D. van Strien, D. I. Adelani, D. Radev, E. G. Ponferrada, E. Levkovizh, E. Kim, E. B. Natan, F. D. Toni, G. Dupont, G. Kruszewski, G. Pistilli, H. Elsahar, H. Benyamina, H. Tran, I. Yu, I. Abdulmumin, I. Johnson, I. Gonzalez-Dios, J. de la Rosa, J. Chim, J. Dodge, J. Zhu, J. Chang, J. Frohberg, J. Tobing, J. Bhattacharjee, K. Almubarak, K. Chen, K. Lo, L. V. Werra, L. Weber, L. Phan, L. B. allal, L. Tanguy, M. Dey, M. R. Muñoz, M. Masoud, M. Grandury, M. Šaško, M. Huang, M. Coavoux, M. Singh, M. T.-J. Jiang, M. C. Vu, M. A. Jauhar, M. Ghaleb, N. Subramani, N. Kassner, N. Khamis, O. Nguyen, O. Espejel, O. de Gibert, P. Villegas, P. Henderson, P. Colombo, P. Amuok, Q. Lhoest, R. Harliman, R. Bommasani, R. L. López, R. Ribeiro, S. Osei, S. Pyysalo, S. Nagel, S. Bose, S. H. Muhammad, S. Sharma, S. Longpre, S. Nikpoor, S. Silberberg, S. Pai, S. Zink, T. T. Torrent, T. Schick, T. Thrush, V. Danchev, V. Nikoulina, V. Laippala, V. Lepercq, V. Prabhu, Z. Alyafeai, Z. Talat, A. Raja, B. Heinzerling, C. Si, E. Salesky, S. J. Mielke, W. Y. Lee, A. Sharma, A. Santilli, A. Chaffin, A. Stiegler, D. Datta, E. Szczechla, G. Chhablani, H. Wang, H. Pandey, H. Strobelt, J. A. Fries, J. Rozen, L. Gao, L. Sutawika, M. S. Bari, M. S. Al-shaibani, M. Manica, N. Nayak, R. Teehan, S. Albanie, S. Shen, S. Ben-David, S. H. Bach, T. Kim, T. Bers, T. Fevry, T. Neeraj, U. Thakker, V. Raunak, X. Tang, Z.-X. Yong, Z. Sun, S. Brody, Y. Uri, H. Tojarieh, A. Roberts, H. W. Chung, J. Tae, J. Phang, O. Press, C. Li, D. Narayanan, H. Bourfoune, J. Casper, J. Rasley, M. Ryabinin, M. Mishra, M. Zhang, M. Shoeybi, M. Peyrounette, N. Patry, N. Tazi, O. Sanseviero, P. von Platen, P. Cornette, P. F. Lavallée, R. Lacroix, S. Rajbhandari, S. Gandhi, S. Smith, S. Requena, S. Patil, T. Dettmers, A. Baruwa, A. Singh, A. Cheveleva, A.-L. Ligozat, A. Subramonian, A. Névéol, C. Lovering, D. Garrette, D. Tunuguntla, E. Reiter, E. Taktasheva, E. Voloshina, E. Bogdanov, G. I. Winata, H. Schoelkopf, J.-C. Kalo, J. Novikova, J. Z. Forde, J. Clive, J. Kasai, K. Kawamura, L. Hazan, M. Carpuat, M. Clinciu, N. Kim, N. Cheng, O. Serikov, O. Antverg, O. van der Wal, R. Zhang, R. Zhang, S. Gehrmann, S. Pais, T. Shavrina, T. Scialom, T. Yun, T. Limisiewicz, V. Rieser, V. Protasov, V. Mikhailov, Y. Pruksachatkun, Y. Belinkov, Z. Bamberger, Z. Kasner, A. Rueda, A. Pestana, A. Feizpour, A. Khan, A. Faranak, A. Santos, A. Hevia, A. Unldreaj, A. Aghagol, A. Abdollahi, A. Tammour, A. HajiHosseini, B. Behroozi, B. Ajibade, B. Saxena, C. M. Ferrandis, D. Contractor, D. Lansky, D. David, D. Kiela, D. A. Nguyen, E. Tan, E. Baylor, E. Ozoani, F. Mirza, F. Ononiwu, H. Rezanejad, H. Jones, I. Bhattacharya, I. Solaiman, I. Sedenko, I. Nejadgholi, J. Passmore, J. Seltzer, J. B. Sanz, K. Fort, L. Dutra, M. Samagaio, M. Elbadri, M. Mieskes, M. Gerchick, M. Akinlolu, M. McKenna, M. Qiu, M. Ghauri, M. Burynok, N. Abrar, N. Rajani, N. Elkott, N. Fahmy, O. Samuel, R. An, R. Kromann, R. Hao, S. Alizadeh, S. Shubber, S. Wang, S. Roy, S. Viguier, T. Le, T. Oyebade, T. Le, Y. Yang, Z. Nguyen, A. R. Kashyap, A. Palasciano, A. Callahan, A. Shukla, A. Miranda-Escalada, A. Singh, B. Beilharz, B. Wang, C. Brito, C. Zhou, C. Jain, C. Xu, C. Fourrier, D. L. Periñán, D. Molano, D. Yu, E. Manjavacas, F. Barth, F. Fuhrimann, G. Altay, G. Bayrak, G. Burns, H. U. Vrabec, I. Bello, I. Dash, J. Kang, J. Giorgi, J. Golde, J. D. Posada, K. R. Sivaraman, L. Bulchandani, L. Liu, L. Shinzato, M. H. de Bykhovetz, M. Takeuchi, M. Pàmies, M. A. Castillo, M. Nezhurina, M. Sänger, M. Samwald, M. Cullan, M. Weinberg, M. D. Wolf, M. Mihaljcic, M. Liu, M. Freidank, M. Kang, N. Seelam, N. Dahlberg, N. M. Broad, N. Muellner, P. Fung, P. Haller, R. Chandrasekhar, R. Eisenberg, R. Martin, R. Canalli, R. Su, R. Su, S. Cahyawijaya, S. Garda, S. S. Deshmukh, S. Mishra, S. Kiblawi, S. Ott, S. Sang-aroonsiri, S. Kumar, S. Schweter, S. Bharati, T. Laud, T. Gigant, T. Kainuma, W. Kusa, Y. Labrak, Y. S. Bajaj, Y. Venkatraman, Y. Xu, Y. Xu, Y. Xu, Z. Tan, Z. Xie, Z. Ye, M. Bras, Y. Belkada, T. Wolf, *Arxiv* **2022**, 2211.05100, https://doi.org/10.48550/arxiv.2211.05100.
[29] C. Alonso, A. Berg, S. Kothari, C. Papageorgiou, S. Rehman, *J. Monetary Econ*. **2022**, *127*, 18, https://doi.org/10.1016/j.jmoneco.2022.01.004.
[30] A. Birhane, *Script-ed.* **2020**, *17*, 389, https://doi.org/10.2966/scrip.170220.389.
[31] S. Kapur, *Nat. Mach. Intell*. **2021**, *3*, 460, https://doi.org/10.1038/s42256-021-00362-7.
[32] D. A. Vyas, L. G. Eisenstein, D. S. Jones, *Obstet. Gynecol. Surv*. **2021**, *76*, 5, https://doi.org/10.1097/01.ogx.0000725672.30764.f7.
[33] Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, *Science* **2019**, *366*, 447, https://doi.org/10.1126/science.aax2342.
[34] R. Schwartz, A. Vassilev, K. Greene, L. Perine, A. Burt, P. Hall, 'Towards a Standard for Identifying and Managing Bias in Artificial Intelligence', National Institute Of Standards And Technology, **2022**.
[35] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, *Acm. Comput. Surv.* **2021**, *54*, 1, https://doi.org/10.1145/3457607.
[36] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, C. Busch, *Ieee Trans. Technol. Soc*. **2020**, *1*, 89, https://doi.org/10.1109/tts.2020.2992344.
[37] J. E. B. da Rocha, H. Othman, G. Botha, L. Cottino, D. Twesigomwe, S. Ahmed, B. I. Drögemöller, F. M. Fadlelmola, P. Machanick, M. Mbiyavanga, S. Panji, G. E. B. Wright, C. Adebamowo, M. Matshaba, M. Ramsay, G. Simo, M. C. Simuunza, C. T. Tiemessen, S. Baldwin, M. Chiano, C. Cox, A. S. Gross, P. Thomas, F.-J. Gamo, S. Hazelhurst, *Front. Pharmacol*. **2021**, *12*, 634016, https://doi.org/10.3389/fphar.2021.634016.
[38] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, W. Sieh, *Sci. Rep-uk* **2019**, *9*, 12495, https://doi.org/10.1038/s41598-019-48995-4.
[39] W. Lotter, A. R. Diab, B. Haslam, J. G. Kim, G. Grisot, E. Wu, K. Wu, J. O. Onieva, Y. Boyer, J. L. Boxerman, M. Wang, M. Bandler, G. R. Vijayaraghavan, A. G. Sorensen, *Nat. Med*. **2021**, *27*, 244, https://doi.org/10.1038/s41591-020-01174-9.
[40] E. Black, R. Richmond, *Globalization Health* **2019**, *15*, 3, https://doi.org/10.1186/s12992-018-0446-6.
[41] M. Sloane, E. Moss, O. Awomolo, L. Forlano, *Equity Access Algorithms Mech. Optim*. **2022**, *1*, https://doi.org/10.1145/3551624.3555285.
[42] A. Chan, C. T. Okolo, Z. Terner, A. Wang, *Arxiv* **2021**, https://doi.org/10.48550/arxiv.2102.01265.
[43] S. Winks, J. G. Woodland, G. 'Colin' Pillai, K. Chibale, *Nat. Med*. **2022**, *28*, 1523, https://doi.org/10.1038/s41591-022-01885-1.
[44] J. Hestness, S. Narang, N. Ardalani, G. Diamos, H. Jun, H. Kianinejad, M. M. A. Patwary, Y. Yang, Y. Zhou, *Arxiv* **2017**, https://doi.org/10.48550/arxiv.1712.00409.
[45] D. Probst, P. Schwaller, J.-L. Reymond, *Digital Discov*. **2022**, *1*, 91, https://doi.org/10.1039/d1dd00006c.
[46] X. Tang, X. Li, F. Ma, *Scientometrics* **2022**, *127*, 181, https://doi.org/10.1007/s11192-021-04207-3.
[47] D. Armenteras, *Nat. Ecol. Evol.* **2021**, *5*, 1193, https://doi.org/10.1038/s41559-021-01496-y.
[48] https://scienceoutreachservers.org
[49] M. I. Zimmerman, G. Bowman, *Biophys. J*. **2021**, *120*, 299a, https://doi.org/10.1016/j.bpj.2020.11.1909.
[50] A. L. Beberg, D. L. Ensign, G. Jayachandran, S. Khaliq, V. S. Pande, *2009 Ieee Int. Symp. Parallel Distrib. Proc*. **2009**, 1, https://doi.org/10.1109/ipdps.2009.5160922.
[51] J. Klinger, J. Mateos-Garcia, K. Stathoulopoulos, *Arxiv* **2020**, https://doi.org/10.48550/arxiv.2009.10385.
[52] J. Priem, H. Piwowar, R. Orr, *Arxiv* **2022**, https://doi.org/10.48550/arxiv.2205.01833.
[53] A. Bender, N. Schneider, M. Segler, W. P. Walters, O. Engkvist, T. Rodrigues, *Nat. Rev. Chem*. **2022**, *6*, 428, https://doi.org/10.1038/s41570-022-00391-9.
[54] V. Balasubramanian, *PNAS* **2021**, *118*, e2107022118, https://doi.org/10.1073/pnas.2107022118.

License and Terms

The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2022.56