DATA20039 Seminar on Virtual Laboratories in Natural Sciences (Autumn 2022)

Jarmo Mäkelä & Kai Puolamäki 06.09.2022

Overview of course contents

This list provides a quick reference to the course contents, separated into three parts:

1. **Building digital twins** — the seminar focuses on replacing physical simulators with robust AI/ML models and exploring, e.g., what benefits, drawbacks, limits, and requirements different approaches have.
    a. Modelling compounds and reactions
    b. Modelling specific instruments
    c. Model selection
    d. Experimental tools
    e. Predictive modelling
2. **Explainable AI and uncertainty quantification** — utilising AI/ML (blindly) is not enough; we need to incorporate human understanding and interactions into these processes, i.e., to build explainable models and explanation interfaces.
    a. Active Learning and user-in loop (interactive models and visualisation)
    b. Explaining decisions of supervised learning methods
    c. Problems with using complex supervised learning models
    d. Quantifying uncertainties in the processes
3. **Constructing a virtual laboratory and HPC** — depending on ambition, building a virtual laboratory (VL) can require extensive resources and expertise, but what would be the minimal requirements, and would you need HPC?
    a. Examining the plans for planned infrastructures, e.g., European Commissions Destination Earth and SWITCH-ON Virtual Water-Science Laboratory
    b. Benefits and requirements of running VL in parallel to actual instruments/models
    c. Dataflow, interactions and (pre)processing
    d. Exploration and visualisation tool for analysing big data
    e. Environments, structures, and feedback loops (e.g., data and model corrections)

About the project work topics

In the course, the idea is that the above contents are split into project topics, which the students will cover in groups of 2 or 3. The students "teach" their topic to the other students, e.g., by making a presentation, grading other students' papers, and making a final report. The learning objective is that after the other students will

1. understand how the concepts and techniques are used in virtual laboratories and natural sciences in general,
2. understand the unique requirements and challenges for these techniques posed by virtual laboratories (or applications in science),
3. understand what is currently known and what are the main open problems, and
4. can continue further studies on the topic and contribute to the domain themselves.

Each project topic follows the same template: projects in topic groups 1–3 focus on a specific machine learning or other technique, while projects in group 4 focus on a specific application area (e.g., quantum physics). You should familiarise yourself with the topic, study the relevant literature, and present the most central issues and relevant items to your peers. Each topic comes with an initial list of references. The list of references is not meant to be final but a starting point: some of the initial references might be more central than others, others not that important, and you can find more references.

Depending on student interest and suggestions, we may still merge some topics, split others, ignore some, and create new ones.

You should consider the topic description as a guideline, not the final truth. Your first task is to make a project plan, where you can specify your topics and indicate the most central references.

List of project work topics

1.1 & 1.2 & 1.3 Regression methods as emulators

Extensive computer models, for example, Earth system or climate models that simulate complex physical relations or quantum mechanical systems, are slow to run. Their correct usage requires in-depth knowledge of the underlying processes. These models can be replaced by computationally faster emulators, which can be used in tangent with the simulators, e.g., to provide more immediate feedback for the user and reduce underlying uncertainties. Naturally, this approach raises such questions as what modelling options we can choose from; is the applied method appropriate and how to confirm this; and how to best combine faster emulators and slower simulators; how to select a good set of features or construct a suitable distance measure to build the emulators (often, data is very high-dimensional)?

This project examines how to replace heavy, accurate simulations with faster machine learning (ML) counterparts. The work should explain the possibilities and caveats of different approaches

— the idea is to present concrete examples of regression methods and how they have been applied in the sciences. This topic has three potential subtopics (that may be combined): **(1.1)** using emulators in tangent with simulators, and **(1.2)** using emulators as replacements for simulators. Typically, this is done, e.g., by training a generative adversarial deep learning network (GAN) to emulate full simulator output.

(1.3) One interesting approach is the multi-fidelity approach, where the idea is that there are, on the one hand, high-fidelity (accurate) but costly-to-run models (e.g., full physics simulation) and, on the other hand, low-fidelity (less accurate) but fast-to-run ML emulators or simple, effective models. These models can be thought to form a "ladder", with the higher fidelity models at the top and lower fidelity at the bottom. Often, we get faster and more accurate predictions by using these models more smartly together than separately. For example, the highest fidelity model could be a full physics simulator that is accurate but slow to run, and the lowest fidelity model could be a simple, effective model built by using the domain knowledge (typically, this model could be a power law relation obtained by the fitting model to the data or something obtained from the domain knowledge). The "middle fidelity" model could be constructed by not trying to model the simulator output directly but instead modelling the difference (residual) between the high and low fidelity models by a regression model, in which case the middle fidelity estimates would be obtained by adding the estimated residuals to the low fidelity models. Notice that this is one way to incorporate domain knowledge into the process (we want to estimate the simulator outputs, but instead of modelling the outputs directly, we model the residuals). See, e.g., Pilania et al. (2016), Ramakrishnan et al. (2015), and Zaspel et al. (2019) to get you started.

References

Anja Butter (ed), Tilman Plehn (ed), Steffen Schumann (ed), et al. Machine Learning and LHC Event Generation, Contributions to Snowmass 2021, https://doi.org/10.48550/arXiv.2203.07460

Garcia-Ruiz F, Sankaran S, Maja J M, Lee W S, Rasmussen J, and Ehsani R. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. Computers and Electronics in Agriculture. 2013, 91: 106-115. http://dx.doi.org/10.1016/j.compag.2012.12.002

Kuan Huang and Huichun Zhang, Classification and Regression Machine Learning Models for Predicting Aerobic Ready and Inherent Biodegradation of Organic Chemicals in Water, Environ. Sci. Technol. 2022, https://doi.org/10.1021/acs.est.2c01764

M. Fernández-Delgado, M.S. Sirsat, E. Cernadas, S. Alawadi, S. Barro, M. Febrero-Bande, An extensive experimental survey of regression methods, Neural Networks 111, 2019, 11-34, https://doi.org/10.1016/j.neunet.2018.12.010

Yukimasa Kaneda, Shun Shibata, Hiroshi Mineno, Multi-modal sliding window-based support vector regression for predicting plant water stress, Knowledge-Based Systems, Volume 134, 2017, 135-148, https://doi.org/10.1016/j.knosys.2017.07.028

Liu, K., Bellet, A. & Sha, F. (2015). Similarity Learning for High-Dimensional Sparse Data. Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research 38:653-662. https://proceedings.mlr.press/v38/liu15.html

Tengyuan Zhao, Yu Wang, Interpolation and stratification of multilayer soil property profile from sparse measurements using machine learning methods, Engineering Geology 265, 2020, https://doi.org/10.1016/j.enggeo.2019.105430

Verrelst, J.; Sabater, N.; Rivera, J.P.; Muñoz-Marí, J.; Vicent, J.; Camps-Valls, G.; Moreno, J. Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis. Remote Sens. 2016, 8, 673. https://doi.org/10.3390/rs8080673

Carleo et al. (2019) Machine learning and the physical sciences. Rev Mod Phys 91: 045002. https://doi.org/10.1103/RevModPhys.91.045002

Pilania et al. (2016) Multi-fidelity machine learning models for accurate bandgap predictions of solids. Comput Mater Sci 129: 156-163. https://doi.org/10.1016/j.commatsci.2016.12.004

Ramakrishnan et al. (2015) Big Data meets Quantum Chemistry Approximations: The $Delta$-Machine Learning Approach. J Chem Theory Comput 11: 2087. https://doi.org/10.1021/acs.jctc.5b00099

Zaspel et al. (2019) Boosting Quantum Machine Learning Models with a Multilevel Combination Technique: Pople Diagrams Revisited. J Chem Theory Comput 15: 1546−1559. https://doi.org/10.1021/acs.jctc.8b00832

Derkach, D., Kazeev, N., Ratnikov, F., Ustyuzhanin, A., & Volokhova, A. (2020). Cherenkov detectors fast simulation using neural networks. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 952, 161804. https://doi.org/10.1016/j.nima.2019.01.031

1.4 & 1.5 Improving observational quality from noisy measurements

Models require data — measurements or observations of the system's current state — to produce estimates of desired variables of interest. Every observation will contain errors, and significant inaccuracies can lead to poor model performance. Hence, the quality of measurements is fundamental for good-quality science. However, the scientific community is often faced with a choice between a few high-quality measurement devices or multiple cheaper but less accurate apparatuses. In some cases, even to most expensive instruments are not

"accurate enough". One approach to improving the quality of observations from low-cost devices is to develop a regression model to evaluate the "true" state from "noisy" measurements (and possibly other complementary information).

On the other hand, the measurements are usually indirect. For example, when a satellite measures gas concentrations, it does not naturally observe them directly; instead, it utilises light reflections from the surface. In other cases, the measurement process may disturb the observed process and cause bias in the measurements. For example, when we measure particle masses in a mass spectrometer, we have to charge the particles, potentially breaking them, changing the mass distribution we try to measure. One solution is to fully model the measurement process, e.g., by combining simulations and machine learning and Bayesian modelling.

This project aims to assess the feasibility and sciences approaches to utilising machine learning and multiple noisy estimates to improve the (or calibrate) measurement accuracy. The topic can be split into two parts**, (1.4)** "simple" calibration of measurements, where the idea is simply to make a regression model to estimate the "true" measurement value, and **(1.5)** full modelling of the measurement process and devices, e.g., by physics simulations and Bayesian modelling.

References

Kauppi, A., Kolmonen, P., Laine, M., and Tamminen, J.: Aerosol-type retrieval and uncertainty quantification from OMI data, Atmos. Meas. Tech., 10, 4079–4098, https://doi.org/10.5194/amt-10-4079-2017, 2017.

Concas, F., Mineraud, J., Lagerspetz, E., Varjonen, S., Liu, X., Puolamäki, K., Nurmi, P. and Tarkoma, S., 2021. Low-cost outdoor air quality monitoring and sensor calibration: A survey and critical analysis. ACM Transactions on Sensor Networks (TOSN), 17(2), pp.1-44. https://doi.org/10.1145/3446005

Georgi Tancev, Federico Grasso Toro, Variational Bayesian calibration of low-cost gas sensor systems in air quality monitoring, Measurement: Sensors, 19, 2022, https://doi.org/10.1016/j.measen.2021.100365

Zheng, T., Bergin, M. H., Johnson, K. K., Tripathi, S. N., Shirodkar, S., Landis, M. S., Sutaria, R., and Carlson, D. E.: Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments, Atmos. Meas. Tech., 11, 4823–4846, 2018, https://doi.org/10.5194/amt-11-4823-2018, 2018

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., and R. Subramanian: A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring, Atmos. Meas. Tech., 11, 291–313, 2018, https://doi.org/10.5194/amt-11-291-2018

Balsamo, G., Agusti-Panareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., Blyth, E., Bousserez, N., Boussetta, S., Brown, A., Buizza, R., Buontempo, C., Chevallier, F., Choulga, M., Cloke, H., Cronin, M. F., Dahoui, M., De Rosnay, P., Dirmeyer, P. A., … Zeng, X. (2018). Satellite and In Situ Observations for Advancing Global Earth Surface Modelling: A Review. Remote Sensing, 10(12), 2038. https://doi.org/10.3390/rs10122038

Shcherbacheva, A., Balehowsky, T., Kubečka, J., Olenius, T., Helin, T., Haario, H., Laine, M., Kurtén, T., and Vehkamäki, H.: Identification of molecular cluster evaporation rates, cluster formation enthalpies and entropies by Monte Carlo method, Atmos. Chem. Phys., 20, 15867–15906, https://doi.org/10.5194/acp-20-15867-2020, 2020.

2.1 Explainable and understandable models

Understanding how models work is essential in (almost) every field of science, hence the applicability of many machines learning (ML) and artificial intelligence (AI) methods in this context is hindered because they are seen as “black box” approaches to modelling. When the physical simulator is replaced by an ML counterpart, the user is often left in the dark on how and why the model has reached certain predictions. This lack of transparency has led to criticism that more efforts should be made to build interpretable models or explanation interfaces.

In recent years there has been lots of work in explainable AI (XAI), which tries to explain typically why trained classification or regression models make the predictions the way they do. The explanation methods can be split into local, which tries to explain why a particular point is classified or regressed as it is, and global, which tries to explain the workings of the entire classifier or regression model. However, most of this work has been in the domain of the non-natural sciences.

This project aims to review methods for interpreting opaque models and approaches to building interpretable models and explanation interfaces. The work should explore the differences between global and local explanations with concrete examples of how they have been applied in sciences, motivate the listener on the importance of explainability and discuss the differences in *explainability* and *interpretability*. Specifically, this project should focus on how explainable AI has been used in sciences and what are the main open problems (and not review XAI in all domains).

References

Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2019) A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys 51(5):1–42, https://doi.org/10.1145/3236009

Labe, Z. M., & Barnes, E. A. (2021). Detecting climate signals using explainable AI with single-forcing large ensembles. Journal of Advances in Modeling Earth Systems, 13, e2021MS002464. https://doi.org/10.1029/2021MS002464

Mark S. Neubauer, Avik Roy, Explainable AI for High Energy Physics, preprint, https://arxiv.org/abs/2206.06632

Simon Letzgus, Patrick Wagner, Jonas Lederer, Wojciech Samek, Klaus-Robert Müller, Gregoire Montavon, 2021, Toward Explainable AI for Regression Models, https://doi.org/10.48550/arXiv.2112.11407

Rasmussen MH, Christensen DS, Jensen JH. Do machines dream of atoms? A quantitative molecular benchmark for explainable AI heatmaps. ChemRxiv. Cambridge: Cambridge Open Engage; 2022; preprint, https://doi.org/10.26434/chemrxiv-2022-gnq3w

Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1, 206–215 (2019). https://doi.org/10.1038/s42256-019-0048-x

Tsai, C.H. and Brusilovsky, P., 2019. Designing Explanation Interfaces for Transparency and Beyond. In IUI Workshops. http://ceur-ws.org/Vol-2327/IUI19WS-IUIATEC-4.pdf

Björklund et al. (2022) SLISEMAP: Explainable Dimensionality Reduction. https://arxiv.org/abs/2201.04455

2.2 Active learning

Active learning is an iterative supervised learning method that levers the knowledge of the user by actively querying the user for data labels. This method is useful in situations, where we have an abundance of unlabelled data, but labelling is expensive. Finding labels might require running costly simulations (e.g., finding molecular orbital energies might require running a quantum chemistry simulation) or performing observations (e.g., when observing the earth system, where should you put your expensive measurement stations and what should they measure?).

In this project, the goal is to review active learning methods used in sciences and explain the caveats and usefulness of different approaches.

References

Karianne J. Bergen, Paul A. Johnson, Maarten V. de Hoop, and Gregory C. Beroza, Machine learning for data-driven discovery in solid Earth geoscience, Science, Vol 363, Issue 6433, 2019, https://doi.org/10.1126/science.aau0323

Berger, K.; Rivera Caicedo, J.P.; Martino, L.; Wocher, M.; Hank, T.; Verrelst, J. A Survey of Active Learning for Quantifying Vegetation Traits from Terrestrial Earth Observation Data. Remote Sens. 2021, 13, 287. https://doi.org/10.3390/rs13020287

Konstantin Gubaev, Evgeny V. Podryabinkin, and Alexander V. Shapeev, "Machine learning of molecular properties: Locality and active learning", J. Chem. Phys. 148, 241727 (2018), https://doi.org/10.1063/1.5005095

Shapeev, A., Gubaev, K., Tsymbalov, E., Podryabinkin, E. (2020). Active Learning and Uncertainty Estimation. In: Schütt, K., Chmiela, S., von Lilienfeld, O., Tkatchenko, A., Tsuda, K., Müller, KR. (eds) Machine Learning Meets Quantum Physics. Lecture Notes in Physics, vol 968. Springer, Cham. https://doi.org/10.1007/978-3-030-40245-7_15

Daniel Reker, Practical considerations for active machine learning in drug discovery, Drug Discovery Today: Technologies, Volumes 32–33, 2019, Pages 73-79, https://doi.org/10.1016/j.ddtec.2020.06.001

Sargsyan, Khachik, Ricciuto, Daniel, and Safta, Cosmin. Earth System Model Improvement Pipeline via Uncertainty Attribution and Active Learning. United States: N. p., 2021. https://doi.org/10.2172/1769699

W. Yao, O. Dumitru and M. Datcu, "An Active Learning Tool for the Generation of Earth Observation Image Benchmarks," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 5720-5723, https://doi.org/10.1109/IGARSS47720.2021.9554198

2.3 & 2.4 Uncertainty quantification

Models are inherently uncertain – both simulators and emulators aim to predict a target quantity, based on, e.g., training data and model parameters. With simulators, we can (in general, at least theoretically) use Bayesian inference to find good approximations for model parameters, produce estimates on model accuracy and ensure the correct behaviour of the model. The performance of an ML emulator is limited by the distribution of the data it was trained on. The degradation of the performance on data from different distributions is called *concept drift* (following the naming by Gama et al. 2014) and it has two main factors: 1) when the modelled relationship changes (*real concept drift*); and 2) when the distribution of the independent variables change (*virtual drift*). The concept drift happens almost always when supervised learning models are used in the real world (including sciences): it is rarely case that a machine learning model is used for data that is exactly from same distribution as it was trained on.t. Since machine learning models are trained on limited data and they are constantly being applied to new data items in changing environments, there is a need to regularly monitor these models due to the above reasons as well as overfitting.

In this project, the goal is to examine and assess different approaches to quantifying uncertainties and to monitor model performance in various fields of science. The project has two sub-topics (which can be divided to separate topics): **(2.3)** how to quantify uncertainty when the underlying distributions do not change (no concept drift), and **(2.4)** how to deal with changing distributions (concept drift).

Notice that there is an inherent connection to topic (2.2) uncertainty quantification. The most straightforward active learning algorithm is to quantify uncertainty in model predictions and then add to the training data those points which have the highest uncertainty.

References

Freno, B.A. and Carlberg, K.T., 2019. Machine-learning error models for approximate solutions to parameterised systems of nonlinear equations. *Computer Methods in Applied Mechanics and Engineering*, *348*, pp.250-296. https://doi.org/10.1016/j.cma.2019.01.024

João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. 2014. A survey on concept drift adaptation. ACM Comput. Surv. 46, 4, Article 44 (April 2014), 37 pages. https://doi.org/10.1145/2523813

Krems (2019) Bayesian machine learning for quantum molecular dynamics. Phys Chem Chem Phys 21: 13392-13410. https://doi.org/10.1039/C9CP01883B

Gabriel A. Pinheiro, Johnatan Mucelini, Marinalva D. Soares, Ronaldo C. Prati, Juarez L. F. Da Silva, and Marcos G. Quiles. Machine Learning Prediction of Nine Molecular Properties Based on the SMILES Representation of the QM9 Quantum-Chemistry Dataset. J. Phys. Chem. A 2020, 124, 47, 9854–9866. https://doi.org/10.1021/acs.jpca.0c05969

Salter JM, Williamson D. A comparison of statistical emulation methodologies for multi-wave calibration of environmental models. Environmetrics. 2016 Dec;27(8):507-523. https://doi.org/10.1002/env.2405

Dominik Seuß. Bridging the gap between explainable ai and uncertainty quantification to enhance trustability. 2021, https://doi.org/10.48550/arXiv.2105.11828

Lange et al. (2020) Machine learning models to replicate large-eddy simulations of air pollutant concentrations along boulevard-type streets. Geosci. Model Dev., in print. https://doi.org/10.5194/gmd-2020-200

Oikarinen et al. (2021) Detecting virtual concept drift of regressors without ground truth values. Data Min Knowl Discov 35: 726–747. https://doi.org/10.1007/s10618-021-00739-7

3.1 Virtual Laboratory in practice

Building a virtual laboratory (VL) can require extensive resources and expertise, but what would be the minimal requirements? What kind of workflow would you need, and how do different infrastructures differ?

In this project, the goal is to assess plans and practicalities (such as HPC workflows) for various virtual laboratory-type infrastructures and to compare the differences and similarities between them (are they compatible). What kind of requirements are typical for a "Virtual Laboratory"? What challenges and opportunities can you see? An interesting question is also related to MLOps, the art of building and running real-world systems that have machine learning components. Is MLOps directly applicable to virtual laboratories in sciences, or do the virtual laboratories have some extra requirements; should there be VLOps or ScienceOps?

References

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundecha, Y., Hutton, C., Lindström, G., Montanari, A., Nijzink, R., Parajka, J., Toth, E., Viglione, A., and Wagener, T.: Virtual laboratories: new opportunities for collaborative water science, Hydrol. Earth Syst. Sci., 19, 2101–2117, 2015, https://doi.org/10.5194/hess-19-2101-2015, 2015

Gürses-Tran, G.; Monti, A. Advances in Time Series Forecasting Development for Power Systems' Operation with MLOps. Forecasting 2022, 4, 501-524. https://doi.org/10.3390/forecast4020028

Klami, Arto; Damoulas, Theodoros; Engkvist, Ola; Rinke, Patrick; Kaski, Samuel (2022): Virtual Laboratories: Transforming research with AI. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.20412540.v1

Lefevre, K., Arora, C., Lee, K. *et al.* ModelOps for enhanced decision-making and governance in emergency control rooms. *Environ Syst Decis* (2022). https://doi.org/10.1007/s10669-022-09855-1

Nativi, S.; Mazzetti, P.; Craglia, M. Digital Ecosystems for Developing Digital Twins of the Earth: The Destination Earth Case. Remote Sens. 2021, 13, 2119. https://doi.org/10.3390/rs13112119

D. A. Tamburri, "Sustainable MLOps: Trends and Challenges," 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2020, pp. 17-23, https://doi.org/10.1109/SYNASC51798.2020.00015

Y. Zhou, Y. Yu, and B. Ding, "Towards MLOps: A Case Study of ML Pipeline Platform," 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2020, pp. 494-500, https://doi.org/10.1109/ICAICE51518.2020.00102

Ratner et al. [Office of Basic Energy Sciences (BES)] Roundtable on Producing and Managing Large Scientific Data with Artificial Intelligence and Machine Learning. United States. https://doi.org/10.2172/1630823

Treveil, M., Dreyfus-Schmidt, L., Lefevre, K., & Omont, N. (2021). Introducing MLOps: How to scale machine learning in the enterprise. O'Reilly Media, Inc.

3.2 Interfacing virtual laboratory

How to interface with a virtual laboratory. How to visualise the data and the models, incorporate human insight and substance area experts (not just machine learning experts) in the model building and so forth.

References

Kobak et al. (2019) The art of using t-SNE for single-cell transcriptomics. Nat Commun 10: 5416. https://doi.org/10.1038/s41467-019-13056-x

Jiang, L., Liu, S. & Chen, C. Recent research advances on interactive machine learning. J Vis 22, 401–417 (2019). https://doi.org/10.1007/s12650-018-0531-1

Verbeeck, N., Caprioli, R.M. and Van de Plas, R. (2020), Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. Mass Spec Rev, 39: 245-291. https://doi.org/10.1002/mas.21602

4.1 Virtual laboratories for physics

This topic is slightly different from the above topics (1.*–3.*), which were centred around specific technical areas (covering many areas of science). Here, you should do the opposite: explore how virtual laboratories and machine learning are (or planned to be used) used in a particular domain, physics, or some of its sub-domains (such as quantum physics).

References

Das Sarma, S., Deng, D.-L., & Duan, L.-M. (2019). Machine learning meets quantum physics. Physics Today, 72(3), 48–54. https://doi.org/10.1063/PT.3.4164

Carleo et al. (2019) Machine learning and the physical sciences. Rev Mod Phys 91: 045002. https://doi.org/10.1103/RevModPhys.91.045002

Dral (2020) Quantum Chemistry in the Age of Machine Learning. J Phys Chem Lett 11(6): 2336–2347. https://doi.org/10.1021/acs.jpclett.9b03664

Interesting articles from CERN about machine learning: https://home.cern/tags/machine-learning

4.2 Virtual laboratories for X

Replace "physics" in 4.1 with some other domain.