ON-PREM

AI's enormous energy appetite can be curbed, but only through lateral thinking

Nothing will change while big tech sets the rules. We'll need someone even scarier

Rupert Goodwins

Tue 27 May 2025 // 08:30 UTC

OPINION How much harm does AI cause the environment? As a report from the MIT Technology Review just confirmed, nobody knows, and almost nobody cares enough to try and find out. Even if lots of people did care a lot, it wouldn't change things. The driver of AI's insane energy addiction is no more amenable to argument than a labrador in possession of an entire roast chicken.

There's an arms race on, and arms races only end in three ways. One side runs out of money, both sides hit an inescapable parity, or the race becomes irrelevant. Nobody's building better castles any more.

Big tech sees reward in AI from capturing users and their data as completely as possible. It fears its competitors doing the same. In the absence of any way to judge the quality or actual usefulness of general AI, the only way to stay in the race is faster token generation, bigger models and granderclaims. There is no enough in sight.

The cost, for now, is not a problem – if anything, it's an advantage in keeping the cost of entry sky high. Nobody is giving a moment's thought to environmental implications beyond the bare minimum of lip service expected by society at large and regulators in particular. You know how well that's going. Nobody flies off on vacation any more, right?

The driver towards sobering up AI from its fondness for the juice has to be economic. Crack that, and everything necessary will follow. The technology is here already. Look at smartphones, which contain decades of engineering effort devoted to power management at every level of the stack. That's because the commercial pressure for more performance and

more features has to be balanced by the ineluctable physics of batteries.

Datacenters have their own version of that equation, except the physics is cooling and power delivery. Until someone invents the pocket datacenter, adding more is a matter of economics, not electrochemistry. That's where big tech's AI obsession is vulnerable to sanity.

The trouble is that while that bit of the industry considers itself in an existential arms race, it will spend existential amounts of money to keep up. There is no consumer pressure to do anything else, any more than anyone knows or cares how much power that always-on smart speaker is taking. In the cogent words of the Dead Kennedys, give us convenience or give us death. A combination of the two is also acceptable.

With little limit to spending more and more on more and more energy guzzling, and no way to judge what parity even means in competing AI, that leaves the third way, to change conditions so that the race becomes irrelevant.

That's what happened in the 1990s when Microsoft twigged that online was going to be huge and the company needed to control it at any cost. It tried offering its own network, its own services, its own browser, its own standards, its own web server, each dug deep into the existing Microsoft ecosystem. It worked for Windows. It worked for Office. It didn't work online, where open standards and open code pushed evolution so quickly that Microsoft's modus operandi was simply irrelevant.

The equivalent neutralizing factor for big tech AI isn't just open source, although that's an absolute necessity. What's needed is an open source based push into all the areas where the big tech AI business model can't follow. Make joules per token a metric, and build more efficient techniques. Even those who don't care much will pick the better option if there's no effort or disadvantage to them.

Make an open ecosystem of components to match what's being hyped, but put the control over data entirely in the hands of those who use or deploy. Build tools entirely untainted by control and capture. Knock away the reasons for the arms race, and we can have as much transparency and control over power as we like.

That's a lot of effort, seemingly with little economic motivation. Fortunately, the arms race to keep control of AI can be recast as a race between control and autonomy. With fears about sovereignty over data suddenly high on the agenda everywhere, the argument for pushing AI that heads off corporate capture makes sense at a nation state level, at least outside the US. It's a very hot topic in Europe, which has a liking for big transnational projects. It's not very good at doing much with the results, though.

So how about China? That feels unthinkable for an authoritarian nation famous for considering all data everywhere the awful prey of the Chinese Communist Party. Openness and freedom seem entirely antithetical for a country you shouldn't trust with your shopping list. But it would be entirely in China's national interests to prevent an American big tech AI monopoly, even if it didn't get one in return - which isn't going to happen in any case, no matter how many switches it infiltrates.

The morality of working with China is hugely difficult, as it always is. Are cheap electric vehicles adding to that nation's reach while ruining Western industries, or an essential part of the path to net zero? Would a huge push into properly open, properly configured AI be an inexcusable attack on Western economies? You can't put tariffs on free and open standards or free and open code, and you don't have to trust that subterfuge isn't hidden within.

back in 2001, Steve Ballmer famously compared open source to cancer. That may not be the best biological analogy here. The last technology that made a big dent in global CO_2 emissions was also open source, free, drove massive global innovation, and came from China too. Even if it was invented by bats. ®