

Prioritize environmental sustainability in use of AI and data science methods

Caroline Jay^{1,2,3}, Yurong Yu⁴, Ian Crawford⁵, Scott Archer-Nicholls⁶, Philip James⁷, Ann Gledson⁶, Gavin Shaddick^{8,3}, Robert Haines^{2,6}, Loïc Lannelongue^{2,9,10,11,12}, Emily Lines^{3,13}, Scott Hosking^{3,14}, David Topping^{3,5}

Artificial Intelligence (AI) and data science will play a crucial role in improving environmental sustainability, but the energy requirements of these methods will have an increasingly negative effect on the environment without sustainable design and use.

Against the backdrop of an implicit assumption that computational resources will continue to increase in availability and reduce in cost, it is rare for researchers to explicitly consider environmental impact when designing or choosing analysis methods. We believe there is an opportunity for the environmental science community to drive a change in approach, optimising energy usage when conducting their own computational research, and advocating for other research domains to do the same. Considering environmental sustainability in computational research will both accelerate innovation and democratise it: regions most affected by climate change – and where local research could have huge benefits – are less likely to have access to significant computational resources. Making energy efficiency and sustainability a primary consideration will also catalyse innovative methodological approaches to scientific research. By aligning these changes with domain-based understanding of the scientific need, we can set standards for best practice in a strategic way.

Energy requirements of computational methods

Net Zero is defined as a state in which anthropogenic emissions of greenhouse gases into the atmosphere are balanced by anthropogenic removals out of the atmosphere. Achieving Net Zero requires the coordination of efforts across social, political, economic and technological spheres¹. AI and data science will play a crucial role in this complex process, helping us to understand and ultimately optimise anthropogenic energy usage².

Competing with this potential benefit, is the fact that AI and data science have huge energy and environmental costs themselves^{3,4}. The resource requirements of AI research, development and application are continually increasing, and nations are under pressure to invest in ever larger computational facilities to keep pace⁵. The potential environmental impact of this is acknowledged, motivating efforts to make computing more sustainable, including the adoption of more energy-efficient hardware, better management of data centres, and the use of renewables to power systems⁶. The role of software architecture is also recognised as important, and that making progress with this will require users skilled in writing efficient code optimised for minimum environmental impact⁵.

There are some initiatives promoting energy efficient software for research (for example, <https://greensoftware.foundation/>), along with efforts to develop high level principles for computational scientists⁷. Nevertheless, the environmental sustainability of methods is not currently a major consideration of any part of the computational science research community, and there are few guidelines for those wishing to develop or use AI and data science in a sustainable way.

Embedding environmental sustainability in AI.

The inaugural meeting of the [Alan Turing Institute Environment and Sustainability Interest Group](#), held in Manchester on March 15th, 2022, culminated in an open discussion about the

relationship between AI and data science and environmental sustainability. It considered the potential for AI and data science to help reach Net Zero, as well as how AI and data science might negatively impact our ability to reach Net Zero and how we might maximise the environmental benefits whilst reducing the energy costs of research. Overwhelmingly, the view was that this issue is not considered in most environmental science research. It was proposed that environmental scientists should not use AI and data science to understand and improve our environment, without also considering the energy efficiency and sustainability of the analyses themselves.

Following this discussion, the Alan Turing Institute Environment and Sustainability Interest Group have worked with the [Software Sustainability Institute](#) to propose a series of recommendations to the environmental science research community, with the aim of helping to embed environmental sustainability in future computational research.

Minimise the environmental impact in methods choice. Rather than relying on increasing computing power, researchers should consider upfront the carbon footprint of their approach and be able to explicitly justify the carbon budget of any computational research. Researchers should continually reflect on the methods used and consider whether data-intensive machine learning is necessary. Whenever researchers develop new methods, or apply existing ones, they should ask themselves whether they are using the most energy-efficient approach. A simpler, less energy-intensive method than machine learning may be sufficient for the aims of the research. They should question whether a small increase in accuracy is worth the extra computing resources required.

Prioritise research on the environmental impact. We know anecdotally that some methods are more efficient than others, and that there are trade-offs between readability of code and computational efficiency, and human vs. computing resources (for example, carefully curated data can be processed faster than noisy raw data). However, the environmental science community lacks proper guidance as to how much carbon will be used or saved as a function of a methodological approach. There is also a lack of understanding of how processor usage correlates with power usage, which makes estimating carbon footprints challenging. Systematically examining the environmental footprint of these design choices would help inform decisions on which approaches to use. Such systematic investigations require an interdisciplinary approach, considering human and organisational aspects of the complex system that makes up the research process, as well as the technical. Environmental scientists have a key role to play in motivating and contributing to this research.

Develop and disseminate best practice. Environmental sustainability of computing methods may not be considered in the wider research community because the environmental impact is not fully understood. Environmental scientists should raise awareness of this issue, both within the domain and when working with other domains – particularly those developing AI and data science tools. This might include introducing the topic when planning a project, suggesting it is considered as a factor when choosing methods, and discussing their own approach to reducing energy consumption. Ultimately, the environmental science community should provide guidance on how to employ tools sustainably.

Champion sustainable science beyond academic research. Researchers should choose industrial partners committed to environmental sustainability themselves, and argue for energy efficiency in AI and data science when working with others. They should also

encourage research funders, government and industry to embed relevant policy and practice across all areas of AI and data science.

Improve the accessibility of methods. Policy change takes place at regional scales and works best when it draws on local scientific data and intelligence. Developing data-driven insights requires access to open and robust digital tools. Yet, areas of the world likely to be most affected by climate change often have much lower investment in digital infrastructure, and as such the energy requirements of AI and data science methods are a critical barrier⁸. Local and municipal authorities in all nations lack access to significant computational resources when making policy decisions about towns and cities⁹. Collecting and analysing data at a local level ensures it is appropriately contextualised and understood. Low energy methods therefore not only help the environment, but democratise access to AI, as they are cheaper to use and have lower infrastructure requirements.

Leading by example

These recommendations are aimed at those working in environmental research, who can lead by example by adopting or championing them as appropriate. However, progress will require more than scientist buy in. We recognise that a better understanding of the environmental impact of computational methods will require a collaborative approach, combining expertise from data scientists, software engineers and domain specialists. We also suggest that funding bodies follow these recommendations as a practical starting point towards minimising the environmental impact of research.

As our understanding of the energy footprint of analysis design decisions improves, the continued development and sharing of best practice will feed into funder policy and AI regulation. We ultimately hope our recommendations serve as a starting point for developing robust standards that will minimise the environmental impact – and facilitate the accessibility – of AI and data science innovation, with benefits for both the global research community and the world at large.

Caroline Jay^{1,2,3}, Yurong Yu⁴, Ian Crawford⁵, Scott Archer-Nicholls⁶, Philip James⁷, Ann Gledson⁶, Gavin Shaddick^{8,3}, Robert Haines^{2,6}, Loïc Lannelongue^{2,9,10,11,12}, Emily Lines^{3,13}, Scott Hosking^{3,14}, David Topping^{3,5}

¹School of Engineering, University of Manchester, Manchester, UK

²Software Sustainability Institute

³Alan Turing Institute

⁴Centre for Environmental Policy, Imperial College London

⁵School of Natural Sciences, University of Manchester, Manchester, UK

⁶Research IT, University of Manchester, Manchester, UK

⁷School of Engineering, University of Newcastle, Newcastle, UK

⁸School of Engineering, Physical and Mathematical Sciences, Royal Holloway University of London, London, UK

⁹Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

¹⁰British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

¹¹Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK

¹²Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK

¹³Department of Geography, University of Cambridge, Cambridge, UK

¹⁴British Antarctic Survey

Corresponding author

Correspondence to Caroline Jay: caroline.jay@manchester.ac.uk

Acknowledgements

The workshop that generated the ideas for this comment was supported by the Alan Turing Institute.

Caroline Jay was supported by the Software Sustainability Institute grant EP/S021779/1.

Caroline Jay and David Topping were supported by the NERC Digital Solutions Programme.

Loïc Lannelongue was supported by core funding from the British Heart Foundation (RG/18/13/33946); the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014; NIHR203312)[*]; the Cambridge British Heart Foundation Centre of Research Excellence (RE/18/1/34212); and the BHF Chair Award (CH/12/2/29428).

*The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

References

1. Fankenhauser, S. et al. *Nat. Clim. Chang.* **12**, 15-21 (2022).
2. Conner, A. et al. *Tackling climate change with data science and AI*. (The Alan Turing Institute, 2023).
3. Al Kez, D. et al. *J. Clean. Prod.* **371** 133633 (2022).
4. Freitag, C. et al. *Patterns* **2** 100340 (2021).
5. Gharamani, Z. *Independent Review of The Future of Compute: Final report and recommendations*. (GOV.UK, 2023).
6. Royal Society. *Digital technology and the planet: harnessing computing to achieve net zero*. (Royal Society, 2020).
7. Lannelongue, L. et al. *Nat. Comp. Sci.* **3**, 514-521 (2023).
8. United Nations. *Connectivity in the least developed nations: status report 2021*. (UN, 2021).
9. Open Data Manchester. *Digital Solutions Hub: User Research Report*. (ODM, 2023).

Competing interests

The authors declare no competing interests.

Figure 1: Recognizing the environmental impact of computational environmental science research. The environmental impact should be considered at every stage of research, from data collection to knowledge dissemination. With the rapid development of AI tools and the increasing energy demands of these methods, there is a particular opportunity to improve the environmental sustainability of the data processing stage, and a particular responsibility for environmental scientists to lead by example.