

Jun 03, 2025

Protocol for Developing the Sustainable Scientific Machine-learning Reporting Toolkit (SSMART)

DOI

<https://dx.doi.org/10.17504/protocols.io.4r3l2pkrqg1y/v1>

Evangelos Danopoulos¹, John A. D. Aston¹, Paul J. Calleja², Sylvain Laizet³, Richard G. McMahon⁴, Jessica Montgomery⁵, Georgios Rigas³, Konstantina Vogiatzaki⁶, Michèle eiland⁷

¹Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, UK;

²Cambridge Research Computing Service, University of Cambridge, Cambridge, UK;

³Department of Aeronautics, Imperial College London, South Kensington, London, UK;

⁴Institute of Astronomy, University of Cambridge, Cambridge, UK;

⁵Department of Computer Science and Technology, University of Cambridge, Cambridge, UK;

⁶Department of Engineering Science, University of Oxford, Oxford, UK;

⁷EPCC, The University of Edinburgh, Edinburgh, UK



Evangelos Danopoulos

University of Cambridge

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

[Create free account](#)

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.4r3l2pkrqg1y/v1>

Protocol Citation: Evangelos Danopoulos, John A. D. Aston, Paul J. Calleja, Sylvain Laizet, Richard G. McMahon, Jessica Montgomery, Georgios Rigas, Konstantina Vogiatzaki, Michèle eiland 2025. Protocol for Developing the Sustainable Scientific Machine-learning Reporting Toolkit (SSMART). [protocols.io https://dx.doi.org/10.17504/protocols.io.4r3l2pkrqg1y/v1](https://dx.doi.org/10.17504/protocols.io.4r3l2pkrqg1y/v1)

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: June 02, 2025

Last Modified: June 03, 2025

Protocol Integer ID: 219319

Keywords: reporting guideline for scientific machine learning, sustainable scientific machine, learning reporting toolkit, scientific machine learning, scientific machine, energy use in sml, reporting toolkit, based reporting guideline, current state of reporting method, transparent reporting for sml, reporting method, sustainable, smart checklist, development of the sustainable, energy use consideration, sustainability, incorporation of energy use consideration, energy use, transparent reporting, smart, researcher, adoption of smart, systematic review, research, comprehensive dissemination plan, physical science, research quality

Funders Acknowledgements:

EPSRC

Grant ID: EP/Y004841/1

EPSRC

Grant ID: EP/Y004450/1

EPSRC

Grant ID: EP/Y005619/1

EPSRC

Grant ID: EP/Y005619/1

Abstract

This protocol describes the development of the Sustainable Scientific Machine-learning Reporting Toolkit (SSMART). This will be a checklist-based reporting guideline for scientific machine learning (SML) based studies in the physical sciences. The development process will consist of five phases. In phase one, a systematic review will be executed to appraise the current state of reporting methods and energy use in SML-based studies. In phase two, an online Delphi exercise will be used to identify the items to be considered for the checklist. In phase three, during a consensus meeting, the items to be included in the SSMART checklist will be finalized. Phase four will involve developing and writing up the main paper where SSMART will be reported, and an explanation and elaboration paper providing more detailed information. During the final phase, a comprehensive dissemination plan will be implemented to promote the adoption of SSMART by a diverse range of users. We anticipate that SSMART will help to establish the incorporation of energy use considerations and transparent reporting for SML-based studies and promote sustainability. Moreover, SSMART will help researchers report key aspects of their work so that readers can interpret their findings, understand limitations, and appraise research quality.

Guidelines

Scientific machine learning (SML) is a branch of AI for science. SML draws from computer science, computational science, physics and engineering^{13,14}. This emerging research area of ML has the unique benefit that by using elements of physics-based modelling (mechanistic modelling), which are grounded on physical laws, it can restrain the boundaries of the analysis and outputs^{15,16}.

The rapid development of SML methods and the use of ever more complex tools creates a tangible problem of model interpretation, transparency, reproducibility and validation¹⁴. The complexity of SML includes aspects of both system complexity and computational complexity³. Many existing publications fail to report key information (e.g. datasets used for training, code, scripts) hindering usability and accessibility¹⁷. These problems are ubiquitous in all fields of AI and ML. Efforts have been made to resolve these issues by developing reporting guidelines (RGs), either focusing on discipline specific scientific literature, such as studies using AI in medical imaging (CLAIM¹⁸), or clinical trials evaluating interventions with an AI component (CONSORT-AI¹⁹), while others are more broad, attempting to encompass all ML-based studies in science²⁰. The development and usefulness of RGs have a long and successful history, with many being endorsed by scientific journals (as a prerequisite for submission), research funders, industry and government^{11,21}. To our knowledge a domain specific RG for SML-based research does not exist.

Moreover, as far as we know, energy consumption, energy usage optimization and sustainability are not part of any existing RGs in this field. Awareness around potential energy related issues as well as the wider environmental and societal impacts is rising^{22,23}. Methods for estimating and measuring ML energy consumption are being developed^{24,25} while estimation of the carbon footprint of computational science is gaining traction²⁶. Being transparent around energy use is key, especially in this field where the extended use of cloud computing might foster hidden environmental impacts, for example where on site and off site processors are used.

A further aspect that is often neglected by scientific papers is quality assurance of products and services, which is the cornerstone of quality management. RGs are closely related to quality assurance which is interlinked to the use of such systems in industry²⁷. There are available quality standards for software systems (e.g. ISO/IEC 25002:2024²⁸, ISO/IEC 25010:2023²⁹, ISO/IEC 25019:2023³⁰). Nevertheless, SML, compared to traditional software engineering, is more complex³¹. Further aspects such as reusability and trustworthiness must be addressed. A recent report produced by scientists working in industry proposed that quality assurance of AI/ML systems or models should look at five different aspects: data integrity, model robustness, system quality, process agility, and customer expectation²⁷. Adhering to RGs constitutes and evidences quality assurance.

Troubleshooting

Before start

A distinction between the concepts of artificial intelligence (AI), machine learning (ML) and scientific-ML (SML), must be made early on, to avoid ambiguity of terms and methods, while recognizing that definitions have evolved over recent years¹. The Royal Society has provided two laconic definitions: AI is "*is an umbrella term for the science of making machines smart.*", while ML is "*a set of rules that allows systems to learn directly from examples, data and experience.*"². ML is a subset or more precisely one of the available pathways to AI.

SML is a separate field of ML which draws from both scientific computing and machine learning to develop new methods and is based on scientific data sets³. It is a cross-cutting research area combining physics-based models, ML and/or AI predominantly in the physical sciences³. SML draws from computer science, computational science, physics and engineering.

Aim of Reporting Guideline

- 1 Here we propose the development of the Sustainable Scientific Machine-learning Reporting Toolkit (SSMART); a reporting guideline (RG) to aid authors to thoroughly describe all elements of a scientific machine learning (SML)-based study. The RG will comprise a set of items concerning reporting and quality recommendations presented in the form of a checklist. The items in the checklist will include areas that are identified as potential sources for introducing risk of bias (systematic errors) and areas for energy use reporting and optimization. The checklist will be structured in three topics: the data used (evidence base), the methods used (models) and the energy consumption (for training and inference). In this sense the first part of the RG should be subject specific i.e. what kind of data and inputs are available and are being used in the specific fields we are interested in. On the other hand, the items included in the second and third part should in principle be very similar to other relevant fields.

The development methodology is taken from Health research, where RGs have been widely implemented and tested for a range of study types in different fields^{4,5}. In addition, the RG will incorporate, the FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles for scientific data management and stewardship⁶.

- 1.1 The benefits from developing and implementing SSMART include:
 - Improving completeness and transparency of research reports.
 - Enabling reproducibility, verification and interpretation of methods and outcomes.
 - Reducing the amount of poorly reported studies and research waste.
 - Improving reporting of energy use in SML-based studies.
 - Improving energy efficiency of ML algorithms (software and hardware)
 - Reducing carbon footprint of the hardware used for SML.
 - Aiding editors and peer reviewers in manuscript reviewing.
 - Improving transparency for funded research by linking clear aims and outcomes.

Methods

2 Working groups

Guideline development requires participation of a broad network of scientists, researchers as well as key stakeholders such as research funders and industry. We will first draw expertise from two ongoing UKRI/ EPSRC funded projects which are interdisciplinary collaborations of four universities in the UK⁷. These were the starting point for creating the scientific community that will inform the development of SSMART. In addition, both projects involve key policymakers and industry partners.

The driving force and validity of the guidance will be the participants. The people participating in the process can be categorized into four groups, which are not mutually exclusive. In fact, it is anticipated that people will participate in more than one group. The executive committee is the steering group, setting the objectives of the RG and in charge of the overall organization of the project. A Delphi panel is going to be used to identify the list of potential items to be included in the RG checklist. The advisory group will consist of the participants of a consensus meeting where the RG will be formally discussed, and consensus will be reached on the items to be included in the checklist. Finally, the writing group is going to undertake the writing of the RG and accompanying documents (Fig. 1).

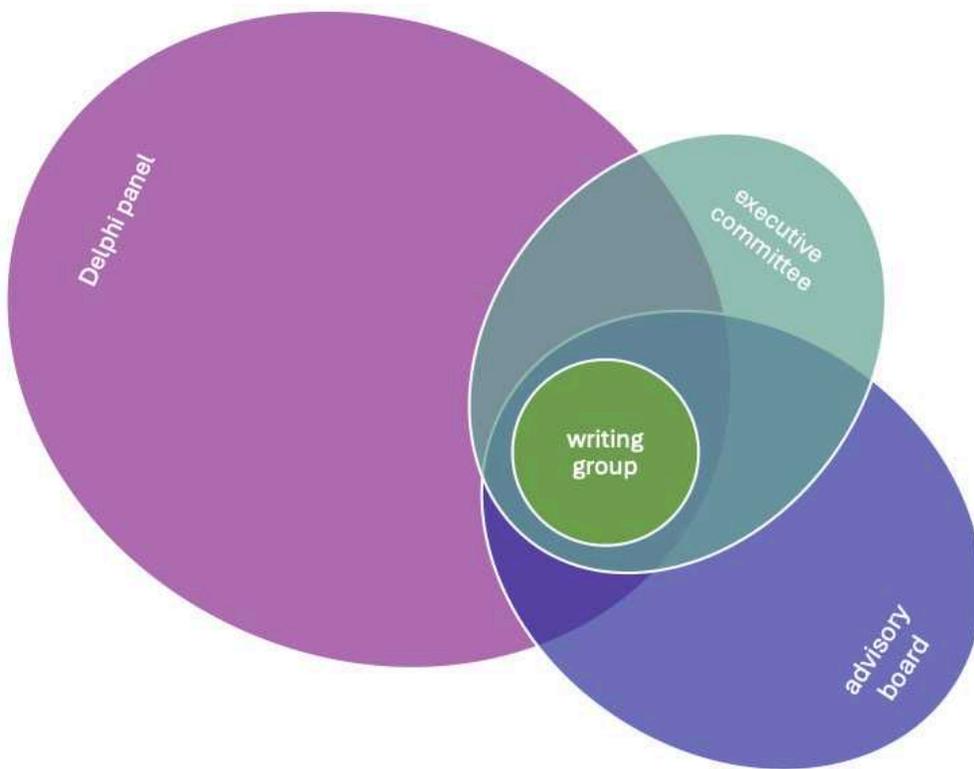


Fig. 1. Sustainable Scientific Machine-learning Reporting Toolkit (SSMART) development groups. The overlapping circles of the Venn diagram illustrate that people might participate in more than one of the four development groups. The size of the circles approximates the anticipated relative size of the groups.

3 Work plan

The development of SSMART will follow the guidance provided by the EQUATOR Network for research RGs and other relevant published protocols^{4,5}. According to the guidance, the development process should be structured in distinct phases, each

comprised of a series of steps. A summary of the five phases we are going to follow, setting out the development timeline is illustrated in Fig. 2.

3.1 Initial steps

RG development is a formal process that should be underpinned by a protocol that is publicly available^{4,5}. Developing and disseminating a protocol ensures transparency and underpins our integrity pledge. In addition, by publishing the protocol we aim to inform the scientific community of this major undertaking and endeavour to attract wider participation.

A key part of the initial steps phase is executing a systematic review to appraise the quality of current reporting within the domain of SML based studies. The first objective of this systematic review will be to evaluate the methodological conduct and reporting of SML based studies in the fields of physical sciences. The second objective will be to identify and evaluate the metrics and tools that are being used for tracking energy use and to appraise the reporting quality of energy use for ML models in this scientific domain. The systematic review will provide important insights on the quality of current reporting practices and will be the primary evidence base for the items to be considered for the checklist of the RG. Existing reporting guidelines for AI/ML based studies will also be identified and used in complementing the evidence from the systematic review.

3.2 Pre-meeting activities

The pre-meeting activities will be planned around the execution of a Delphi exercise. This is a methodology for developing consensus between members of a panel through structured communications^{8,9}. Delphi is based on sequential questionnaires that are refined by feedback provided from the previous round. We aim to include two rounds so that the members can achieve consensus. If consensus has not been reached after the second round, an optional third round will be executed at the discretion of the executive committee.

A ranking-type questionnaire will be sent to all the participants forming the Delphi panel, in which they will be asked to rank the importance of a series of reporting and quality items to be considered for the checklist of the RG. In the second round, the results from the first round will be presented and participants will be asked to agree on the top-ranking items. Free text boxes will be available in both rounds for providing feedback. The results of the Delphi panel, from all rounds, will be summarised in tabular form and accompanied by a narrative summary prepared by the executive committee and presented to the advisory board during the consensus meeting. Items where ranking agreement was not reached will also be included in this summary for further consideration during the consensus meeting.

The members of the Delphi panel should be as diverse as possible to achieve maximum participation from all interested parties. This is one of our key goals as it will foster maximum consensus within the community and ensure future adoption and usefulness of the RG. As such, we will invite researchers, academics, funders, journal editors, policymakers and industry actors that have a direct or indirect interest in SML to participate in the Delphi panel.

The Delphi exercise will be executed according to the methods of empirical survey-based research¹⁰. As such ethical approval will be sought before the start of the survey according to the ethics of research policy of the academic institution of the lead author/s. Participants will be provided with all the necessary information on the purpose and the particulars of the survey at the time of the invitation via email. If they chose to take part, a further participant information sheet and an adjacent consent form will be provided.

3.3 **Consensus meeting**

The purpose of the consensus meeting is for the advisory board to discuss the outcomes of the Delphi exercise and agree on the final list of items to be included in the checklist. The meeting will be hybrid with members joining both in person and online in order to accommodate different needs and be inclusive.

The advisory group will be formed by members of the executive committee and members of the Delphi panel. As such, all the different categories of the Delphi panel members' groups should ideally be represented. The development process is anticipated to have a long timeline. Acknowledging that circumstances may change, further field experts may be invited to participate in the advisory group according to availability. The outcome of the consensus meeting will be the final list of items to be included in the SSMART checklist.

3.4 **Post-meeting activities**

The writing group will use the items agreed during the consensus meeting and will operationalise them in a checklist. The structure of the checklist and the possible inclusion of an accompanying flow diagram will also be discussed. The writing group will consist of members of the executive committee while additional members from the Delphi panel and/or the advisory group may be invited.

SSMART will be reported in two papers, as is the norm for RGs dissemination⁵. The first paper will describe the development of SSMART and provide the checklist, while a second paper will offer further explanation and elaboration on how SSMART may be implemented in practice. This second paper will go into further detail around the rationale of the items in the checklist and provide specific examples.

3.5 **Post-publication activities**

The dissemination plan will be instrumental in the visibility and adoption of SSMART. We aim to publish the papers in impactful journals in the fields of SML. Key existing reporting guidelines^{11,12} are endorsed by numerous journals and are actively used in the peer-review process as a prerequisite for submission. We will contact journals that include SML-based studies in their scope and propose the use of SSMART. A website will also be built specifically for the promotion of the guideline. In our publications and on the website, we will actively encourage feedback from the scientific community and all interested stakeholders.

In order to further promote the adoption of SSMART, we will contact the relevant research funding councils in the UK, such as UKRI, and put forward a case of using SSMART specifically as a measure of energy use transparency and sustainability. Furthermore, we will encourage including SSMART in traditional research dissemination plans including presentations in conferences and workshops.

We recognise that there are potential barriers to the adoption of SSMART stemming for example, from the added burden of reporting energy use. We believe that our combined post-publication strategy will help alleviate some of the barriers by convincing all the interested parties of the benefits of using SSMART.

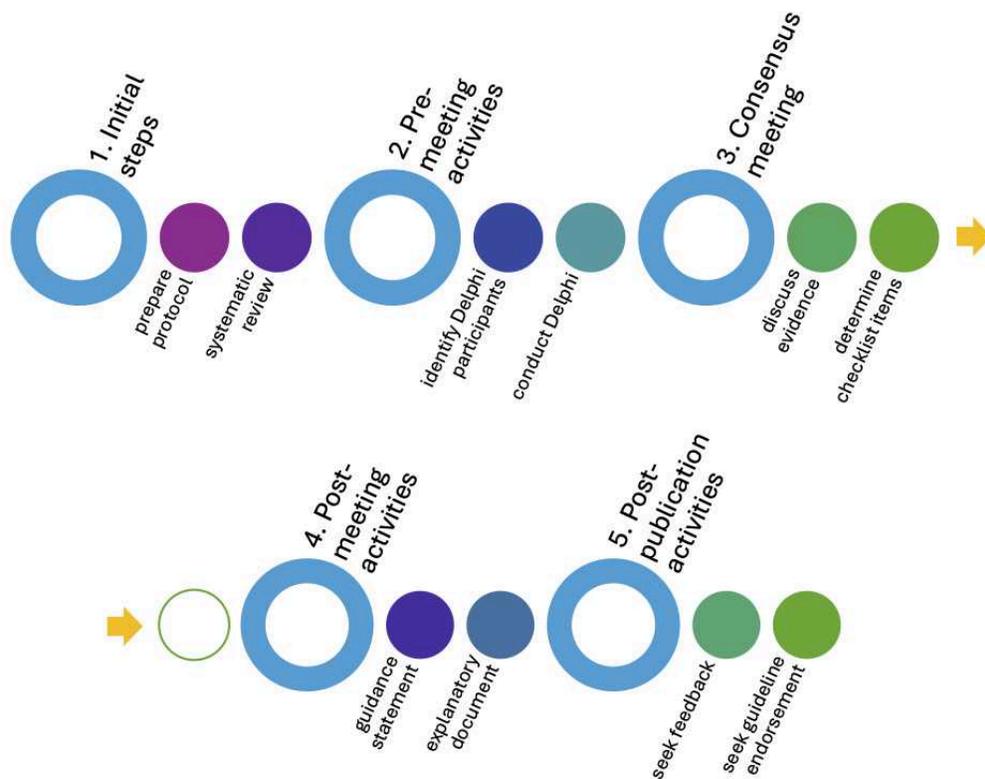


Fig. 2. Sustainable Scientific Machine-learning Reporting Toolkit (SSMART) development timeline.

Conclusions

- 4 SML is increasingly being used in physical sciences. Recent data explosion availability has propelled a revolution in SML methods in an array of diverse applications. The use of SML opens up an exciting new world of opportunities, however, there are associated risks when limits and pitfalls are not recognized. Consistent reporting of the highest standard based on transparency and reproducibility is vital for all concerned audiences. Moreover, consideration and planning around the energy required to run the modeling should start at the beginning of the research planning and continue until the publication of results.

We expect that SSMART will help researchers report transparently the rationale, methods, and results, while at the same time it will put SML energy use and sustainability center stage. Moreover, it will empower peer-reviewers, funders, policymakers, and industry actors to better understand and critically appraise SML-based research outputs.

Protocol references

1. Kühl, N., Schemmer, M., Goutier, M. & Satzger, G. Artificial intelligence and machine learning. *Electronic Markets* 32, 2235-2244, doi:10.1007/s12525-022-00598-0 (2022).
2. The Royal Society. *Machine learning: The power and promise of computers that learn by example.* (2017).
3. Baker, N. et al. *Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence.* Medium: ED; Size: 109 p. (United States, 2019).
4. Gary, S. C. et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 11, e048008, doi:10.1136/bmjopen-2020-048008 (2021).
5. Moher, D., Schulz, K. F., Simera, I. & Altman, D. G. Guidance for Developers of Health Research Reporting Guidelines. *PLOS Medicine* 7, e1000217, doi:10.1371/journal.pmed.1000217 (2010).
6. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018, doi:10.1038/sdata.2016.18 (2016).
7. Imperial College London. *AI for Net Zero; Using Artificial Intelligence (AI) to help the UK's energy and transport sectors achieve carbon net zero,* <<https://www.imperial.ac.uk/ai-net-zero/>> (2024).
8. Nasa, P., Jain, R. & Juneja, D. Delphi methodology in healthcare research: How to decide its appropriateness. *World J Methodol* 11, 116-129, doi:10.5662/wjm.v11.i4.116 (2021).
9. Okoli, C. & Pawlowski, S. D. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management* 42, 15-29, doi:10.1016/j.im.2003.11.002 (2004).
10. Hasson, F., Keeney, S. & McKenna, H. Research guidelines for the Delphi survey technique. *Journal of advanced nursing* 32, 1008-1015 (2000).
11. Schulz, K. F., Altman, D. G. & Moher, D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 8, 18, doi:10.1186/1741-7015-8-18 (2010).
12. Von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The lancet* 370, 1453-1457 (2007).
13. Thiyagalingam, J., Shankar, M., Fox, G. & Hey, T. Scientific machine learning benchmarks. *Nature Reviews Physics* 4, 413-420, doi:10.1038/s42254-022-00441-7 (2022).
14. Hey, T., Butler, K., Jackson, S. & Thiyagalingam, J. Machine learning and big scientific data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378, 20190054, doi:doi:10.1098/rsta.2019.0054 (2020).
15. Iwema, J. *Scientific Machine Learning,* <<https://sciml.wur.nl/reviews/sciml/sciml.html>> (2023).
16. Cuomo, S. et al. Scientific Machine Learning Through Physics-Informed Neural Networks: Where we are and What's Next. *Journal of Scientific Computing* 92, 88, doi:10.1007/s10915-022-01939-z (2022).
17. Roscher, R., Bohn, B., Duarte, M. F. & Garcke, J. Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* 8, 42200-42216, doi:10.1109/ACCESS.2020.2976199 (2020).
18. Mongan, J., Moy, L. & Charles E. Kahn, J. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiology: Artificial Intelligence* 2, e200029, doi:10.1148/ryai.2020200029 (2020).
19. Liu, X. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nature Medicine* 26, 1364-1374, doi:10.1038/s41591-020-1034-x (2020).
20. Kapoor, S. et al. REFORMS: Consensus-based Recommendations for Machine-learning-based Science. *Science Advances* 10, eadk3452, doi:doi:10.1126/sciadv.adk3452 (2024).

21. Page, M. J. et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLOS Medicine* 18, e1003583, doi:10.1371/journal.pmed.1003583 (2021).
22. Bojic, L., Bala, K., Medojevic, M. & Talanov, M. in 2024 9th International Conference on Smart and Sustainable Technologies (SpliTech). 1-4.
23. van Wynsberghe, A. Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics* 1, 213-218, doi:10.1007/s43681-021-00043-6 (2021).
24. García-Martín, E., Lavesson, N., Grahm, H., Casalicchio, E. & Boeva, V. in *ECML PKDD 2018 Workshops*. (eds Carlos Alzate et al.) 243-255 (Springer International Publishing).
25. Yang, T. J., Chen, Y. H., Emer, J. & Sze, V. in 2017 51st Asilomar Conference on Signals, Systems, and Computers. 1916-1920.
26. Lanelongue, L. et al. GREENER principles for environmentally sustainable computational science. *Nature Computational Science* 3, 514-521, doi:10.1038/s43588-023-00461-y (2023).
27. Fujii, G. et al. Guidelines for Quality Assurance of Machine Learning-Based Artificial Intelligence. *International Journal of Software Engineering and Knowledge Engineering* 30, 1589-1606, doi:10.1142/s0218194020400227 (2020).
28. ISO/IEC 25002: . (ISO (the International Organization for Standardization); IEC (the International Electrotechnical Commission), 2024).
29. ISO/IEC 25010: . (ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission), 2023).
30. ISO/IEC 25019: . (ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission), 2023).
31. Siebert, J. et al. in *Quality of Information and Communications Technology*. (eds Martin Shepperd, Fernando Brito e Abreu, Alberto Rodrigues da Silva, & Ricardo Pérez-Castillo) 17-31 (Springer International Publishing).

Acknowledgements

This work was supported by: the UKRI - EPSRC grant: Real-time digital optimisation and decision making for energy and transport systems; the UKRI - EPSRC grant: Towards a more sustainable High Performance Computing sector: a hardware/software co-design proof-of-concept; and the an ERC Starting Grant PhyCo