CARBON FOOTPRINT:
THE (NOT SO) HIDDEN COST
OF HIGH PERFORMANCE COMPUTING

Scientists, empowered by huge amounts of computing power, storage and memory are making world changing discoveries — including to help combat climate change. Loïc Lannelongue, a PhD student at the University of Cambridge, explores how high performance computing itself can lighten its carbon contribution.

There is something a bit surreal about having dozens of CPUs and gigabytes of memory at your fingertips. Yet, this is what high-performance computing (HPC) is all about: by centralising servers in data centres and opening safe channels of communications with the user, HPC facilities have drastically reduced the burden of purchasing and maintaining supercomputers. For users, it has made it easier than ever to access these resources; any laptop, tablet or even smartphones can do it painlessly.

Cloud computing platforms, such as Amazon's AWS, Google's GCP or Microsoft's Azure are the most popular options, but most research institutions and companies have their own data centre and follow the same principles.

On the plus side, such developments in computing have enabled discoveries like the first direct image of a black hole 55 million light-years away, weather forecasts more accurate than ever and the discovery of thousands of genetic variants related to diseases. To measure the magnitude of HPC usage, we can look at the usage statistics of XSEDE, the Extreme Science and Engineering Discovery Environment, a network of data centres in American universities used for scientific research.

In 2020, every hour, one million compute hours were performed on the network, for a total of nine billion compute hours.

What's the issue with this? After all, this leads to ground-breaking discoveries. It also requires mountains of hardware and the electricity to power it, which comes at a significant, yet largely ignored, environmental cost.

It is estimated that data centres have a yearly carbon footprint of 100 megatons of CO2e (carbon dioxide equivalent — a way to quantify the global warming impact of an activity) just from electricity production — similar to the entire American commercial aviation sector.

Unsurprisingly, this will only increase in the next few years, by two-to-nine-fold in the next decade, some studies say.

A variety of environmental impacts
There are numerous ways in which large computing facilities impact the environment: energy production, hardware manufacturing, long-term storage management, cooling, maintenance, etc.

notoriously bad for the environment due to the extraction of precious metals. In the case of consumer devices, as much as 70 to 80% of the total carbon footprint is from manufacturing, with usage and disposal only responsible for the remaining 20 to 30%. This shows the importance for us to try to keep, repair and reuse our devices as much as possible. In data centres, it's also important to factor in the environmental impact in the renewing cycle of hardware.

Artificial intelligence: big models, huge carbon footprint
'Training a single AI model can emit as much carbon as five cars in their lifetimes.'

> **"It is estimated that data centres have a yearly carbon footprint of 100 megatons of CO2e just from electricity production — similar to the entire American commercial aviation sector."**

Energy usage is perhaps the aspect most discussed in the media. The environmental cost of powering data centres depends directly on the carbon footprint of energy production, which varies greatly with the energy mix of the country where the data centre is located. For example, producing 1 kWh of electricity in Switzerland (powered mainly by hydro) emits 12g CO2e on average, but 253g CO2e in the UK and 880g CO2e in Australia (where coal is the main source of energy). Consequently, this means that exactly the same task will have a carbon footprint 73 times greater in Australia compared to Switzerland.

However, this is far from being the only downside. Manufacturing IT hardware is

This headline was on the front of all the tech newspapers in the summer of 2019, from the *MIT Technology Review* to *New Scientist*. This followed an article studying the carbon footprint of algorithms trying to comprehend natural language, an extremely difficult task where it is common for algorithms to run for days or even weeks.

Similar concerns were raised in articles such as 'Green AI' and 'On the dangers of stochastic parrots: can language models be too big?' — in which researchers discuss the issue of accessibility (what happens to research if only a handful of tech companies can afford to develop such models?) and the risks posed by these technologies. In particular, they point out

that the populations suffering most from the environmental cost of AI also benefit the least from innovations such as Apple's Siri or Amazon's Alexa.

Notwithstanding the ethical issues arising if such tech companies are the only ones overseeing language models underpinning numerous aspects of society. Almost to illustrate this point, Google disbanded the Ethical AI team, who participated in writing the stochastic parrots article shortly after its release.

These different studies led to the development of several tools aimed specifically at estimating the carbon footprint of machine learning models, such as the ML CO2 Impact and the Experiment Impact Tracker.

AI may be the tree that hides the forest

The ever-growing computational needs of artificial intelligence are a real source of concern, but we shouldn't forget about all the other fields of science that also rely heavily on computations.

As discussed at the start, complex algorithms are everywhere, from genomics to physics and astronomy. This realisation that carbon footprint should be accounted for by all scientists using algorithms, not only in AI, is what led the development of a broad initiative: the Green Algorithms Project. This is a theoretical framework (another one!) to estimate the carbon footprint of any computation, but more importantly, an online calculator to do the estimations easily. This helped shed light on the carbon footprint of bioinformatics for example, where tools in genomics or molecular simulations emit kilograms of CO2e at each utilisation.

Since then, Australian astronomers have also investigated their environmental impact and found that when taking into account all sources of emissions (travel, computations, offices, etc.) the carbon footprint of an astronomer averages over 37 tonnes of CO2e per year — five times the global average. In several recent publications: 'An astronomical institute's perspective on meeting the challenges of the climate crisis' and 'The ecological impact of high-performance computing in astrophysics', astronomers urge their community to reduce this impact.

What about cryptocurrencies? The power usage of the so-called "mining farms" is estimated by the Cambridge Bitcoin Electricity Consumption Index at ~100 TWh per year. For comparison, if Bitcoin was a country, it would rank just behind the Netherlands in terms of energy usage, or alternatively, these 100 TWh could power all the kettles in Britain for 22 years. There is also the rapid renewing cycle of the dedicated application-specific integrated circuits (ASIC) which creates mountains of technological waste that's difficult to recycle.

This massive carbon footprint can't be ignored and, similarly to other computation, it will be important to weigh the pros and cons of this technology moving forward.

Why don't we care?

When the computer is a loud tower in a corner of the room, the countless cables and the noisy fan are a reminder of the sort of energy needed to power it. On the other hand, the physical distance between users and data centres, which makes running demanding algorithms easier than ever, also makes us forget about the resources used behind the scenes.

Besides, energy efficiency and carbon footprint are rarely broached when discussing computing hardware. Fridges, dishwashers, TVs, cars — all these devices promote their low energy needs and sustainability. Being (or claiming to be) environmentally friendly is now a key element of an advertising campaign (as well as a legal requirement in many cases). This is not yet the case for computers, for which mostly speed and performance is discussed, and when energy efficiency is brought up in non-IT-specialist communities, it's usually to promote battery life and fast charging, not sustainability.

If you add to that the fact that computing time is virtually free for a lot of scientists, it's no surprise that it hasn't been addressed earlier. Although, until recently, it was excessively difficult for scientists who would like to estimate their carbon footprint to do so, as it required extensive research on the environmental impact of each hardware component. Without data, the problem has mostly been ignored, but hopefully with tools such as Green Algorithms, this will start to change.

What now?

Does that mean that scientists should stop using algorithms for their work? Of course not, it has and will continue to enable amazing discoveries, including in the fight against climate change. However, just as a financial cost-benefit analysis is included prior to the inception of any project, the environmental cost should similarly be considered.

Besides estimating and acknowledging their impact, there are a number of things scientists can do to limit the carbon footprint of their work, such as factoring in sustainability in hardware and software choices, optimising (or using optimised) code and avoiding unnecessary computations.

In the coming years, high performance computing will certainly both continue to contribute to global warming and help to fight it. Only by being transparent about the carbon footprint of computation and by doing our best to reduce it, can we ensure that the net result benefits society, globally.

Ten simple rules to make your computing more environmentally sustainable*

Rule 1: Calculate the carbon footprint of your work.
Rule 2: Include the carbon footprint in your cost–benefit analysis.
Rule 3: Keep, repair, and reuse devices to minimise electronic waste.
Rule 4: Choose your computing facility.
Rule 5: Choose your hardware carefully.
Rule 6: Increase efficiency of the code.

Rule 7: Be a frugal analyst.
Rule 8: Releasing new software? Make its hardware requirements and carbon footprint clear.
Rule 9: Be aware of unanticipated consequences of improved software efficiency.
Rule 10: Offset your carbon footprint.

*Read the full paper: https://doi.org/10.1371/journal.pcbi.1009324