How to Accelerate Ethics for Innovation and Against Precaution in Generative AI

Keywords:
Acceleration AI ethics, Precautionary principle, innovation, safety, generative AI

1. Dilemma

The precaution principle has been applied across recent regulation of artificial intelligence, and stipulates: "If an action might cause harm to the public and if there is no scientific agreement on the issue, then the action should not be carried out." Effectively, precaution sets a default between safety and AI development, and pulls toward safety.

Against the default, there is the argument that safety can be self-defeating: because the precautionary impulse itself creates risks, the principle prohibits what it simultaneously requires. Concretely, well-intentioned guardrails limiting AI development could inadvertently prevent safety benefits from driverless cars, industrial efficiencies, medical breakthroughs. The list goes on.

Narrowing the scope to generative AI, the precautionary dilemma is afflicting companies struggling with their own trust and safety employees. While this sentence is being written, Microsoft is being scorched on academic social media for eliminating one of its responsible AI teams while expanding public access to its most powerful large language model. And, there is no reason to believe anything but that subsequent releases of ChatGPT and related platforms will do anything but intensify the conflict between risk and potential.

2. Acceleration response

This essay is about potential. It describes an innovation principle to counter the precautionary approach. This does not mean safety will be ignored, only that the burden shifts to those who believe that risk factors and safety concerns require slowing, pausing, or stopping AI innovation. Now they are the ones who need to show why: instead of

information engineers demonstrating that their work should advance despite the risks, the risk adverse must demonstrate that their concerns justify halting further advance.

Five elements – five values and attitudes – compose the ethics.

- **Uncertainty is encouraging**. The fact that we do not know where generative or linguistic AI will lead is not perceived as a warning or threat so much as an inspiration. More broadly, the unknown itself is understood as potential more than risk. This does not mean potential *for* something, it is not that advancing into the unknown might yield desirable outcomes. Instead, it is advancing into uncertainty in its purity that attracts, as though the unknown is magnetic. It follows that AI models are developed for the same reason that digital-nomads travel: *because* we do not know how we will change and be changed.
- **Innovation is intrinsically valuable**. A link forms between innovation in the technical sense and creation in the artistic sense: like art, technical innovation is worth doing and having independently of the context surrounding its emergence. Then, because innovation resembles artistic creativity in holding value before considering social implications, the ethical burden tips even before it is possible to ask where the burden lies. Since there exists the positive value of innovation even before it is possible to ask whether the innovation will lead to benefits or harms, engineers no longer need to justify starting their models, instead, others need to demonstrate reasons for stopping.
- **The only way out is through: more, faster**. When problems arise, the response is not to slow AI and veer away, it is straight on until emerging on the other side. Accelerating AI resolves the difficulties AI previously created. So, if safety fears emerge around generative AI advances like deepfakes, the response is not to limit the AI, but to increase its capabilities to detect, reveal, and warn of deepfakes. (Philosophers call this an excess economy, one where expenditure does not deplete, but increases the capacity for production. Like a brainstormed idea facilitates still more and better brainstorming and ideas, so too AI innovation does not just perpetuate itself, it naturally accelerates.
- **Decentralization**. There are overarching permissions and restrictions governing AI, but they derive from the broad community of users and their uses, instead of representing a select group's preemptory judgments. This distribution of responsibility gains significance as technological advances come too quickly for their corresponding risks to be foreseen. Oncoming unknown unknowns will require users to flag and respond, instead of depending on experts to predict and remedy.
- **Embedding**. Ethicists work inside AI and with engineers to pose questions about human values, instead of remaining outside development and emitting restrictions.

Critically, ethicists team with engineers to cite and address problems as part of the same creative process generating advances. Ethical dilemmas are flagged and engineers proceed directly to resolution design: there is no need to pause or stop innovation, only redirect it.

Together, these attitudes and practices remake ethics as accelerating rather than restraining artificial intelligence.

[In the conference paper, they will be illustrated with reports and papers exploring the innovation and safety characteristics of generative AI development.]

3. Practical reality

A practical argument supplements the theoretical foundation. Regardless of whether academics, regulators, developers, and users find the acceleration approach intellectually satisfying, it will come. It will because human nature, the inability to regulate globally, and economic necessity together produce this reality: responding to the dangers of AI innovation with restrictions and limitations means getting left behind.

OpenAI discovered this after imposing walls of conscientious restrictions on the outputs of the image generator Dall-E. The result was not more safety, just a stampede to of users – and then money – to the less restrictive Stable Diffusion. The mistake was not repeated with OpenAI's ChatGPT. The lesson learned is irrelevant, however, because even if OpenAI had learned nothing, nothing would have changed. It is just that another platform would have blazed forward to grasp the public imagination of large language models.

Inescapably, people do not like being told what they are allowed to create. That goes for developers and for users, and in a digital and electric world without borders and therefore without centralized government, the only guidelines that exist – the only ones that *can* exist – will be those enforced by the rules of innovation. Precaution today, consequently, does not mean taking a taking a stand for safety, it means opting out of AI development, and blindness to the only kind of ethics that can keep humans in control of the machines.

End