The Ethical Imperative to Identify and Address Data and Information Asymmetries

Stefaan G. Verhulst, The GovLab

Francis Bacon is generally known as the man who developed the scientific method in the 16th century, but he is also credited with coining the aphorism that "knowledge is power" (Meditationes Sacrae, 1597). As economies and society in general have grown more complex, the truth of this aphorism has only increased. In 1970, George Akerlof published "The Market for Lemons," his seminal paper on information asymmetries and how they can define markets and establish winners and losers. Although Akerlof focused on the used car market, his insights applied across sectors and industries. In 2001, Akerlof, Michael Spence and Joseph Stiglitz were awarded the Nobel Prize in Economics "for their analyses of markets with asymmetric information."

The insight that knowledge is power is more relevant today than ever before. The data age has redefined the very notion of knowledge (as well as power), leading to a greater reliance on dispersed and decentralized insights as well as to new forms of innovation and learning, such as artificial intelligence (AI) and machine learning (ML). As Thomas Piketty (among others) has shown, we live in an increasingly stratified world, and our society's socio-economic asymmetries are often grafted onto data and information asymmetries. As we have documented elsewhere, data access is fundamentally linked to economic opportunity, improved governance, and citizen empowerment. The need to address information asymmetries—and their resulting inequalities of political and economic power—is therefore emerging as among the most urgent ethical challenges of our era.

Even as recognition grows of this imperative, society and policymakers lag in their understanding of the underlying issue. Just what are information asymmetries? How do they emerge, and what form do they take? What forces and power structures perpetuate or deepen these asymmetries? I argue that it is a mistake to treat this problem as homogenous. In what follows, I analyze three types of asymmetries. Although closely related, each one emerges from a different set of contingencies, and each is likely to require different policy remedies. The focus of this short essay is to describe the different types of asymmetries. Further research could help define solutions that are contextually appropriate and fit for purpose.

1. Data Asymmetries

Data asymmetries are the classic—and most commonly recognized—form of asymmetry in today's digital marketplaces. They occur whenever there exists a divide or disparity in access to data. The nature of this divide can take many forms, however, depending on the relationship between information holders ("owners") and information users.

Consumer-to-business (C2B) asymmetries dominate much current public discussion. Such asymmetries have grown increasingly common with the datafication of consumption patterns and typically occur when companies collect data on their users while providing services or

selling goods. For example, companies might collect data related to transaction or browsing histories or a variety of socio-demographic markers. As a result, companies often possess a disproportionate amount of data on their users—information that users may not even be aware of having surrendered. This data allows companies to target advertisements, optimize internal operations, train AI systems, and pursue other objectives. This increased "dataveillance" prompts ethical discussions around "data sovereignty" and how to provide agency to individuals about how others use their data as well as a search for new policy solutions.

In addition to C2B data asymmetries, various other forms of asymmetry also exist. Among the most consequential are business-to-business (B2B) asymmetries. Recent years have witnessed the emergence of a number of large data monopolies that dominate their sectors and the broader economy. These companies have access to huge amounts of data collected and processed across various domains (e.g. search data, location and mobile phone data, consumer spending data) and their ability to combine and derive insights from this data or train ML algorithms results in de facto barriers to entry. There are concerns that B2B data asymmetries may be stifling innovation and competition as well as hurting the rights of consumers, leading to calls for greater regulation and better enforcement of antitrust law, perhaps extending so far as to the breakup of some of these large players.

Other data asymmetries worth considering relate to business-to-government (B2G), in which government may be hampered in developing data-driven policies or providing targeted services without access to data and insights that the private sector may possess—a topic considered by the High Level Expert Group to the European Commission on B2G Data Sharing (of which I am a member). Other data asymmetries relate to government-to-citizens (G2C), in which data collected by the government is siloed and hoarded without transparency or without making it accessible to society at large through, for instance, open data platforms.

Concerns for privacy harms and increased surveillance in all of the above cases are real and require careful consideration and mitigation. Yet too often, the spectre of harms to privacy and civil liberties is used to justify limiting access to data that could provide for transformative public value. Limiting access may, however, increase the asymmetries without fully addressing the privacy concerns or establishing trust in how data is being collected and handled.

2. Information Asymmetries

Data asymmetries occur when there are disparities in access to the raw material of the information society. But even when such disparities are overcome, there often exist pervasive inequalities in the extent to which individuals and groups can actually benefit—for example, by deriving new insights or informing innovation—from their formal access. In short, stakeholders differ in their abilities to translate data into actionable information.

The data ecology is highly complex and rapidly growing in complexity. The legibility of data depends significantly on the technical, financial, and human resources of organizations that collect, store, and access data. Smaller firms and organizations, as well as individuals, may be at a particular disadvantage, as they frequently lack the know-how to surface the signal in noisy

data. Smaller organizations may similarly be at a disadvantage at a time when so much of the data ecology depends on collaboration and data integration. Their relative lack of resources may not only limit their ability to derive insights from data but also to operate as equal partners when defining terms and other parameters of collaboration. Information asymmetries are an under-recognized yet growing problem in the data ecosystem and, if left unaddressed, can undermine trust in those organizations and initiatives that are focused on re-using data for other purposes than collected (such as Sidewalk Labs in Toronto).

3. Intelligence Asymmetries

The datafication of life and the economy has spurred the development of a number of new technologies, among the most prominent of these being ML and AI. While offering tremendous potential, these technologies also pose risks and are leading to new forms of asymmetry. My colleagues and I call these intelligence asymmetries.

Intelligence asymmetries occur when there is a discrepancy among actors in their ability to understand (i.e., "look under the hood of") the algorithms and processes that are responsible for an increasingly broad range of automated decisions. Today, AI plays a role in determining the outcome of loan applications, bail and parole requests, and other important matters. How these decisions are made is increasingly opaque and contested. The fact that different stakeholders have differential insight into both the algorithmic processes themselves and the underlying data (used to derive those algorithms through ML) is an increasingly worrying form of asymmetry. It is now widely recognized that AI algorithms often contain inherent biases; addressing intelligence asymmetries through greater traceability is therefore often a matter of social justice and wider socio-economic equity.

It is worth pointing out, too, that intelligence asymmetries may not exist simply among human stakeholders. The opaqueness of AI decision-making means that humans are often at a disadvantage relative to machines in several industries and in completing several job functions as well as in the realm of conducting evidence-based decision-making more generally. While grim prophecies about an imminent takeover of the world by robots are surely exaggerated, there are clear and emerging signs of smaller-stakes rivalries between humans and machines in the areas of AI and ML. These emerging human–machine asymmetries, as much as conventional asymmetries, should be proactively addressed by policymakers. Overall, there is a need for new norms and standards regarding how training data is collected and used, and a general insistence on greater transparency when it comes to AI-driven decision-making.

Conclusion

The above has outlined three types of asymmetries that increasingly define the data age. As I show, these asymmetries are important not only to achieve greater equity in access to data but also because they are at the root of many pernicious and growing socio-economic power inequalities. This essay has been primarily descriptive. While I have described the problem, I do not have sufficient space to prescribe solutions.

Nonetheless, there does exist a growing set of possible solutions that could provide a toolbox, and these possibilities should be explored more fully in future papers. For example, some argue for a greater focus by policymakers and others on data liquidity (and portability), which would enhance citizens' and others' agency over data. This focus could also generate an ecosystem of responsible data exchanges. Data holders and demand side actors could also experiment with new and emerging operational models and governance frameworks for purpose-driven, cross-sector data collaboratives that bring to bear previously siloed datasets. In addition, there is a need to further define and professionalize the notion of "data stewards", individuals or groups who are tasked with managing data and its ethical and responsible re-use within organizations. The potential of emerging technologies such as Distributed Ledger Technologies and others to address information asymmetries also need further attention.

Beyond these specific possibilities, there is a need to demystify data, provide transparency into existing relationships - how digital systems collect and process data, their intended purpose and who is responsible for that data activity- and crucially, provide mechanisms for feedback. No citizen or stakeholder is removed from the ramifications of datafication, including its resulting asymmetries. Francis Bacon was right all along: Knowledge is power, and the extent to which organizations can address these asymmetries will depend in large part on public engagement (such as data assemblies), which, in turn, relies on greater literacy and awareness on the part of citizens and policymakers.

The above—and more generally our work at The GovLab—are small steps in the direction of raising public awareness, and thus empowering citizens with knowledge about the data age and the need for improved data governance. Much more needs to be done to address the ethical ramifications of data and information asymmetries.