

## How Ethics Diverges between American and European AI

James Brusseau

Philosophy Department, Pace University, New York City. Mail: [jbrusseau@pace.edu](mailto:jbrusseau@pace.edu)

### 1. Innovation creates safety, or safety allow innovation?

The overarching ethical question asks about the best way to ensure innovation and safety in AI. Are innovation's accompanying problems best solved with still more innovation? Or is the best way to innovate, at least in the long term, to guarantee safety before allowing technological advance?

Starting on the American side, the acceleration approach to AI ethics means that risks deriving from AI innovation are best managed with still more and faster innovating.<sup>1</sup> For example, AI image analysis in medicine can increase efficiency and diagnostic accuracy, but it simultaneously requires extensive personal information, and that creates a privacy violation risk.<sup>2</sup> To respond, privacy-by-design strategies may inject noise into datasets, or machine unlearning may increase users' effective control over the release of their information.<sup>3</sup> Regardless, what matters on the ethical level is that these are technical innovation solutions responding to innovation risks. On the ethical level, safety understood in terms of privacy is subordinated to innovation because

the *same* innovative potential that initially created the privacy vulnerabilities surges ahead to address them.

As displayed by the AI Act, the European attitude leans the other way. Safety ethically precedes innovation and consequently the burden is on innovators to slow down their work until adequate protections can be demonstrated. Specifically around privacy, the AI Act imposes stringent requirements on data use, and system capabilities are limited until personally identifying information can be effectively protected.<sup>4</sup>

More legal requirements could be added, but what matters is the ethics underneath. As opposed to the acceleration model where innovation provides safety, the AI Act implies an ethical dynamic where safety enables trustworthy innovation. It is because privacy risks have been addressed that innovation may subsequently push ahead.

Ultimately, there is no certain answer to the question about whether innovation creates safety or safety allows innovation. More, there's no reason to think that a single answer covers the entire range of AI applications. Still, the question remains, and the answers are divisive.

### 2. Does innovation hold intrinsic value?

For acceleration ethics, innovation is worth having intrinsically. It holds value independently of its effects in society, and regardless of subsequent applications.

<sup>1</sup> J. BRUSSEAU, *Mapping ai avant-gardes in time: Posthumanism, transhumanism, genhumanism*, in *Discover Artificial Intelligence*, 3, 1, 2023, 32.

<sup>2</sup> Y. CHEN, P. ESMAELZADEH, *Generative AI in medical practice: in-depth exploration of privacy and security challenges*, in *Journal of Medical Internet Research*, 26, 2024, e53008.

<sup>3</sup> J. JIAO et al., *Navigating llm ethics: Advancements, challenges, and future directions*. 2024, arXiv preprint

arXiv:2406.18841; M. CHEN, et al., *When machine unlearning jeopardizes privacy*, in *Proceedings of the 2021 ACM SIGSAC conference on computer and communications security*, 2021, 896-911.

<sup>4</sup> G. MALGIERI, F. PASQUALE, *Licensing high-risk artificial intelligence: toward ex ante justification for a disruptive technology*, in *Computer Law & Security Review*, 52, 2024, 105899.

Technical progress, consequently, resembles artistic creation. We do not grade an artwork in terms of how it gets *used*. Our consideration of a painting is not diminished because it was commissioned by criminals who then exploited the opaque art market to launder illicit money. Ethically, the art and the crime are separate considerations. For acceleration ethics, a parallel divergence splits AI: technological creativity separates from the technology's subsequent uses. Taking the infamous case of Bostram's AI paperclip machine, it wipes out humanity while prodigiously fulfilling its operational design.<sup>5</sup> For the acceleration approach, the value of innovation subsists, even while its effects in the world are condemned.

Positing value directly in innovation tips a significant ethical burden. Since there exists positive value *before* asking whether a particular innovation will lead to benefits or harms, engineers are liberated to freely experiment. Because the value is there initially, innovation is *always* justified, the only question is whether subsequent consideration leads to more powerful counterarguments against continuing forward. Stronger, acceleration ethics means that engineers resemble artists in this way: they no longer need to justify *why* they are creating, others are responsible for explaining why they should stop.

By contrast, the ethical foundation supporting the AI Act indicates no intrinsic value for technological creation. Instead, value is found in the consequences for society. So, innovation is permitted—and even encouraged—when its downstream uses demonstrably uphold human dignity, fairness, and accountability. However,

when these outcomes are uncertain or negative, the AI Act imposes restrictions (Novelli et al. 2024).<sup>6</sup> Overall, the process is outcome-oriented in the sense that innovation must earn ethical legitimacy through its effects, and not claim it by default. It follows that technical creativity is restricted because engineers must establish the beneficence of their research before they can begin experimenting.

The American attitude toward innovation is deontological (innovation is worth having on its own), while the European approach is consequentialist (innovation is as good or bad as its effects).

### 3. Does the unknown hold value?

The third element of acceleration AI ethics is that positive value exists in the unknown. For many AI developments, there exists considerable uncertainty about how the innovation will affect the larger world. One response to the void is to find it attractive. Stated differently, uncertainty can be taken to imply potential, which does not mean potential *for* something, it is not that advancing into the unknown might yield desirable outcomes and *therefore* it is desirable. Instead, advancing into uncertainty in its purity is alluring, as though the unknown were magnetic.

It follows that AI can be developed not only to serve predetermined purposes, but also less directionally. It can be developed for the same reason that nomads travel: *because* there is no way to know what will emerge afterward (Brusseau 2021).<sup>7</sup> For both nomadic travelers and AI engineers, advances do not exist despite the

<sup>5</sup> N. BOSTROM, *Superintelligence: Paths, dangers, strategies*, Oxford, Oxford University Press, 2014.

<sup>6</sup> C. NOVELLI, et al., *Taking AI risks seriously: a new assessment model for the AI Act*, in *AI & SOCIETY*, 39, 5, 2024, 2493-2497.

<sup>7</sup> J. BRUSSEAU, *What Does It Mean To Be A Digital Nomad?*, in *Turkish Policy Quarterly*, 22., 2021, <http://turkishpolicy.com/article/1104/what-does-it-mean-to-be-a-digital-nomad> (last access 21 August 2024)



unknown, but *because* there is no knowing where changes will lead, and what they will catalyze.

By contrast, the AI Act recoils from the unknown. Legally, this translates into a regulatory structure that does not wait for empirical evidence of harm before intervening. Instead, when credible concern exists that AI *might* cause damage, early restrictions are justified.

For example, there is the AI Act's ban on real-time remote biometric identification in public spaces, such as the facial recognition used by police to scan crowds. Although comprehensive empirical evidence of harm from such systems in Europe is still limited—partly because these systems have not been widely deployed—the AI Act nonetheless preemptively prohibits their use under most circumstances (Gikay 2023).<sup>8</sup> The legal and social effects of that ban will become more evident as time passes, but what is already clear is the ethical underpinning. It is that the unknown in itself is not valuable. Uncertainty is not something inherently worth having.

#### 4. Centralized or decentralized governance?

Acceleration AI ethics decentralizes governance. Permissions and restrictions surrounding an AI tool derive from the community of users employing the technology. There are, consequently, two vectors of decentralization. First, governance comes after a tool has been deployed, as opposed to being a condition requiring fulfillment before deployment. Second, governance emerges from users working inside and with the technology, as opposed to being imposed from outside and by overseers.

One decentralization example is the Community Notes feature on X (Pilarski et al. 2024). To manage misinformation on the platform,

contributors can propose to add corrections to posts, and others vote on the note's helpfulness. If consensus is reached across a range of diverse users, the note is appended to the original post. While the governance strategy remains imperfect, it is clearly decentralized. What counts as misinformation is determined after a posting, not before, and the judges are common users, not pre-ordained overseers.

On the contrary, the AI Act supports centralization and preemption. For AI applied to hiring in the EU market, for example, a prospective tool must first undergo a conformity assessment which tests for data quality, fairness, transparency, and human oversight. Then, the accompanying regulatory enforcement is provided by national supervisory authorities in each member state, which in turn defer to the newly formed European AI Office (Thelisson and Himanshu 2024.).

While the burgeoning legal complexities will be significant, the underlying ethics remains simple. What is worth having is centralization. Power is exercised preemptively, and from above those actually using the technology.

#### 5. Summary

Neither American nor European AI is monolithic, and the rapidity of change along with the subjectivity of ethical perspectives threatens all conclusions. Still, contrasts exist, and four are summarized in Table 1.

<sup>8</sup> A.A. GIKAY, *Regulating use by law enforcement authorities of live facial recognition technology in public*

*spaces: an incremental approach*, in *The Cambridge Law Journal*, 82, 3, 2023, 414-449.

**Table 1. How Ethics Diverges between American and European AI**

How Ethics Diverges between American and European AI	
ACCELERATION	PRECAUTION
Innovation creates safety	Safety allows innovation
Innovation is intrinsically valuable	Innovation is only as valuable as its consequences
Unknown is encouraging	Unknown is discouraging
Decentralization	Centralization

