

DATA WASTE

*Elettra Bietti & Roxana Vatanparast**

This Post is the fifth in a new Frontiers series that critically explores the connection between international law and emerging technology, featuring the writing of scholars from a variety of disciplines affiliated with the Institute for Global Law and Policy (IGLP) at Harvard Law School.

There is no small irony in the fact that a company sharing a name with the largest tropical rainforest in the world is significantly contributing to environmental destruction. While Amazon has announced intentions to be carbon neutral by 2040, its platform operations contribute to significant carbon emissions.¹ Amazon also contributes to environmental harms in indirect ways through its provision of cloud-computing services to the oil and gas industry,² and its donations to political candidates denying climate change.³

Amazon's practices are just one example of the environmental impacts of data-driven technologies. They also reflect a broader phenomenon related to data-intensive commercial practices today. Empirical findings have shown that digital technologies contribute to 4% of overall

* Elettra Bietti is a Doctoral Candidate at Harvard Law School, a Kennedy-Sinclair Scholar, and an Affiliate at the Berkman Klein Center for Internet and Society. Roxana Vatanparast is a Visiting Researcher at the Institute for Global Law & Policy at Harvard Law School (IGLP) and a PhD Candidate at the University of Turin.

¹ See Carbon Footprint, AMAZON SUSTAINABILITY—US (2019), <https://sustainability.aboutamazon.com/carbon-footprint> (last visited Oct. 6, 2019).

² See Brian Merchant, *Amazon Is Aggressively Pursuing Big Oil as It Stalls Out on Clean Energy*, GIZMODO, <https://gizmodo.com/amazon-is-aggressively-pursuing-big-oil-as-it-stalls-ou-1833875828>.

³ See David McCabe & Karen Weise, *Bezos and Zuckerberg Take Their Pitches to Washington*, N.Y. TIMES, (Sept. 19, 2019), <https://www.ny-times.com/2019/09/19/business/bezos-zuckerberg-washington.html>.

greenhouse gas emissions, a number expected to double by 2025.⁴ Other estimates indicate that training a single artificial intelligence (“AI”) model emits carbon dioxide in amounts comparable to that of five cars over their lifetimes,⁵ and that bitcoin mining consumes more electricity than the country of Switzerland.⁶

While much legal scholarship relating to data collection and processing has focused on individual privacy and autonomy, far less attention has been paid to the environmental impacts of data. In this contribution, we highlight a phenomenon we call “data waste,”⁷ or the carbon emissions, natural resource extraction, production of waste, and other harmful environmental impacts directly or indirectly attributable to data-driven infrastructures. These include platform-based business models, the programming

⁴ See Maxime Efoui-Hess, *Climate Crisis: The Unsustainable Use of Online Video*, SHIFT PROJECT (2019).

⁵ See Karen Hao, *Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes*, MIT TECH. REV., <https://www.technologyreview.com/s/613630/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/>; Emma Strubell, Ananya Ganesh & Andrew McCallum, *Energy and Policy Considerations for Deep Learning in NLP*, ANN. MEETING ASS'N COMPUTATIONAL LINGUISTICS (2019).

⁶ See James Vincent, *Bitcoin Consumes More Energy Than Switzerland, According to New Estimate*, THE VERGE (July 4, 2019), <https://www.theverge.com/2019/7/4/20682109/bitcoin-energy-consumption-annual-calculation-cambridge-index-cbeci-country-comparison>.

⁷ On the wastefulness of the vast amounts of data generated on a daily basis, see Ron Bianchini, *Are Data Centers the New Global Landfill?*, WIRED (Oct. 2012), <https://www.wired.com/insights/2012/10/data-centers-new-global-landfill/>. Others have referred to the generation of data as a “digital pollutant,” a “a new kind of trash for the information age.” Tyler Elliot Bettilyon, *How Data Hoarding Is the New Threat to Privacy and Climate Change*, MEDIUM, (Aug. 1, 2019), <https://onezero.medium.com/how-data-hoarding-is-the-new-threat-to-privacy-and-climate-change-1e5a21a49494>. Nanna Bonde Thylstrup frames “digital traces” as a question of waste, which create social and environmental problems that “follow gendered and colonial structures of violence.” Nanna Bonde Thylstrup, *Data Out of Place: Toxic Traces and the Politics of Recycling*, BIG DATA & SOC., July-Dec. 2019, at 1, 2. Similarly, interdisciplinary workshops and research projects have been organized around addressing issues of waste in the context of big data. See, e.g., *Workshop Data Associations in Global Law & Policy II: Making Waste, Talking Trash*, DIGITAL HUMANITARIANISM: L. & POLY CHALLENGES (Dec. 14, 2018), <http://datapolicy.law.unsw.edu.au/resources>.

and use of AI systems, and blockchain-based technologies.

We first describe data-driven infrastructures as reflecting an ideology of permissionless innovation and as producing network effects that tend toward monopolization. We then explain how these factors have shaped data-driven infrastructures' impact on the climate. Next, we discuss how law and legal institutions have facilitated these developments and call for a more systemic understanding of the factors contributing to data waste. We end with a call to frame the issue of data waste as a sociotechnical controversy that raises broader questions about the power dynamics that underlie data-driven infrastructures and the need to imagine alternatives.

I. DATA-DRIVEN INFRASTRUCTURES

Data-driven infrastructures are here defined widely as technologies that are fueled by data and that generate data. Data in other words is both an input and an output of these technologies, in that it is both produced by such infrastructures, and what allows them to function. Examples of data-driven infrastructures include content sharing platforms such as Facebook, YouTube, Twitter, Google, and Amazon, AI systems that both rely on data for their training and functioning and also produce new data, relying on the idea that “bigger is better”,⁸ and blockchain-based systems which rely on and produce information in the form of ledgers. Data-driven infrastructures also depend on material infrastructures for storing, transporting, and processing data, such as data centers and data servers, which depend on

⁸ “Bigger is better” in this context refers to the belief that AI models that use mass computation are more accurate, and thus “better.” Roel Dobbe and Meredith Whitaker, *AI and Climate Change: How They're Connected, and What We Can Do About It*, MEDIUM (Oct. 17, 2019), <https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c>.

intensive energy⁹ and water¹⁰ usage. Data-driven infrastructures therefore exist at multiple infrastructural layers.¹¹

Data-driven infrastructures possess two specific characteristics which set them apart from most other infrastructures. The first is that they are based on logics of “permissionless innovation” and “generativity.”¹² Permissionless innovation is the principle according to which a given infrastructure allows others to innovate on top of it without asking those who develop or control the infrastructure for permission to do so. Generativity is the capacity to build upon an existing system in ways unforeseen by the original designers.¹³ The second characteristic of data-driven infrastructures is that they facilitate the monopolization of digital markets. These infrastructures enable and incentivize business models based on connecting increasing numbers of people, capturing their attention, and using their data for profit. Data-driven infrastructures give rise to network effects; their functionality improves with the greater number of people using them and the greater amount of data collected through them, driving ever-increasing data collection. The success of these business models thus depends on market power concentration, achieved by ousting competitors for valuable user data and attention, and making other businesses dependent on their essential

⁹ See SEAN CUBITT, *FINITE MEDIA: ENVIRONMENTAL IMPLICATIONS OF DIGITAL TECHNOLOGIES* 16–21 (2017).

¹⁰ See Mél Hogan, *Data Flows and Water Woes: The Utah Data Center*, *BIG DATA & SOC.*, July-Dec. 2015, at 1.

¹¹ See Yochai Benkler, *From Consumers to Users: Shifting the Deeper Structures of Regulation Toward Sustainable Commons and User Access*, 52 *FED. COMM. L. J.* 561 (2000).

¹² See generally JONATHAN ZITTRAIN, *THE FUTURE OF THE INTERNET AND HOW TO STOP IT* (2008).

¹³ See *id.*; see also Jonathan Zittrain, *The Generative Internet*, 119 *HARV. L. REV.* 1974 (2006).

infrastructure.¹⁴ In what follows, we explore these two aspects in turn, showing how they have contributed to data waste.

II. “PERMISSIONLESS” DATA WASTE

The way the Internet was initially developed, including its layered end-to-end networked structure, enabled innovation to take place at the user-interface level without interference from those who control the physical backbones of the internet, including cables and transmission channels.¹⁵ This led to a vision of the Internet in its early stages as an open ecosystem for permissionless innovation—a site where new entrants were able to develop projects and initiatives that were thought to significantly challenge the status quo, through peer-to-peer networks and collaborative platforms. A similar principle applies to blockchain technologies which rely on an open distributed ledger managed by a peer-to-peer network of distributed nodes. Behind the appearance of a frictionless infrastructure, however, these and other permissionless technologies hide a darker side. They have enabled a culture of unreflective innovation which, if left unchecked, is too often blind to long-term systemic effects. A culture of “moving fast and breaking things,” as expressed in the famous Facebook slogan,¹⁶ can lead to negative externalities—including the harms on society and the environment imposed by data waste—while responsibility for those harms remains diffuse.

¹⁴ See Lina M. Khan, *Amazon’s Antitrust Paradox*, 126 YALE L.J. 710 (2017).

¹⁵ See YOCHAI BENKLER, *THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM* 59 (2006).

¹⁶ See, e.g., Hemant Taneja, *The Era of “Move Fast and Break Things” Is Over*, HARV. BUS. REV. (Jan 22, 2019), <https://hbr.org/2019/01/the-era-of-move-fast-and-break-things-is-over>; JONATHAN TAPLIN, *MOVE FAST AND BREAK THINGS: HOW FACEBOOK, GOOGLE, AND AMAZON CORNERED CULTURE AND UNDERMINED DEMOCRACY* (2017).

III. NETWORK EFFECTS AND MARKET POWER

A second aspect of data-driven infrastructures is their network effects. Network effects are the positive effects that adding a new user of a good or service can have on the value of the good or service itself.¹⁷ For instance, the more individuals use Facebook, the more attractive it becomes to join and spend time on the site. Similarly, the more searches are performed on Google Search, the more accurate their search algorithms become. Network effects go hand-in-hand with so called multi-sided markets. These are markets where two complementary services become easier to offer through the intermediation function of a platform. For example, Amazon makes it less costly for buyers to find sellers, and also cheaper for sellers to find buyers. Similarly, on Google Search, advertisers and websites are a source of content for users, and users the source of attention that websites and advertisers seek. Google stands to gain from the transactions between websites/advertisers and users because it obtains both users' data and advertisers' money.

Network effects in multi-sided data-driven markets can lead to monopolization, as corporations increasingly compete for the data and scarce attention of individuals from which they can generate revenues and profits. Such monopolization leads to an attention-based economy.¹⁸ In the attention economy, platforms try to outcompete each other by offering users appealing content, capturing their attention, collecting their data, and locking them into the habit of using their platform instead of another. The more data that a company can collect, the more successful they will be at locking-in

¹⁷ See generally CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY (1999).

¹⁸ See, e.g., TIM WU, THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS (2016).

users and driving out competitors.¹⁹

The combination of permissionless innovation and network effects that concentrate market power is driving ever-increasing data processing and storage, content (and particularly video) streaming in real-time, and complex data analytics systems—all of which require intensive energy usage.²⁰ To compete and generate revenues, tech giants such as Google and Facebook fuel an aggressive culture of constant virtual connection for all—a culture largely powered through video content, as made salient by the success of platforms such as YouTube, Instagram, and Twitch—one that may also be advancing in the Global South through cheap mobile devices and phone contracts.²¹ The constant connection inherent to these business models comes with a substantial increase in data collection and waste.

Big tech companies' market power also has other environmentally detrimental effects. Consider Amazon. Amazon Web Services is the largest provider of web storage and cloud services worldwide, far ahead of its largest competitors: Google, Microsoft, and IBM. Not only does the internet largely run on Amazon servers, but Amazon also

¹⁹ It has been argued extensively, for example, that users' consent in digital environments is illusory, because platforms and other digital actors own infrastructures and impose the terms of their relationship on users without users' ability to opt-out and look elsewhere. See, e.g., Julie E. Cohen, *Turning Privacy Inside Out*, 20 THEORETICAL INQUIRIES L. (2019); Elettra Bietti, *Consent as a Free Pass: Platform Power and the Limits of the Informational Turn*, 40 PACE L. REV. 310 (2020). Further, once infrastructure owners have data about individuals, they can leverage that data to nudge and behaviorally manipulate users to remain on their platform. See, e.g., MIREILLE HILDEBRANDT, *SMART TECHNOLOGIES AND THE END(S) OF LAW: NOVEL ENTANGLEMENTS OF LAW AND TECHNOLOGY* (2015); SHOSHANA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER* (2019).

²⁰ See CUBITT, *supra* note 9.

²¹ *Buying a Smart Phone on the Cheap? Privacy Might Be the Price You Have To Pay*, PRIVACY INT'L (Sept. 20, 2019), <https://privacyinternational.org/long-read/3226/buying-smart-phone-cheap-privacy-might-be-price-you-have-pay>.

provides cloud-computing services to the oil and gas industry, helping them extract fossil fuels more efficiently.²² Amazon also donates to political candidates that deny climate change, thereby shaping public discourse on the issue.²³ Amazon's example shows how network effects and concentrated market power are tied to environmental harms, as well as the power to shape public debates on climate change through political donations. Thus, data-driven infrastructures, market power, and the political discourse around environmental issues cannot be disentangled from data waste.

IV. THE ROLE OF LAW

The early history of permissionless innovation is linked to the ideology of "lawlessness," or the idea that innovation flourishes in the absence of regulatory constraints and that there exists an independent lawless space called cyberspace.²⁴ Indeed, in the United States, for instance, the dominant discourse in the tech industry is that self-regulation and self-governance promote innovation. This might be attributable to the assumption that law perpetually lags behind innovation, and therefore law should take a reactive role and defer to the authority of technologists in order not to hinder innovation.²⁵

This idea, of course, reflects a limited view on the interaction between law, technology, and society. Cyberspace and data-driven infrastructures are and have always been far

²² See Merchant, *supra* note 2.

²³ See McCabe and Weise, *supra* note 3.

²⁴ See John Perry Barlow, *A Declaration of the Independence of Cyberspace* (1996), <https://www.eff.org/cyberspace-independence>.

²⁵ This idea refers to the "law lag" narrative prevalent in dominant discourses on law's relationship to science and technology, which fails to take into account their interactive relationship in ordering society. See Sheila Jasanoff, *Making Order: Law and Science in Action*, in *THE HANDBOOK OF SCIENCE AND TECHNOLOGY STUDIES* 761–786, (Edward J. Hackett et al. eds., 3rd ed. 2008).

from lawless. Lawyers use private law mechanisms to code assets into capital.²⁶ Tech companies, as owners of data-driven infrastructures, use these mechanisms to create and protect the wealth they obtain through data markets.²⁷ At the same time, legal rights, privileges, and immunities have assured the reproducibility of economic and political power through information technologies.²⁸ Changes in U.S. antitrust doctrine, and the challenges brought by technology corporations that no longer fit within the “consumer welfare” antitrust model based on short-term price effects, have also contributed to the concentration of market power of companies like Amazon.²⁹ Moreover, in the United States, campaign finance laws protect corporate financing of political candidates based on freedom of speech principles,³⁰ giving tech companies opportunities to shape public discourse and policy. At a more fundamental level, law affects the relative bargaining power between technology corporations and users,³¹ which facilitates tech corporations’ endless accumulation of data from users and determines how gains from the data economy are distributed. Finally, international environmental law’s role in managing global waste and the externalization of waste by market mechanisms shapes the global distributive effects of data waste.³² To say that data-driven infrastructures are self-regulated or that data waste is a result of lawlessness would therefore overlook a wide variety of legal mechanisms and institutions that have contributed to the phenomenon, and the broader legal

²⁶ See generally KATHARINA PISTOR, *THE CODE OF CAPITAL: HOW THE LAW CREATES WEALTH AND INEQUALITY* (2019).

²⁷ See generally *id.*; JULIE E. COHEN, *BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM* (2019).

²⁸ See COHEN, *supra* note 27.

²⁹ See Khan, *supra* note 14.

³⁰ See *Citizens United v. Federal Election Commission*, 558 U.S. 310 (2010).

³¹ See Duncan Kennedy, *The Stakes of Law, or Hale and Foucault!*, 15 *LEGAL STUD.* F. 327 (1991).

³² See Olivier Barsalou & Michael Hennessy Picard, *International Environmental Law in an Era of Globalized Waste*, 17 *CHINESE J. INT’L L.* 887 (2018).

arrangements which form part of the social fabric which technology embeds and in which it is embedded.³³

V. DATA WASTE AS SOCIOTECHNICAL CONTROVERSY

The issues surrounding data waste reflect a sociotechnical controversy. The problem is not just how to rein in the harmful environmental effects of data processing by creating more so-called “sustainable tech” or “green tech,” or by creating newer and better legal and policy frameworks around these particular environmental harms. Rather, as a sociotechnical controversy, it raises broader questions.

Calls for technical fixes, such as moving to greener technologies to address climate change, reflect the same visions of social progress that are embedded in the very technologies that are contributing to harm in the first place.³⁴ The vision that technology can fix social and political problems without the need for social and political engagement is deeply flawed. What is needed instead as a first step toward political engagement on these issues is a systemic understanding of how data-driven infrastructures, the social and legal constructs that contribute to their flourishing, and the people building and designing them are collectively shaping futures for humanity and ecology. Data waste, at its core, is a democratic and distributional problem. It requires engagement not only from experts and policymakers, but also democratic publics. This political engagement should also consider the ways in which different ownership, production, and distribution models might impact data waste.

³³ See Jasanoff, *supra* note 25; STATES OF KNOWLEDGE: THE CO-PRODUCTION OF SCIENCE AND THE SOCIAL ORDER, (Sheila Jasanoff ed., 2004).

³⁴ See STATES OF KNOWLEDGE, *supra* note 33.

VI. CONCLUSION

While many questions remain to be explored in future research on data waste, this contribution serves as a call to consider not only the environmental harms resulting from data-driven infrastructures, but also broader questions about the forms of social life we would like to promote. The technological and economic powers associated with data-driven infrastructures are also political powers. The issues they raise will not be solved through technical fixes. We must ask whether we want technology companies, data-driven infrastructures, and the people behind them to have the power to shape the social and ecological conditions for our futures, and if not, who ought to be exercising such power, and what role law and public political engagement can play in shaping alternatives.