Research, reuse, repeat

A research paper makes the most impact when its methods, data and code are available for others to use and build on. We highlight the benefits of good sharing practices with a new type of article, reusability reports.

Artificial intelligence approaches have found many uses in scientific applications. In the past decade, machine learning developments in particular have transformed research areas such as medical imaging, protein structure prediction and materials discovery. We are often impressed by authors who embrace these opportunities and develop high-quality machine learning tools and pipelines to tackle specific challenges and applications in their main area of research.

An essential part of research progress is that these new tools are made available to others who can replicate the results and build on the work. It is our editorial policy that code, when essential to the main findings of the paper, should be shared with reviewers and upon publication with all readers. Many of our authors are keen to share code, and to do this well. They often expend substantial efforts to present well-structured and complete code repositories with user-friendly instructions (see, for instance, the code repository associated with this recent paper published in *Nature Machine Intelligence*). Several authors produce a containerized version of their code, or 'compute capsules', so that interested users can run the code without the need to install various software dependencies. We provide assistance to authors with these efforts. In particular, we have partnered with Code Ocean to assist authors in setting up compute capsules. This not only has advantages for our readers, but also facilitates code peer review. In our collaboration with Code Ocean, referees are given access to a private peer review capsule that contains the code, data, results and computational environment to reproduce the results in the paper. Authors can opt for single blind or double blind review of the

code. Upon publication, a permanent Code Ocean capsule is made available free to all, minted with a DOI (see an example here).

To highlight and amplify the benefits of making high-quality code — or compute capsules — available to other researchers, we introduce a new type of article, reusability reports. These articles are short research papers that analyse and report on the reusability of code from a previously published paper in the context of new data or in a new scientific application. The current issue sees a first example of a reusability report. The code is from a paper published earlier this year and concerns a generative approach to design new molecules for potential use in medical chemistry. In the original paper, the authors learn the text representation of compounds that have a high likelihood of showing bioactivity against a given dopamine receptor. The reusability report authors explore the possibility to use and extend the code in a different area of molecular design, namely for photovoltaic materials. They discuss advantages and limitations of applying the code in a new setting, as the size of the considered molecules, as well as the desired features, are very different. They also explore future directions for extending the code, such as efficient transfer learning strategies, the use of alternative string representations of molecules or employing neural network models that capture more long-range relationships.

We intend to publish several reusability reports next year and are keen to receive feedback on this format. The articles are peer-reviewed and currently written by invitation. Our initiative takes place in the wider context of a growing focus on reproducibility in the machine learning community. In an important development

last year, the Neural Information Processing Systems (NeurIPS) annual meeting, the largest machine learning conference, introduced a reproducibility checklist for submissions. Although it is not yet mandatory, this year NeurIPS authors are strongly encouraged to share code. A further step is the NeurIPS Reproducibility Challenge, which was also introduced in 2019. For this challenge, which runs again this year, researchers are invited to 'claim' a paper that is accepted at NeurIPS (and where authors have shared their code) and to test the reproducibility of the paper's results. The 2019 version was deemed a success with 173 papers claimed for testing and over 70 institutes taking part. The organizers observe that the reports submitted in the challenge go beyond straight reproducibility checks; rather, challenge authors often provide a detailed and nuanced account of their efforts, offering valuable information for other researchers interested in reproducing or building on previous machine learning work.

This is a theme that we will focus on with our introduction of reusability reports. They are not intended as reproducibility or replication checks of the original paper's claims. Rather, they explore how robust, generalizable and reusable the code is and highlight new possibilities for researchers who are interested in further developing the code and using it on new research data. Research outputs make the most impact when they are shared and presented so that others can reuse and build on them. We will continue to support authors in their efforts to achieve this goal. ❐