



This working group will address how blockchain technology can help resolve Big Data issues as they currently exist and into the future. The purpose of this working group is to collaborate, propose, and implement blockchain solutions to the expanding Big Data issues in order to permit both governments and societies to improve their functions through systems that handle the massive data (which is projected to be about 44 zettabytes and 50 billion nodes [I.P. Addresses] by 2020). To that end, we will explore, experiment, and test innovative technology opportunities. We will hold lectures with thought leaders, technologists, and government leaders. The following is a primer which is intended to frame the terms we will encounter most often.

### A. What is a Blockchain?

A blockchain is a growing list of cryptographically inter-related electronic ledgers called “blocks”<sup>1</sup>. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (generally represented as a merkle tree root hash). Blockchain was invented by Satoshi Nakamoto in 2008 to serve as the public transaction ledger for the cryptocurrency he named “Bitcoin”. The invention of the blockchain for Bitcoin made it the first digital currency to solve the double-spending problem without the need for a trusted authority or central server.

For that reason, blockchain is said to be “decentralized” because there is no one physical place, network server, or stand-alone computer which is responsible for maintaining and verifying the cryptography. In contrast, search engines like Google and social media platforms like Facebook operate from a centralized networked server. Being *networks* open to public use via the internet, these platforms may well “look and feel” decentralized, but they are not. It is *crucial* to distinguish a network from a decentralized publicly distributed blockchain ledger.

By design, a blockchain is resistant to modification of the data it contains and controls. It is “an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way”. A blockchain is typically managed by a peer-to-peer (“P2P”) network collectively adhering to a protocol for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks; yet this requires the consensus of the network majority.

New blocks are disseminated by a process known as “mining.” Mining adds transaction records to a public ledger of past transactions. This serves to validate all transactions to the decentralized network. Blockchain nodes can thus distinguish legitimate transactions from illegitimate attempts to modify the blockchain in a way that is inconsistent with the publicly distributed ledger. Blockchain mining concretely sets the history of blockchain transactions in a way that is computationally impractical to modify by any one central entity. A blockchain is thus “autonomous” because it is able to reach a consensus about the order of its own cryptography.

---

<sup>1</sup><https://en.wikipedia.org/wiki/Blockchain>

## B. What is Big Data and What are the Issues Related to Big Data?

A report<sup>2</sup> from McKinsey & Co. stated that by 2009, companies with more than 1,000 employees already had more than 200 terabytes of data of their customer's lives stored in some form of digital electronic database. In contrast, we should consider the startling amount of additional stored data which the rapid growth of *social* media platforms has since accumulated. There are trillions of tweets, billions of Facebook likes, and other social media sites like Snapchat, Instagram, and Pinterest are only adding more to this social media data deluge.

In an economic vacuum, social media accelerates innovation, drives cost savings, and strengthens brands through mass collaboration. Across every industry, companies rely on social media platforms to market and hype up their services and products, along with monitoring what the audience is saying about their brand. The convergence of social media and Big Data has prompted technology driven by profit but powered by algorithms and artificial intelligence. This is now collectively called “surveillance<sup>3</sup> capitalism.” Big data abuse is blamed for creating social and professional epistemic<sup>4</sup> bubbles and eco-chambers which have caused damages to society.

Data is useful for capitalism. That's not new. What's new is the scale and significance of Big Data, all thanks to, or unfortunately from, breakthroughs in information technology<sup>5</sup>. Technology giants have become so influential that they now function like transnational governments charting the future to a greater degree than any classical national government had previously accomplished. Facebook and Google, for example, have effectively become central mediators<sup>6</sup> unilaterally determining the balance between free speech and election manipulation for all major developed democracies. At the same time as the widespread decline in the agency of market participants, rhetoric from the technology sector suggests a coming wave of underemployment due to artificial intelligence (AI) and automation of a cybernetic<sup>7</sup> variety.

The fear of a future in which people are increasingly treated as valueless and devoid of economic agency has elevated the ambitions of universal basic income advocates. Their rhetoric leaves room for only two outcomes: 1. Either there will be mass poverty despite technological advances, or 2. Much wealth will have to be taken under central, national control through a social wealth fund to provide citizens a universal basic income. Yet both dramatic inequality and what we might call “fully automated luxury communism” are dystopias that hyper-concentrate power and undermine or ignore the value of data creators (in the broadest sense) in a way similar to how the market's value of “women's work” in the home has long been ignored and debased.

---

<sup>2</sup><https://www.simplilearn.com/how-facebook-is-using-big-data-article>

<sup>3</sup>[https://en.wikipedia.org/wiki/Surveillance\\_capitalism](https://en.wikipedia.org/wiki/Surveillance_capitalism)

<sup>4</sup><https://philpapers.org/rec/NGUECA>

<sup>5</sup><https://logicmag.io/04-the-data-is-ours>

<sup>6</sup>Previously posted link went down; see attached copy (pp. 5 – 22).

<sup>7</sup><https://en.wikipedia.org/wiki/Sociocybernetics>

### C. What Exactly is the Internet, and What is the *Legal Nature* of the Internet?

The history of the Internet begins with the development of electronic computers in the 1950s. Initial concepts of wide area networking originated in several computer science laboratories in the United States, United Kingdom, and France. The U.S. Department of Defense awarded contracts as early as the 1960s, including for the development of the ARPANET<sup>8</sup> project. The Internet protocol suite, TCP/IP which is still in widespread use today, was developed in the 1970s and became the standard networking protocol on the ARPANET.

In the 1980s, research at CERN in Switzerland resulted in the World Wide Web (also commonly referred to as the Internet), by linking hypertext documents into an information system accessible from any node on the network. Since the mid-1990s, the Internet has had a revolutionary impact on culture, commerce, and technology, including the rise of near-instant communication by electronic mail, instant messaging, voice over Internet Protocol (VoIP), interactive phone and video calls, forums, blogs, social networking, and online shopping sites.

In July 2006, the Secretary-General of the United Nations announced the establishment of the Internet Governance Forum<sup>9</sup>, a multi-stakeholder forum for policy dialogue on issues of Internet governance. The IGF brings together all stakeholders in the Internet governance debate, whether they represent governments, the private sector or civil society, including the technical and academic community, on an equal basis and through an open and inclusive process. To this day, the *internet itself* remains unregulated<sup>10</sup> while nation States attempt to regulate *the activities* of Tech Giants whose work is broadcasted on the internet. These “external regulations” ignore the need for internal<sup>11</sup> Constitutional laws<sup>12</sup> which stem from, and are thus native to the Internet.

### D. What is a Government, and who are the Societies that Governments Govern?

Before the advent of computers in the 1950’s, governments were simply ideologies in control of sovereign and autonomous societies such as democracy, socialism, and communism, all of which were limited and defined by *physical* borders and *natural* resources, the collective whole of which is called a country or a nation<sup>13</sup> State. The United Nations divides the 206 listed nation States into three categories: 193 member States, 2 observer States, and 11 other States.

Compiling such a U.N. list is controversial as there is *no binding definition* on all the communities of all nations concerning *the criteria for statehood*. The U.N. list includes entities

---

<sup>8</sup>[https://en.wikipedia.org/wiki/History\\_of\\_the\\_Internet](https://en.wikipedia.org/wiki/History_of_the_Internet)

<sup>9</sup>[https://en.wikipedia.org/wiki/Internet\\_Governance\\_Forum](https://en.wikipedia.org/wiki/Internet_Governance_Forum)

<sup>10</sup><https://www.theguardian.com/technology/2017/jul/02/is-it-time-to-rein-in-the-power-of-the-internet-regulation>

<sup>11</sup><https://community.humanetech.com/t/sociocybernetics-and-the-civics-law-of-cyberspace-on-the-internet/2746>

<sup>12</sup><http://airofusion.com/ghi>; also see <https://www.eff.org/cyberspace-independence>

<sup>13</sup>[https://en.wikipedia.org/wiki/List\\_of\\_sovereign\\_states](https://en.wikipedia.org/wiki/List_of_sovereign_states)

recognized to have a de facto status as sovereign states, and inclusion should not be seen as an endorsement of any specific claim to *statehood in legal terms*. Consider further that the Republic of Estonia is on the U.N. list only once and only with reference to its *physical* borders and *natural* resources. Yet Estonia claims the existence of a “digital nation” of which Pope Francis is a citizen<sup>14</sup>. This separate “digital nation” is not on the U.N.’s list. The logic is that if Estonia’s “digital nation” is not separate from Estonia itself, then how can the Pope be a remote citizen?

The use of computers (and all their networked physical devices, collectively called “the Internet of Things<sup>15</sup>”) to plan and manage the economy of a nation State is not new. The Soviets briefly experimented with such cybernetics<sup>16</sup> in the 1960s, Salvador Allende’s Chile explored it in the 1970s with cyberneticist<sup>17</sup> Stafford Beer’s project Cybersyn<sup>18</sup>, and since the 1990s western leftists<sup>19</sup> have been concerned about how the law of all nation States impacts such technology.

It’s not just Estonia who claims to have established an autonomous and sovereign ecosphere on the Internet; non-profit block chain organizations<sup>20</sup> (without an underlying nation State structure, such as a 501(c)(3)), have also appeared on the Internet. The country of China has started ranking its citizens with a creepy “social credit” system<sup>21</sup> which has been criticized as a type of dystopian digital<sup>22</sup> authoritarian prison. Journalist Fareed Zakaria reported<sup>23</sup> on how the European Union has fined Google billions of dollars and the country of India has its own static biometric identification system<sup>24</sup> whose low tech equivalence is a national telephone book.

Static biometric ID systems rely on fingerprints, facial scans, retina scans, and other static data to identify a person. Static biometrics differs from behavioral biometrics. The latter are controlled by algorithms designed to identify a person’s behavior which can vary according to the type of exposure to digital stimulus. The behavioral biometric algorithms of Facebook et seq. are at the center of legal disputes concerning the privacy rights<sup>25</sup> of individuals in California.

---

<sup>14</sup><https://medium.com/e-residency-blog/welcome-to-our-digital-nation-your-holiness-pope-francis-4df91a6e4336>

<sup>15</sup>[https://en.wikipedia.org/wiki/Internet\\_of\\_things](https://en.wikipedia.org/wiki/Internet_of_things)

<sup>16</sup>[https://en.wikipedia.org/wiki/Suppressed\\_research\\_in\\_the\\_Soviet\\_Union](https://en.wikipedia.org/wiki/Suppressed_research_in_the_Soviet_Union)

<sup>17</sup><http://www.cybsoc.org>

<sup>18</sup><https://www.nytimes.com/2008/03/28/world/americas/28cybersyn.html>

<sup>19</sup><https://www.eff.org/about/board/john-gilmore>

<sup>20</sup><https://virtualcommodities.org>

<sup>21</sup><https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4>

<sup>22</sup>[https://www.youtube.com/watch?v=eViswN602\\_k](https://www.youtube.com/watch?v=eViswN602_k)

<sup>23</sup><https://www.youtube.com/watch?v=Zar-RCyNTMU>

<sup>24</sup><https://en.wikipedia.org/wiki/Aadhaar>

<sup>25</sup><https://www.consumerwatchdog.org/privacy-technology/privacy-bills-progress-california-senate-victory-consumers-consumer-watchdog>

# A Blueprint for a Better Digital Society

- Jaron Lanier
- E. Glen Weyl

September 26, 2018

Digital transformation is remaking the human world, but few are satisfied with how that's been going. That's especially true in media, where the dominant model of targeted advertising derived from data surveillance and used to fund free-to-the-public services like social media and search is increasingly viewed as unsustainable and undesirable.



**Illustration by Michael McQuaid**

Today, internet giants finance contact between people by charging third parties who wish to influence those who are connecting. The result is an internet — and, indeed, a society — built on injected manipulation instead of consensual discourse. A system optimized for influencing unwitting people has flooded the digital world with perverse incentives that lead to violations of privacy, manipulated elections, personal anxiety, and social strife.

It has also made many of the largest tech companies immensely powerful. A classic example of online behemoth power, what we call a “siren server,” is YouTube, owned by Google. The network effects that always accompany digital entities allow YouTube to control both the production and the consumption of digital video. They are at once a monopoly and a monopsony (a sole purchaser of data), deciding which content producers will be paid, in the manner of a communist central planner, and determining what content billions of users will consume.

Tech giants have become so influential that they function like transnational governments charting the future to a greater degree than any national government. Facebook and Google, for example, have effectively become central mediators unilaterally determining the balance between free speech and election manipulation for all major developed democracies.

At the same time as the widespread decline in the agency of market participants, rhetoric from the tech sector suggests a coming wave of underemployment due to artificial intelligence (AI) and automation. The fear of a future in which people are increasingly treated as valueless and devoid of economic agency has elevated the ambitions of universal basic income advocates. Their rhetoric leaves room for only two outcomes: Either there will be mass poverty despite technological advances, or much wealth will have to be taken under central, national control through a social wealth fund to provide citizens a universal basic income. Yet both dramatic inequality and what we might call “fully automated luxury communism” are dystopias that hyper-concentrate power and undermine or ignore the value of data creators in a way similar to how the market value of “women’s work” in the home has long been ignored and debased.

As we wait helplessly for more elections to be compromised, for more nasty social divisions to be enflamed, for more invasive data surveillance, and for more workers to become insecure, the widespread assumption that no other models are possible leads to a state of despair.

But there is an alternative: an emerging class of business models in which internet users are also the customers and the sellers. Data creators directly trade on the value of their data in an information-centric future economy. Direct buying and selling of information-based value between primary parties could replace the selling of surveillance and persuasion to third parties. Platforms would not shrivel in this economy; rather, they would thrive and grow dramatically, although their profit margins would likely fall as more value was returned to data creators. Most important, a market for data would restore dignity to data creators, who would become central to a dignified information economy.

These models have been discussed widely for years. Here, we describe a future based on them by exploring the business and societal structures that will be required to bring them to life. In the process, we will advocate for a more coherent marketplace. Without one, no corrective measure stands a chance.

## Data Dignity

A coherent marketplace is a true market economy coupled with a diverse, open society online. People will be paid for their data and will pay for services that require data from others. Individuals’ attention will be guided by their self-defined interests rather than by manipulative platforms beholden to advertisers or other third parties. Platforms will receive higher-quality data with which to train their machine learning systems and thus will be able to earn greater revenue selling higher-quality services to businesses and individuals to boost their productivity. The quality of services will be judged and valued by users in a marketplace instead of by third parties who wish to influence users. An open market will become more aligned with an open society when the customer and the user are the same person.

Glen has called this idea of a true market economy for information “data as labor” and “liberal radicalism,” while Jaron has called it “humanistic digital economics” and “entrepreneurial democracy.” Here we’ll use the less politically charged term “data dignity.” This translates the concept of human dignity that was central to defeating the totalitarianisms of the twentieth century to our contemporary context in which our data needs to be protected from new concentrations of power.

We understand the term “data” to include most digital activity. It is intentionally created entertainment data, like a YouTube video or a social media meme, as well as less deliberately produced data gathered through surveillance or biological sensors, such as location or metabolic logs. Other examples are language provided to a translation engine to train software, and real-time data flows such as a music lesson delivered over Skype.

All of this has a value to the producer, and when the producer gains control over that value, incentives will be transformed; a market participant will try to persuade the buyer to spend money with them instead of paying monopolistic platforms to manipulate a targeted person.

For instance, automated language translation services have challenged the employment prospects of professional human translators, yet these services require a vast amount of fresh data every day from the people being put out of work (to keep up with current events, pop culture, and so on). Translators might think that they’re voluntarily subtitling foreign films for online friends; they have no idea of the extent of the value they are providing. Once the people providing this data are honestly informed that they are needed, they will earn compensation for their service, take pride in providing better data, and help the automated services to function better.

The entire architecture of the digital world will gradually become clearer and less sneaky. The result we hope for is emphatically not utopian, but we believe this solution is the only viable one yet articulated for the problem of excessive, erratic, and unsustainable power concentration on digital networks.

## MIDs

The foremost challenge in implementing data dignity is the yawning gap between big tech platforms and the individuals they harvest data from. If we asked big tech alone to make the change, it would fail: Too many conflicts of interest exist, and the inevitable concentration of power these platforms create is inimical to competitive markets and an open society. Nor can individuals demand data dignity on their own, even by petitioning governments for action, because network effects have given platforms disproportionate power, and the complexity of the digital economy makes it impossible to regulate in detail.

For data dignity to work, we need an additional layer of organizations of intermediate size to bridge the gap. We call these organizations “mediators of individual data,” or MIDs. A MID is a group of volunteers with its own rules that represents its members in a wide range of ways. It will negotiate data royalties or wages, to bring the power of collective bargaining to the people who are the sources of valuable data. It will also promote standards and build a brand based on the unique quality and identity of the data producers they represent. MIDs will often perform routine accounting, legal, and payment duties but might also engage in training and coaching. They will help focus the scarce attention of their members in the interest of those members rather than for an ulterior motive, such as targeted advertising.

The concept of MIDs is not terribly revolutionary. Entities of their shape and necessity in the physical world could hardly be more familiar. Organizations like corporations, labor and consumer unions, farmers’ cooperatives, universities, mutual funds, insurance pools, guilds, partnerships, publishers, professional societies, and even sports teams are all critical to dignified societies and effectively serve the MID function.

Some of the most important thinkers about democratic market societies have emphasized the need for precisely these kinds of organizations. Alexis de Tocqueville observed that community organizations were critical to sustaining liberty in the United States. Beatrice and Sidney Webb argued that labor unions were critical to making large corporations operate effectively as they gave productive workers a voice. Hannah Arendt highlighted that it was the extreme individualism and collapse of social institutions in the interwar years that paved the way for the rise of totalitarianism in the 1930s.

Some of the models for MIDs have been traditionally more associated with the political left, while others have traditionally found sympathy with the right. In an advanced information economy, that distinction will be less important. If we are to use the language of the left: Some MIDs will, like traditional artisans’ guilds, redistribute the successes of the greatest stars and broadly share earned revenue. Others, like artist royalty collection agencies such as ASCAP, may allow a broader range of individual payouts. But we can also use the language of the right: Some MIDs might be hard to join, analogous to becoming a partner in a prestigious law firm.

MIDs will reverse current trends in the information economy, where from the early days platforms tended to profess an ideology of extreme individualism, which, echoing Arendt’s warnings about the totalitarian consequences of extreme individualism, tragically paved the way for the rise of increasingly concentrated platform power. The slogan “Move fast and break things” from the early days of Facebook, for example, meant in practice the weakening of pre-internet MIDs such as publishers and unions for creative professionals.

Excessively concentrated power was not the only problem. The societal structures that were broken were supplanted by algorithms that target people for advertising; these

tend to corral individuals into divergent groups. Incentives to increase online “engagement” can then result in heightened social rifts as suspicions are raised about the “other.” This tendency of the current network architecture is so prominent that it has become a favorite tool for information warfare; both rich and poor societies have been disrupted by malicious social media campaigns that emphasize and encourage societal divisions.

Self-organizing MIDs will give rise to different incentives. Individuals will have memberships in many separate MIDs. While MIDs will compete, individuals will have tangled allegiances. An analogy from the pre-internet world is that two people might work for competing stores but attend the same church, or might choose different car insurance companies while investing in the same mutual fund. In the same way, individuals will seek memberships in many MIDs rather than a few platforms; the result will be complex identities and interests instead of managed, corralled identities that are ripe for targeting.

Since we first started talking about the idea of MIDs a few years ago, we’ve received thousands of unsolicited queries from entrepreneurs attempting to launch MIDs of their own. It appears that billions of dollars have already been invested, though that investment is scattered. Most of the communication with us has come from tech startups, but non-governmental organizations, labor unions, nonprofits, corporate initiatives within existing tech companies, and possible new government agencies have all been proposed.

We will not discuss specific proposals here, but we can describe the trends we see in them. Some of the classes of proposals include the following:

- Entrepreneurs seeking to create groups based on common interests to negotiate a fair price for access to that group’s data. Two that come to mind are medical patients with a similar condition, and language translators whose work informs automated translation engines. We are skeptical of many such efforts, however, as they plan to derive all funding from data sales, creating incentives similar to existing data brokers to abuse member privacy and trust.
- Technically focused groups that want to implement tools like blockchain to manage data provenance, access, and flow — the first step in managing its value. We are skeptical that the extreme decentralization emphasized by these projects will offer the bargaining power or informational security required to obtain a fair deal for data creators.
- Groups that hope to counter large tech companies by becoming large and powerful enough to operate like the tech companies — running a nonprofit social network platform, for instance — but with a pledge to acknowledge and respect members. Call them “enlightened siren servers.” We are skeptical of these because we don’t think any siren server can escape perverse incentives without MIDs.
- Collectives that hope to replace gig economy platforms like Uber with similar operations owned by those who do the work and who would get paid for the data they generate. We are concerned that these efforts exclude most data creators who do not currently

view themselves as working in the gig economy and thus are unlikely to recruit the critical allies necessary to create a broad-based social movement.

- Champions of a gift-giving or patronage economy who want to grow that market exponentially, to the point that most human sustenance would be derived from it. We worry that historical experience has shown that, beyond relatively narrow and exclusive communities, gift economies become chaotic and are often dominated by powerful agents who take advantage of others' generosity.

## Principles for MIDs

It's gratifying to see the high level of interest and activity around MIDs so far. Unfortunately, while some of the proposals are creative and substantial, nearly all of them thus far fall short on crucial dimensions. It has become clear that for MID initiatives to succeed, those building them need clearer guidance on how to structure them. Here, we present eight principles or requirements to give those starting MIDs a way to test their designs against a reasonable guess at future requirements.

**1. Fiduciary duty.** A MID should be a true fiduciary for individuals who create data or from whom data is measured, in a legal, economic, and structural sense.

Legally, MIDs should have an exclusive and overriding fiduciary responsibility to serve the true best interests of data creators, even when these creators do not necessarily fully understand their best interests.

Economically, a MID should be funded in a way that avoids conflicts of interest that make it impossible to serve as a true fiduciary. For example, it should not be purely funded by fees proportional to the volume of data exchanged, as this would compromise its incentive to protect privacy in the best interests of its members.

Structurally, a MID should be separate from other organizations with naturally opposite interests, such as data consumers. One particularly attractive structure, though by no means the only desirable one, would be a mission-driven nonprofit or data worker cooperative. Organizations in this mold (under the banner "data union") have formed in both Europe and the United States.

Siren servers cannot be fiduciaries; they serve too many masters simultaneously, just as financial advisers who are part of a mutual fund family are motivated to push associated funds that may not be in the best interest of the customer. This type of conflict of interest not only exists on the current internet, it's foundational. It must be reformed. It's simply unrealistic for a business funded primarily by advertisements (Google, Facebook) to be a fiduciary for the targets of that advertising. In highly opaque domains like data governance or finance, dedicated, independent, unconflicted fiduciaries are critical. This role cannot be covered by a world-spanning, centralized fiduciary, any more than a single lawyer can represent both sides in a legal dispute.

It is also unrealistic to expect regulators to micromanage these conflicts through regulation. Indeed, when tech executives testify before government bodies, they typically make a show of how politicians and regulators can't keep up with fast-changing technology well enough to understand it. And yet, precisely because of the complexity, dynamism, technical sophistication, and psychological potency of the modern digital experience, it is essential that individuals have access to representatives and advisers with fiduciary duties. MIDs can be those fiduciaries. Far from suffering conflicts of interest, they encourage competition and will represent opposing interests and philosophies.

Without a fiduciary to check them, siren servers are fated to take on a dystopian perspective of controlling society from above. This is seen in the Chinese Social Credit System, but is found equally in certain American platform thinking, such as the notorious Google concept of the "Selfish Ledger," in which users are described as passive servants of Google's true client, their data.

**2. Quality standards.** MIDs will foster decency, high standards, accountability, and acknowledged achievement in terms that they will largely define themselves.

Consider the problem of what we have come to refer to as fake news. It is impossible for a siren server to select a preferred set of news sources that adhere to standards, because that would be viewed as unfair. One idea in vogue to combat fake news is for Twitter, Google, or Facebook to use crowdsourcing or a large number of low-paid workers to demote or annotate certain classes of undesirable speech. But there is no known way to do this without displeasing some interests, often powerful ones. Certain politicians, for instance, notoriously disagree about which news is fake.

More broadly, platforms have agreed to demands that they attempt to restrict obscenity, gore, incitement, cruelty, and so on. However, it is impossible to come up with one set of standards that please billions of people; even the most well-meaning siren server cannot appease everyone about everything.

Governments can enforce rules that really do benefit everyone, such as criminal codes and food safety regulations, but as soon as rules exceed the bounds of acknowledged universal necessity, enforcement becomes authoritarian. When critics demand that a platform like Facebook ban certain forms of speech, they also make it more authoritarian, just as a government that demands that people be polite must be authoritarian. The root problem is that siren servers have a dysfunctional excess of concentrated power. MIDs will distribute that power and open a path out of what is otherwise a hopeless dilemma.

MID antecedents such as corporations, unions, and universities nurture progress for society where it may not happen without them. Other entities that serve the role of MIDs also enforce quality standards, like science journals that demand scientific method through peer review, and professional societies that enforce codes of conduct and work

standards. Critics have observed that these societal institutions, responsible in part for increased civility, reporting of truth, and tolerance, have been weakened in the digital age. MIDs can restrengthen them.

Not every MID will be elite, but a successful elite MID will have rigorous and fair systems for evaluating and tracking the quality of data provided by members and maintaining reputations and incentives for members to provide quality data and improve it over time. This will be critical not only to ensuring a strong and credible bargaining position with data consumers, but also to allowing data producers to “level up” by increasing their abilities and reputation, and thus to earn more money. Empowered MIDs will have enough clout to sell their users’ data to data consumers relying on their standards and quality, just as prestigious universities like Harvard trade on their prestige to market publications such as the *Harvard Business Review*.

**3. Inalienable provenance.** While a MID should facilitate the efficient flow of data to high-value uses, it must not allow data (especially sensitive personal data) to be permanently sold or alienated from the control of its members. While intellectual property may be licensed, authors cannot sell their moral interest in their works.

Similarly, wherever technologically possible, transactions should involve selling access to the data for a defined use, in a way that does not allow the purchaser to retain any access to that data beyond that use. Recent advances in cryptography and the field of “differential privacy” increasingly make possible a separation between uses of data for important AI applications and control over the underlying data for broader purposes. Any use of data that is not so clearly contained should be clearly and indelibly marked as such, so the data creator can claim a share of future revenue and the right to refuse future uses that conflict with her legitimate privacy interests. It should be impossible for data to be incorporated into some intermediate system and then continuously and increasingly ambitiously used, without at least some of the associated value flowing back to the original creators.

It should be noted that the engineering infrastructure currently in place to track users and target personalized advertising is functionally similar to what would be needed to calculate what should be paid to each individual based on data originating from them. While fresh engineering will be required to implement data dignity, particularly for payments, security, and provenance, much of the engineering in the two paradigms is similar.

**4. Benefit sharing.** MIDs will become vital parts of the society and economy. To that end, regulation will be needed and should be welcomed.

For instance, a MID should ensure that a fair share of the value of data is returned to its creators. A rough approximation fair share is 70%, the historical portion of national income accruing to workers. At present, the labor share of the tech industry is far lower — 5–20%, depending on what company you look at. Any intermediary model that

doesn't have the power to bring data workers' share close to 70% and isn't structured to allow most of that value to pass to the actual data creators (and not to the MID itself) will be ineffectual, or may become a siren server, with too much concentrated power over the data it manages.

Regulation of MIDs will have to borrow from labor law, antitrust, and other precedents that aim to simultaneously protect the power of organizations representing those with little power and prevent them from becoming overbearing.

**5. Competence and professionalism.** A MID must have adequate expertise to accomplish its mission. It must possess sufficiently discreet management so that it can credibly engage in negotiations with data customers and be entrusted by those customers with confidential business details necessary to put the two parties on a position of parity. It will require technical expertise to build systems that support its unique attributes. The best MIDs will develop intelligence capacities to understand the ways data consumers use data in order to negotiate terms and conditions of use from a position of rough informational parity. All MIDs will require staff to audit and understand what is happening with the data that is already licensed on behalf of members.

**6. Biological realism.** Siren servers radiate risk outward; the effect is seen in the gig economy, for example, in which workers must “sing for their supper” — for every meal. It is hard to manage sick days or plan for old age. This is especially true in the United States, where health insurance is not universal.

MIDs should strive to create outcomes for members that reflect the true nature of a human life cycle, and not just aim for a quantitative deal measured as “fair” against some imaginary, robotic worker who doesn't age and whose needs don't change over time.

A well-designed and well-administered MID will encourage members to build up more and more royalty streams from data, so that as old age approaches, a diverse portfolio of data royalties, buffered by memberships in multiple MIDs, will provide necessary security. This is analogous to how a fiduciary financial adviser is likely to encourage a client to diversify investments.

A future person of retirement age will earn a small sum from each of hundreds of data schemes they participated in over their lifetimes (captioning pictures, commenting on products, and the like). These payments only cover what we think could already be calculated for a typical person. But most people will find a few areas of specialization, and these will vary over the course of their lives. For example, a nurse who joined a MID to provide data to machine learning schemes for nursing robots will continue to earn royalties, even as new generations of nurses with new ideas and new data gradually supplant the older contributions. The same nurse might also have joined a MID that tagged and promoted a new esthetic in sushi that drove robotic sushi chefs everywhere for a few years and still has a few fans. Another retirement-age person

might earn royalties from a canonical virtual reality experience they created during college, and also from managerial training that was incorporated into business artificial intelligence, as well as from a long chain of smaller cultural tidbits they entered during their many years on social networks. (Scenarios for the future cannot help but sound unlikely, but if we only prepare for a likely future, we're preparing for no future at all.)

**7. Cognitive realism.** We can't saddle MID members with impossible-to-understand terms and too-complex decisions. Vast terms and conditions, or choices so large and complex that members concede to whatever they are presented with, will not work. (Unfortunately, this happens frequently today with consent-driven relationships online.)

This will require innovation in algorithms and design, but some old ideas will be equally helpful. For instance, if each member of a MID can use a single virtual knob to set the price of their data from their smartphone, a great many decisions can be compressed into a single parameter. The setting would not determine an absolute price, but would simply be a bias to be added to calculations performed by the MID on behalf of members. An individual who values privacy over wealth might set the knob to the highest price, thus making personal data too expensive for companies to buy. A young person who is starting out and wishes to self-promote might set the price low. A MID might place upper and lower limits. Someone who wants to maximize profits will probably keep the knob near the center, but will tweak the setting frequently, trying to predict market fluctuations, and might pay services to assist with that. Some MIDs might have one or two other knobs, related to the value of a member's time, for example, but the effort for the individual must remain manageable, with elegant and minimized options.

Without clarity, there can be no agency, and MIDs must commit to maximizing the agency and dignity of their members. The philosophy of informed consent in health care can serve as a precedent here — but not the implementation, because our information systems have become more complex than our health systems.

**8. Longevity.** MIDs should not be designed to last forever (like a nation, for example), but to last longer than a human lifetime (like an insurance company). That is because MIDs will become the guardians of intergenerational digital wisdom and context.

MIDs must be able to form long-lasting and reliable relationships between each other. A MID that represents nurses should maintain a great multi-decade relationship with a MID that represents biological data scientists, which itself would have a wonderful relationship with a MID that represents people willing to be trial subjects for new biological sensors. MIDs should form themselves into value chains, just like players in any advanced market.

What will keep interdependent MIDs from merging into "mega-MIDs," which would be effectively siren servers themselves? While early MIDs might be motivated to become large and concentrated to offset the power of siren servers, our society and economy

will be best served if they continually work toward a world in which their unilateral discretion is reduced while their bargaining power is maintained.

Perhaps antitrust law will come to play a role in reining in large MIDs. Perhaps rules that restrict the degree to which MIDs can pit members against each other will spawn a larger number of smaller MIDs, in a manner similar to the way law firms can't represent opposing clients. Blockchain-based transparent coordination devices for data strikes (in which users simultaneously disconnect access to their data to force a platform to the bargaining table) and seeding of new internal entrepreneurship are other promising but still speculative methods of enshrining moderation.

A useful metaphor here is neural networks, which require intermediate layers of neurons that function as accumulators of feedback. Without these middle layers, a neural network cannot learn. The intermediate layers become the most persistent elements of a machine learning system, the bearers of value. MIDs will function as just such an intermediate layer, bearing the value for a whole economy and, indeed, a whole society.

## Other Ideas for Improving the Internet

Correctives other than MIDs are already widely promoted in the marketplace and the political debate, two realms that are increasingly intertwined. Comparing MIDs to the ideas below will further illuminate the motivation to adopt MIDs as a solution.

**Self-regulation.** It has become common for users to make demands on siren servers to regulate the speech and behavior of both free users and advertising customers. We see this approach in action when consumers and activists demand that platforms ban hate groups, sadists, pornographers, and so on.

Platforms have had some success checking the worst excesses of cruelty through self-regulation. But while that might address certain psychological and social degradations, it increases a platform's power over society rather than limiting it. This can only lead to some mixture of censorship and chaos.

The naive version of openness that has characterized large platforms does not work and cannot be self-regulated. Content is currently forged out of an unrelenting short-term contest for views. Inflammatory content is an example of what gets the most attention in an economy in which advertising and persuasion are the only paid products. Indeed, the softer clichés of the internet, such as cat videos or videos to soothe toddlers, will often flow into disturbing or misleading content as recommendation engines take the lead. The result is that civil dialogue is crowded out by an overwhelming amount of uncivil communication, damaging civil society in the process.

Unfortunately, it's not damaging to the companies. If anything, platforms are punished for their efforts to self-regulate. Twitter has seen valuations fall after fake accounts or

accounts associated with information warfare campaigns are purged. Facebook has seen similar drops in valuation when it has attempted to push back on fake news. This is because the metrics that investors have used since Facebook's IPO to value the platforms — such as daily active users, clicks, and so forth — are suppressed by platforms attempting to regulate activity. Indeed, fake accounts and their content are engineered to drive attention, clicks, and responses in a way that also benefits the platform. This typically results in the kind of engagement Wall Street views as “healthy” activity. Bizarrely, regulation is currently structured to motivate companies to promote this pattern, not remove it.

But if hate speech and other damaging content can be voluntarily excluded by popular MIDs (see Principle #2 above, “Quality Standards”), then there is less reason to attempt to repress it absolutely, much less to ask the platforms to do it. It becomes possible to boost high-quality content instead of repressing nasty content, because the high-quality content will have a chance, rather than being subsumed by a global mashup algorithm.

Rather than petition large corporations to govern speech and behavior, we appeal to common sense and a commitment to free speech. Horrible content existed before the internet, but it didn't crowd well-intended content off the newsstand. The reason was that when given a chance to reflect, most people turned out to not be horrible. A human-scale magazine rack turned out to be a better filter for useful content than an unbounded digital one.

MIDs retain values that have always been treasured on the internet and in open societies: All associations are voluntary, and no one is censored. At the same time, however, MIDs are an alternative to a unitary platform in which everyone competes for attention simultaneously. MIDs are democratic, not anarchic.

**Privacy regulation.** Digital privacy regulations can help protect individuals from abuses in specified cases, and we do not oppose such regulations, but they aren't systemic or future oriented. They don't redirect economic incentives to dissuade ever more innovative privacy violations. They stand in opposition to most of our principles for MIDs, and especially to Principle #7, “Cognitive Realism.”

What's more, a privacy violation is not a sufficiently coherent or complete characterization of a problem. Privacy rights without economic rights rely on concepts of consent that aren't meaningful when the uses of data have become highly technical, obscure, unpredictable, and psychologically manipulative. Cognitive realism is the best answer to privacy concerns. Once individuals have access to tools that explain the data in their lives, they will demand control over that data and be able to control it.

The concept of privacy has not transferred gracefully to the networked world, because siren servers have come to define the environment, and they cannot incorporate truly private places. This is one reason why abstract rights to data privacy are hard to enforce. No one knows if they are enforced. We are asked to trust the platforms, for

there is no room or time for a routine, independent auditing function. Siren servers are often motivated to be stealthy and to obscure the provenance of data. Even if we could trust the platforms, the nature and uses of data will continue to evolve in unpredicted ways, making it hard to write rules in advance.

Linking data privacy rights to commercial rights, though, creates an incentive for accountants and lawyers to track how data is used, and to negotiate over its use. Accountants and civil litigators can be annoying, but we should remember why those professions were invented. In a nonmarket society, there are only police to enforce rules, while in a market society, there is also civil litigation. Without contracts, every intervention is from above. Enforcement of privacy rights is a form of centralized power; distributed power is less likely to be corrupted.

Ultimately, what people need in their digital lives is not maximized privacy per se, any more than what they need in their work lives is maximized leisure. In both cases, people need, in essence, the right to be left alone: a reasonable ability to construct what is seen and known about themselves by others, reasonable limitations on what efforts are demanded of them, an accessible means of self-determination, fair compensation for what they do give up, and an affirmative environment in society for seeking meaning and happiness.

MIDs are the natural structures to help individuals realize these benefits.

**Technical decentralization.** Internet reformers are eagerly proposing new architectures to decentralize influence on digital networks, often invoking blockchain as the mechanism to do it.

Decentralization through technical architecture is an appealing idea where possible, but it is often an inadequate idea in the face of network effects and cognitive load, forces that create centralization in the first place. It doesn't always work: Open-source software and the ideal of "free" media was supposed to lead to a radical decentralization of power. Unfortunately, while they encouraged *labor* to be free, they were not able to achieve the same for capital or control of platforms. The result was a unilateral disarmament of labor to the benefit of the seductive monolithic corporations that manipulate us to extract our data.

Today, many put their faith in blockchain's push toward decentralization. The clear emphasis on economic incentives within blockchain communities holds more promise than open software communities, but until we can overcome network effects and the difficulty for any individual to navigate terms of use on his own, any push to fully decentralize will further weaken the individual's ability to resist networks. To the extent that centralized authorities, whether governmental or computational, do not or cannot surrender their power, we must aim to bolster countervailing forces, like MIDs.

Furthermore, despite techno-utopian visions for blockchain, a social contract probably can't be enforced by code alone. At some point, code must be coupled to the physical world to have any effect, and that point of coupling is where corruption, fraud, and manipulation can appear. MIDs affirmatively strengthen a social contract.

## Good for Business or Bad for Business?

A future of MIDs should be embraced, not feared, by big tech. (We work for — and we like — big tech companies.) Ultimately, platforms will benefit. But we anticipate resistance. Here we lay out some of the counterarguments:

**People want free.** Some will argue that a marketplace for data and a world of MIDs are impossible in a world where consumers are used to the “free” internet — that they will not pay for what has previously been free. We believe that objection has been answered empirically. It was once widely believed that no one would pay for online video, and yet Netflix built a business anyway. And now, many free video sources like YouTube are following with for-pay options. Similarly, Apple proved that customers will pay into an app store.

Furthermore, free isn't really free. Consumers may not have yet made it part of their cost analysis (MIDs can help them do this), but they pay higher data rates than they need to in order to support the vast overhead of the surveillance and advertising economy, not to mention costs in lost privacy, and distorted information.

**The value of data is negligible.** Some economists argue that the amount of payment available for data would be negligible, that even if Jane User could sell her data, the return would be less valuable than the effort to do it. The cost of managing a MID may outstrip the value it can return. Even without changing existing business models, attempts to calculate the value of data suggest that many Americans could earn \$500 to \$1,000 a year.

But we think that's a low estimate. This modest calculation ignores how much of the present value of data is simply off the books. It's like arguing that because women were unpaid for home work there would be no market for domestic labor. In fact, once women entered the workforce, a large market for home care was created. No one disputes that digital systems add value to the world. How much added value exists is an open question.

We believe that once the value of individual data contributions is properly calculated, the overall economy will be revealed and will grow greatly as the quality of data inputs increases. Furthermore, if the AI economy grows as anticipated, the value of data is likely to explode, just as new applications for oil (such as cars) made the value of that resource explode. We have calculated that even if AI ends up as only 10% of the economy, just the AI-feeding aspect of data dignity could deliver \$20,000 in annual

income for an average American family of four (though they might also have added expenses to pay for digital services).

But that is not all. In an advanced information economy with MIDs, individuals will be motivated to specialize in ways that are not identified as economic niches at present. For instance, a botanist who likes to hike might join a MID for people who compile useful photographs and data about trees in less developed regions. This data will be valuable to forestry companies, property managers, truffle hunters, regulators concerned with climate change, AI services that automate drone deliveries in forested areas, and many others. There are many types of valuable data that will come into existence or become more precisely, more clearly annotated and thus more useful. The point of a market is not just to distribute a finite pie, but to grow the pie. Those who dismiss the value of what people do online have forgotten this most basic benefit of open markets.

We believe that MIDs will have the potential to identify and market approximately enough new value to offset job losses due to automation and restore a labor share comparable with historical levels. There is no guarantee that this will happen, but to see the potential, suppose that there will someday be total unemployment due to total automation in a society that is advanced enough to have overcome material scarcity. (We do not propose that this situation will come about; this is only a thought experiment — though a near-universal one in tech culture.)

As we have argued elsewhere here, the automation in this scenario would not function were it not for data derived from the society. Will such a society value its data highly enough to finance itself? This is a matter of business plan philosophy. If there are only customers for advertising and persuasion, then of course the data in such a society will never be valuable enough. The society will have to function primarily on nonmarket principles, with highly centralized power. If there are customers associated with *all* the ways in which data is valuable to the society, however, then a market will have a value commensurate with the society it supports, since data improvement will be the only remaining scarcity.

Will many people provide valuable data, or only a few? At present, siren servers such as Google that define themselves as competing in an AI race are sparing no effort to gather data from everyone. If they could identify people whose data was not worth collecting, they would have every incentive to ignore those people. But that is not what is happening.

**MIDs are an existential threat to platforms.** Economists often perceive a zero-sum fight between digital platforms and data producers — who gets what slice of what's there now. We see it differently.

Tech giants will do well in a world of data dignity: They will take a smaller share of a far larger pie. Demand for ever-better computational resources and data tools will remain

intense. The burdens of content curation and data verification will be lifted from them and outsourced to an ecosystem of MIDs better positioned to manage it, both politically and economically. Rather than having everyone engaged in sneaky electronic conflict with everyone else — trying to trick each other into sharing this or believing that — data dignity will bring the interests of most market players into better alignment.

Data dignity, in fact, depends critically on big platforms. It is not a repudiation of the inevitable rise of network effects, but instead an attempt to maximize their benefits. If anything, a rational marketplace facilitated by MIDs will help resolve existential challenges posed by an anti-technology backlash that is likely to follow if we don't find a better solution than what we have today.

**AI will obviate the need for data in the future.** It is sometimes claimed, cynically, that the AI of the future will no longer need more data from people, so we might as well not worry about any lost opportunity to pay people for the data that is taken during this brief period before AI becomes self-sustaining. That claim can be debated, but it sidesteps a deeper issue.

Any advanced economy includes ever-more sectors that trade in subjectively valued goods and services, such as cosmetics, sports, entertainment, design, fashion, tourism, art, journalism, commentary, and gaming. All of these are comprised of valued expressions between people. Indeed, the value exchanged on siren servers owned by companies like Google and Facebook comes almost entirely from this category.

Rhetoric about AI can distract from the fact that these subjective sectors will eternally need data from people, because exchanging data value between people is what they do. Indeed, even if AI someday directs all the movies, and robots apply all the makeup, the tastes that drive future AI will still be derived from people. If people are to retain any volition, at core the value will still be about people providing data to each other.

Any dignified future economy that relies heavily on information technology must value the people who add the data. This cannot just be an idea; there needs to be a structure to make it so. We will need true symmetrical market bargaining and insurance that stands up for and protects the value of creators, which is only possible if pools of data creators are organized in groups like MIDs.

**If the internet isn't free, the poor will be excluded.** Another objection, frequently expressed by Facebook and other large platform companies, is that services that cost money will exclude the poor. This is a problem, but it is not a new one. Books that cost money exclude the poor. Private vehicles that enable people to seek more and better jobs exclude the poor. Markets in combination with robust democratic policies are better solutions to these problems than simply failing to account for value. Just as public libraries make books available and public transportation offers a reasonable alternative to a car, analogous institutions can and will exist on the monetized internet. Most companies that charge find ways to make their products reasonably broadly accessible.

For example, our employer, Microsoft, charges for most of the products it sells, yet these same products are widely adopted in poorer countries at reasonable, tailored prices.

Moreover, the existence of free internet services has coincided with an extreme increase in income inequality and declines in the share of income paid to work. What we've witnessed in the past two decades is new wealth attaching itself to those who are close to the largest computer resources; the free internet isn't really serving the interests of the middle class, much less the poor.

. . .

It is not easy to talk about ways to improve the internet, largely because conversations on the topic typically happen *within* the internet, a venue that has been overtaken by siren server systems that are motivated to increase engagement by making people upset. This means that conversations quickly become agitated and ugly.

We have therefore deliberately made our argument slowly, in detail and at length. This length aims in some ways to serve as a corrective for our inability to answer the many queries we have received from nascent MIDs, and is our own attempt to pool together all those responses.

No single proposed venture we have seen has met all the requirements for data dignity that we've set forth here. Invention and a sense of adventure will be required to meet them all. The argument for MIDs flows from fundamental principles. The requirements will surely evolve, however, as more is learned.

And they must continue to evolve. The influence of the internet on all aspects of human experience is so great that we must demand data dignity if we are to retain any dignity at all. The Big Idea

**About the authors:** *Jaron Lanier* is Office of the Chief Technology Officer Prime Unifying Scientist (OCTOPUS) at Microsoft, a musician, and the author of several books, including *Who Owns the Future?* *E. Glen Weyl* is principal researcher at Microsoft Research in New York City and teaches economics at Princeton. His most recent book, coauthored with Eric Posner, is *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*.