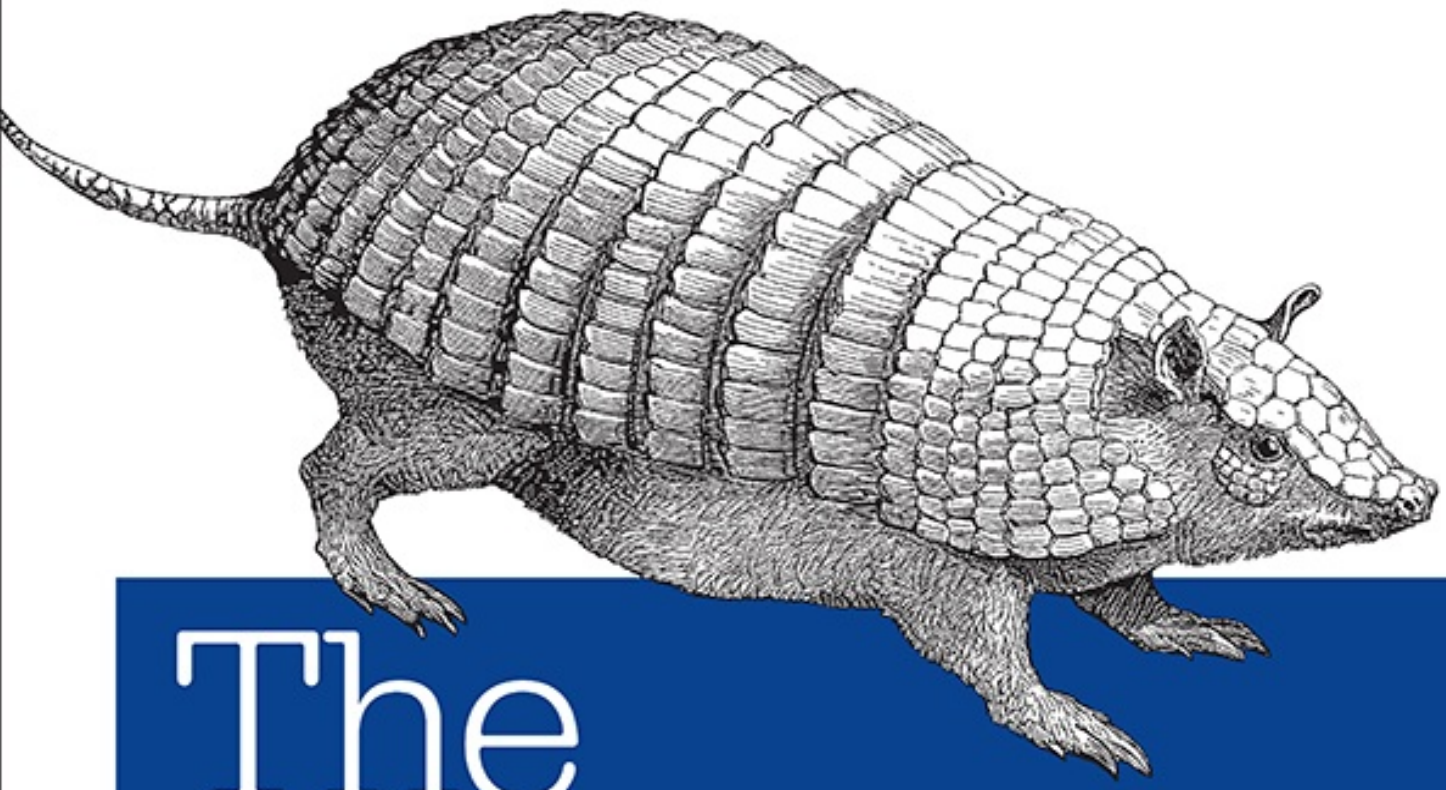
O'REILLY®



The Architecture of Privacy

ON ENGINEERING TECHNOLOGIES THAT CAN DELIVER TRUSTWORTHY SAFEGUARDS

Courtney Bowman, Ari Gesher,
John K. Grant & Daniel Slate

The Architecture of Privacy

Courtney Bowman, Ari Gesher, John K. Grant, and Daniel Slate

Edited by Elissa Lerner

O'REILLY®

Beijing • Boston • Farnham • Sebastopol • Tokyo

The Architecture of Privacy

by Courtney Bowman, Ari Gesher, John K. Grant, and Daniel Slate

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (http://safaribooksonline.com). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com.*

- Editors: Elissa Lerner, Heather Scherer, and Mike Loukides
- Production Editor: Colleen Lobner
- Copyeditor: Christina Edwards
- Proofreader: Gillian McGarvey
- Indexer: WordCo Indexing Services
- Interior Designer: David Futato
- Cover Designer: Ellie Volckhausen
- Illustrator: Rebecca Demarest
- September 2015: First Edition

Revision History for the First Edition

- 2015-08-26: First Release

See http://oreilly.com/catalog/errata.csp?isbn=9781491904015 for release details.

978-1-491-90401-5

[LSI]

Foreword

When I was an undergraduate majoring in computer science a few decades ago, books published by O'Reilly and Associates possessed talismanic power to me. As it happened, some of the earliest O'Reilly books were being published during my freshman year, and their mix of great writing, beautiful production value, and hyper-specificity were tailor-made for a young geek learning about Unix, perl, and the Internet for the first time. My dorm room bookshelves were lined with a rainbow of brightly colored book spines. Across my desk roamed a veritable menagerie of cover illustrations, from camels to grasshoppers, from crabs to crowned pigeons. I read every line of Dale Dougherty's book, and Cricket Liu's book; the tattered pages of my copy of Larry Wall's *Perl Programming* bega to fall apart in my hands. I even splurged (well, my parents did) and bought the entire set of pricey X Window System guides, although I confess that I didn't read most of those.

I tell you this history to come clean: I gladly would've written a Foreword to contribute text to an O'Reilly book to honor my twenty-year-old self's obsession even if that book was just average. Wha a happy moment it is for me, then, to be able to contribute front matter to an O'Reilly book that is much more than just average. You hold in your hands (or view on your screen) a fantastic contributio to the burgeoning literature of privacy engineering.

Privacy requires a dialogue between two types of people: those who speak policy and those who spea engineering. The most important word of that sentence—and the part that many people fail to understand—is "dialogue." In many other spaces where tech touches policy, these two tribes stand across a chasm, reacting *to* one another but not conversing *with* one another. Thus, in modern digital copyright policy, creators create, technologists protect and circumvent, and lawyers create laws and spur lawsuits reacting to these actions. In telecommunications policy, engineers engineer and lawyer react and respond.

And even in a field that many people—including many experts—mistakenly think relates closely to privacy—information security—the dialogue is hardly essential. Security folks traffic in the impossible and possible—this crypto works or it is broken. The benchmarks for "victory" and "defea are entirely internal to the discipline. And the law and policy folks sit on the sidelines and react and respond.

Privacy doesn't work this way. A privacy engineer, at least a good one, cannot live in ignorance of la and policy because the ideas of "victory" and "defeat" for privacy cannot be subjected to correctness proofs and measurements of algorithmic complexity. Engineers can tell you how to dial down or dial up a particular information flow, but it requires a source external and foreign to the engineer's core training—maybe the law department, public relations, the shareholders, or the engineer's moral compass—to determine right and wrong, acceptable risk or not, privacy violation or not.

As only one example, take the topics of data anonymization and re-identification, topics central to work I have done. This much we now know: "data can either be useful or perfectly anonymous but never both." I said this once, and much ink has been spilled trying to prove me wrong. I'm not wrong but at the same time, I am not being very interesting when I say it. Of course scrubbed data can be unscrubbed. You would be foolish indeed (or worse, trying to sell anonymization consulting services to fail to realize that modern improvements in data processing, auxiliary data, and storage could lead to any other result. But recognizing this boring truth is far from knowing what to do about it. The

lesson of powerful re-identification isn't that we take our ball and go home. But it is just as unacceptable to continue to act as nothing has changed.

You cannot "solve" the re-identification problem without lawyers who understand tech and techies who understand policy. (I try to be both, as I went to law school a few years after obtaining that CS degree and now teach law.) It might be enough to delete eighteen identifiers or it might not. It might be enough to encrypt the data and leave the key with "Joan in the front office," or it might not. Mayb you can distribute the data to a trusted third party, or maybe you shouldn't. It's nuance and hard choices and a dialogue between engineers and lawyers all the way down. We need to train a new bree of privacy engineer, and it starts with creating a literature elaborating this new discipline.

This bringing together of engineering and law means that it takes an exceptional group of people to come together to write a proper book on this topic. Luckily for you, and for the privacy community a a whole, the authors of this book compose such a group. They include top-notch engineers and good lawyers. But more importantly, they include people steeped in the weird mental gymnastics, arcane training, and time spent in rooms in Silicon Valley and state and national capitals required to be calle privacy experts.

It is even luckier for you that they happen also to be extremely engaging writers. This is a very well-produced and organized book. It has the virtues of clarity and modesty, two virtues often lacking in books written by engineers. I call the book modest, because it recognizes that this field is new and th we don't really even yet understand what we mean when we call somebody a privacy engineer.

I'm not sure I'm ready to call this book a classic or a new entrant into the canon. I think time will tel and I hope I am invited back to update this Foreword for the second edition, when I can trot out those labels, if they stick. But this seems to me at least to be a very useful book, one that fills a gaping hole in the current literature. I'll happily place my copy of this book on my shelf. I have a particular spot mind where I think it will fit in well.

Paul Ohm

Professor of Law,
Georgetown University Law Center

Boulder, Colorado

July 2015

Preface

Who Should Read This Book

This book is not for privacy experts.

If you are looking for an in-depth discussion of the legal implications of the *Kyllo v. United States (2001)* Supreme Court decision or thorough exploration of how to implement differential privacy in a database, then you should look elsewhere. There is no shortage of invaluable literature on these and many other privacy-related topics, and we recommend it to those readers.

This book is for privacy beginners. Those who have a niggling worry that the technology they are creating raises privacy concerns and want to do something about it, but who also are not going to spend the next 10 years perusing privacy case law and academic papers trying to figure out how to port those lessons into lines of code.

It's also for those who have a basic understanding of law and privacy policy, but who cannot read lin of code to save their lives. It's OK—you don't need to be an engineer to read this book. And even aft you read it from cover to cover, you still won't be able to write code for access controls. But you wil have an understanding of the range of possibilities when it comes to basic privacy-protective technic capabilities. You'll know what to ask your coders to build.

You may be surprised how frequently what you build has privacy implications, but we live in a time increasing capabilities for personal re-identification. This book will help you be familiar with how to spot privacy questions. If you read nothing else but Chapter 1, you'll understand better how to judge whether or not what you're doing is connected to data privacy.

Whether you're building a new smartphone app in your dorm room or a database empire from your garage, *The Architecture of Privacy* will be your first step into the world of privacy engineering.

Why We Wrote This Book

Decisions made by engineers can unleash technology upon the world that can significantly affect fundamental rights. In some cases, this can yield positive outcomes such as the creation of new platforms to exchange ideas that catalyze change in the world's most oppressive regimes. In other cases, new technologies can become tools of repression and control, enabling governments and corporate interests to track and manipulate individuals with surprising subtlety and at remarkable scale. With such high stakes, it must be in the interest of more than just lawyers and bureaucrats to recognize, promote, or guard against these potential outcomes as needed.

This book is, in part, an effort to empower the engineer. Successful technology is not just technology that works; it is technology that works while simultaneously offering capabilities that protect privacy and civil liberties. Readers of this book will not have to watch helplessly as their technology is misused, nor will they have to wait for others to try to curb that misuse. Instead, they will have the tools to recognize potential risks and design against them, sparing much headache and heartache.

This book is distinctive in the realm of privacy literature as it is written by technical authors who approach privacy and civil liberties from what is currently a highly atypical perspective: how to engineer technologies that will deliver trustworthy safeguards capable of supporting liberal-democratic principles. By contrast, most privacy books are written by professional scholars who take law and policy as their starting point and treat technological concerns as ancillary at best and menacing at worst, which is hardly a perspective that will encourage the engineers of the world.

But this book is not just for engineers. For the non-engineers who read this book—the academics, lawyers, and policymakers—we offer a new perspective. The policy choice is not simply to build or not to build, to ban or not to ban. Instead, these readers will find that engineers can offer an arsenal o technical tools that can form the building blocks of nuanced policies that maximize both privacy protection *and* utility. This book provides a menu of what to demand in a new technology.

A Word on Privacy and Technology Today

Over the past few years, the public has become aware of the vast scale of data collected and held by governments and corporations. As we produce more data about ourselves through the ubiquitous use of electronic payment systems, mobile devices, and cloud computing services, the institutions aroun us have concluded that this data holds tremendous value. Unfortunately, the private companies and government agencies that hold data about us do not always put appropriate safeguards in place to prevent deliberate or accidental privacy violations. Sometimes this is because of gaps in their interna policies, or because they misjudged risk or their ability to mitigate it. But sometimes it's because these organizations don't have data management systems that offer the technical capabilities necessary to support robust, privacy-protective policies.

That need not be the case. Today we know enough to design systems that build in, from the beginnin appropriate safeguards that can substantially reduce the chances of abuse or mistakes when handling people's sensitive data. We believe it is time to move away from all systems that don't have these straightforward and sensible protections in place. We have become heavily reliant on advanced information technology, and we need to be able to trust our systems and each other with our data.

Effective privacy-facilitating technology is designed to minimize the friction between a person and their work. Capabilities can and should be designed in such a way that they enable privacy-protective policies and procedures while creating as few hurdles as possible in using the system. The easier privacy protections are to use and the more unnoticeable they are to everyday users, the more likely these protections will be embraced. As soon as a privacy-protective feature becomes cumbersome, some users will look for ways to avoid it or develop shortcuts that will undermine its overall effectiveness. We advocate reducing potential friction by adhering to what is sometimes referred to a "privacy by design"—an approach that incorporates thinking about privacy-protective features and implementing them as early as possible. Capabilities that are part of the core functionality of the product are far less likely to cause friction than those simply grafted on to the technology late in the development process. Specific advice on how to incorporate privacy by design into your product can be found in extensive documentation on the topic elsewhere.

It is important to note that nothing described in this book could be said to *automatically* protect privacy. Simply having these capabilities in your system won't guarantee that privacy is protected.

Rather, these capabilities must work in concert with legal frameworks and policy in order to be effective. Privacy law is an extremely nuanced field that often depends on subjective evaluations of the legitimacy of certain actions (and those evaluations can change rapidly depending on outside factors)—something that is very difficult to hardcode into a technology.

Access controls, for example (see Part II), are a powerful tool for managing data use, but a user must configure those controls in order to ensure that data is accessed by those who have the authority to se it, and denied to those who do not. Meanwhile, the mere existence of audit logs (see Part III) is not enough to ensure rigorous oversight of system usage—someone must actually read those logs and tak effective action when they see misuse of the data. Though just about anything is possible in the world of technology, we should maintain healthy skepticism of any technology that claims to automatically protect privacy while maximizing data utility.

Most likely, any attempt to automate privacy protection is going to lead to a system that is either unnecessarily restrictive, thereby undermining the utility of the system, or too permissive, thereby leaving ample room for misuse of the data (which might not be caught because oversight is reduced d the erroneous assumption that the system can govern itself).

Navigating This Book

We have tried to write this book in a way that allows readers to skip around, focusing on the topics most relevant to their needs. But we've also tried to ensure that the book hangs together as whole. Ou narrative thread therefore goes something like this:

Whenever you collect and process personal and/or sensitive data, you have an obligation—moral in a cases, legal in most—to protect that data from theft, misuse, and abuse. You are directly responsible for designing and implementing security-enhancing and privacy-protective technologies and policies This is hard! Understanding the different ways in which data can be personally identifying, recognizing the privacy risks associated with different technologies and use cases, implementing measures to mitigate those risks without compromising your original goals, and staying up-to-date on relevant law and policy are complex challenges, and there's no guaranteed recipe for success. There are, however, several broad categories of technology and policy that are foundational to protecting privacy and civil liberties, and you'll want to build on these strong foundations.

The opening four chapters of this book focus on the fundamental building blocks necessary to create privacy-protective system. Chapter 1 is a brief history of the intersections of informational privacy, technology, and privacy law, which situates the reader in the context surrounding these issues. Chapters 2 and 3 cover the data collection technology, policy, and practices that should be transparen to your users or data subjects and should ensure that the kind and amount of data collected is proportional to your product's or service's stated purposes. Chapter 4 addresses high-level information security technology and policy needed to protect data from theft and other forms of *unauthorized access*.

Privacy technology and policy should ensure that data accessed through *authorized* means is protecte from misuse and abuse. This goal is best achieved through some combination of access control and oversight measures.

Chapters 5, 6, and 7 address various ways of restricting and controlling authorized access to data. We describe how to grant differentiated access to the various levels of your system (e.g., application, network, hardware, etc.) and apply controls to varying levels of granularity in your data (e.g., system level, record-level, cell-level, etc.). We describe different types of access (e.g., read, write, discovery etc.) and conditions under which access is granted (selective-, purpose-, and scope-driven revelation). We describe federated system architectures that delegate some access-control decisions to the owner of systems separate from your own.

Chapters 8, 9, and 10 center on oversight, the necessary counterpart to access control. In order to hold the system and your users accountable, we present techniques for logging user activity in a way that makes data use auditable. We explain how data retention policies and data-purging technologies should be designed and implemented in a way that complies with regulations and minimizes privacy risks without compromising the usefulness of the system.

In Chapter 11, we walk through several case studies that demonstrate how these various building blocks can be assembled to solve real problems. In Chapter 12, we describe the role and responsibilities of the Privacy Engineer, an individual who will become increasingly critical to companies that process personal information. Finally, in Chapter 13, we share some thoughts on the future of privacy and how you can prepare for it.

In general, think of the capabilities described in the chapters that follow as a set of building blocks. They can be combined in a variety of ways to support different privacy imperatives. However, not all of these capabilities need to be used in every information system, and not all privacy issues that might arise from the use of those systems can be solved by these technologies.

Safari® Books Online

NOTE

Safari Books Online is an on-demand digital library that delivers expert content in both book and video form from the world's leading authors in technology and business.

Technology professionals, software developers, web designers, and business and creative professionals use Safari Books Online as their primary resource for research, problem solving, learning, and certification training.

Safari Books Online offers a range of plans and pricing for enterprise, government, education, and individuals.

Members have access to thousands of books, training videos, and prepublication manuscripts in one fully searchable database from publishers like O'Reilly Media, Prentice Hall Professional, Addison-Wesley Professional, Microsoft Press, Sams, Que, Peachpit Press, Focal Press, Cisco Press, John Wiley & Sons, Syngress, Morgan Kaufmann, IBM Redbooks, Packt, Adobe Press, FT Press, Apress, Manning, New Riders, McGraw-Hill, Jones & Bartlett, Course Technology, and hundreds more. For more information about Safari Books Online, please visit us online.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

- O'Reilly Media, Inc.
- 1005 Gravenstein Highway North
- Sebastopol, CA 95472
- 800-998-9938 (in the United States or Canada)
- 707-829-0515 (international or local)
- 707-829-0104 (fax)

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at *http://bit.ly/architecture-of-privacy*.

To comment or ask technical questions about this book, send email to *bookquestions@oreilly.com*.

For more information about our books, courses, conferences, and news, see our website at *http://www.oreilly.com*.

Find us on Facebook: *http://facebook.com/oreilly*

Follow us on Twitter: *http://twitter.com/oreillymedia*

Watch us on YouTube: *http://www.youtube.com/oreillymedia*

Acknowledgments

All the authors wish to acknowledge the extraordinary efforts of Elissa Lerner, editor of this book. Pu simply, this book would not exist without her tireless efforts to herd this unruly band of cats as they blew through deadline after deadline and were readily distracted by new and exciting tangents.

We also wish to thank Palantir Technologies and its CEO, Dr. Alex Karp, for encouraging us in this effort. Although the book and its contents in no way represent the views of Palantir Technologies or any of its other employees, these authors would not have met and set out on this course without the support of the Palantir family and their tireless dedication to making the world a better place through technology.

We acknowledge a huge debt of gratitude to Dr. Lawrence Lessig, whose work in this space inspired both the title of this book and our whole approach to the interaction of legal and technical code. We also thank Paul Ohm for contributing an insightful Foreword to this book. Do yourself a favor—find everything that these two have ever written and read it.

We also wish to thank those who provided invaluable comments on this book at various stages of its life: Asher Sinensky, Kyle Erickson, Brendan Cooney (and the legal ninjas), Andy Oram, Nathan Good, and Seth Schoen.

Special thanks to the rest of the Privacy & Civil Liberties Engineering team at Palantir for acting as our research arm and teachers on all things privacy. Special thanks to the distinguished members of the Palantir Council of Advisors on Privacy and Civil Liberties (PCAP) for providing encouragement and fodder for this effort through so many enlightening discussions. Special thanks to the engineering teams at Palantir, for spending years imagining, building, and perfecting many of the architectures that we describe in this book. In a very real way, the authors are just the messengers, the documenter

of the hard technical work that go into creating these systems.

Special thanks to Mike Loukides at O'Reilly for being as excited about this book as we were and helping us to make it happen. Inspiration for this book came out of meetings at O'Reilly's Foo Camp specifically a session run by Brian Fitzpatrick and Harper Reed on Internet privacy in the post-Snowden era.

Courtney Bowman

To my co-authors, who enrich my working life with their erudition and passion for the content of this text, and whose company and good humor were the epidural to this protracted labor of love. To Kyle Erickson and especially Elissa Lerner for knowing when humor, gentle prodding, good-natured publi shaming, and other more medieval editorial machinations were needed to prod me along; I quite literally would not have done it without your indefatigable encouragement. And, most of all, to Sarah whose support, understanding, kindness, and unwavering affection remind me daily that the sacred spaces we aim to preserve and protect through privacy engineering matter most when the personhood cultivated therein can ultimately also be shared.

Ari Gesher

To my wife, for indulging yet another professional distraction that draws me out of our happy home and for taking care of the bedtimes that I'll miss while I'm out playing author. To my children: you are the inspiration for my wanting to make the world a better place to live in. Consider this a small part of my efforts to build a world that is safe for you to live in. To Chris Dibona and Tim O'Reilly f your encouragement, sometimes intentional and sometimes incidental. And to my parents: for never being impressed enough with my work to let me feel satisfied. You keep me moving. And finally, to my co-authors for tolerating my rewrites and doing the bulk of the work in writing this book.

John K. Grant

To Mike van Opstal, the engineer in my life. To my family and friends, who have to listen to me rant and rave on privacy and civil liberties on a daily basis and who hoped that this effort might purge my soul (no such luck). To Dr. Karp and Palantir for believing that privacy engineering could be a real jo and letting me turn my passion into my life. To those who risk their lives every day to protect the ideals of a free society and labor to bring freedom to all those who want for it.

Daniel Slate

To my teachers, who shared with me the love of a good sentence. To my friends, Elissa Lerner and Kyle Erickson, from whom I have learned so much, and who, with their erudition, refined linguistic sensitivity, and dauntless persistence for excellence, made this a far better book than it would otherwise have been; they are as much authors of this book as any of us, regardless of what ended up on the cover page, and all should know this. To John, Ari, and Courtney, for their wisdom and good humor along the path of shared suffering. To my family and those who came before them who left th lands of tyranny to build lives of ordered creativity in a new, free land, and without whom none of th

would have been possible.

Part I. Getting Started

You have decided to build a new technology that processes data about people. Where do you start? In Part I, we walk you through the initial steps that lay the foundation upon which your privacy-protective framework will be built. Chapter 1 defines the concept of privacy and the critical role of th engineer in shaping that concept through technology. We then raise some preliminary questions regarding when and how data is collected, which can be explored in great depth in privacy literature. While our book largely focuses on the management of data *after* it has been collected, data-collectio considerations themselves do shape privacy architecture. We therefore provide a high-level discussic of data collection in Chapters 2 and 3. Finally, protecting data privacy necessarily involves ensuring that data is secure. Since the literature on information security techniques is substantial, Chapter 4 provides a basic discussion of the topic as background.

Chapter 1. What Is Privacy?

> *Privacy incursions in the name of progress, innovation, and ordered liberty jeopardize the continuing vitality of the political and intellectual culture that we say we value.*
>
> *Julie E. Cohen, professor of law, Georgetown University*

Privacy is important. These three words comprise the philosophical compass for this book, and summarize (albeit inelegantly) the eloquent description above regarding the consequences of ignorin privacy. For us, privacy serves not only as a bulwark against threats to individual liberty and society as we know it, but also as a cornerstone of a thriving economy rife with innovation.

There has long been and continues to be roiling societal debates on the topic of privacy. Every reader of this book will come with their own conception of why privacy is important. What do you see as a threat to privacy? How significant are those threats? And most importantly, what role will your technology play in shaping the world in which those threats exist? If you are reading this book, then you have probably asked yourself these questions and, in your own way, reached the same conclusion we have: privacy is important.

Proceeding from that premise, we then assert that engineers can and should take privacy into account when designing and building technology. There is a long history of interaction between policy and technology that demonstrates just how important a role engineers can play. Thinking carefully about the architecture of privacy will show that it is possible to build systems that make it substantially easier to protect privacy and much more difficult to violate it, intentionally or otherwise. This book will help you do that.

How to Think About Privacy

In order to build technology that can help protect privacy, we must first understand privacy and how is shaped by law, policy, and technology. Though we often take its meaning for granted, privacy is neither a simple concept nor can it be assumed that everyone defines it the same way. Privacy can encompass a broad swath of sometimes interrelated and often overlapping ideas. It is also a moving target—the concept changes and adapts over time.

In this section, we define privacy for the purposes of this book. We also examine how technology has interacted with legal and policy development (and vice versa) to shape the concept of privacy. This is not meant to be a comprehensive history of privacy, but rather to provide some context to this complex interaction that will help you understand the broader environment in which your technology must operate.

Defining Privacy

A single definition of the word “privacy” has been historically difficult to pin down. Definitions of privacy have always been reflections of contemporary contexts, resulting, perhaps unsurprisingly, in

what legal scholar Daniel Solove describes as a "concept in disarray."[1] Consequently, this concept ca plausibly encompass no less than the "freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations."[2]

Even documents regarded as essential bulwarks against encroachment on individual privacy turn out to be surprisingly vague on the topic. The United States Constitution, for instance, does not contain the actual word "privacy."[3] Other documents do not eschew the word, but they do not offer much hel in defining it. The Universal Declaration of Human Rights, a component of the United Nations-create International Bill of Rights, asserts in Article 12 that "No one shall be subjected to arbitrary interference with his privacy." The European Convention on Human Rights, for its part, was only abl to muster a "right to respect [the individual's] private and family life, his home and his correspondence." Consequently, it has been left to legislatures, courts, advocates, and academics to actually flesh out the elusive meaning behind these seven letters.

Broadly speaking, experts sort conceptions of privacy into two categories—*informational privacy*, which concerns "the collection, use, and disclosure of personal information," and *decisional privacy*, which relates to "the freedom to make decisions about one's body and family."[4] Given that this book focuses on technologies that work with data, we will concentrate our discussion on informational privacy. However, this should not be read to suggest that the adoption of these capabilities will only affect informational privacy. If information truly is power (as is frequently asserted), then the ability to control information about oneself can have a direct effect on the freedom one has to think and act independently. Thus, informational privacy cannot be wholly divorced from decisional privacy, and addressing one necessarily implicates the other.

A Short History of U.S. Informational Privacy

The concept of informational privacy has evolved over time. Tracing its history shows that the development of technology and privacy law and policy are closely intertwined. Changes in one area can have significant effects on the other. A historical review also illustrates how, in many cases, the same core issues we face today are merely the latest permutation of long-standing challenges. Understanding how informational privacy has developed in one jurisdiction will not only help us understand its current state but also its potential evolution in the future.[5]

More than 120 years before the seeming omnipresence of information-sharing platforms like Facebook, Instagram, Snapchat, and their kin, there was Kodak. In 1888, the Kodak camera was introduced to the American public, allowing anyone to capture and share moments of peoples' lives like never before. Concerns about the privacy implications of the new technology quickly followed. "Beware the Kodak," lamented *The Hartford Courant*, "The sedate citizen can't indulge in any hilariousness without incurring the risk of being caught in the act and having his photograph passed around among his Sunday School children."[6]

The legal community soon took notice. In 1890, Samuel Warren, a prominent Boston attorney, and Louis Brandeis, later to serve as a Supreme Court justice, published "The Right to Privacy" in the *Harvard Law Review*, an article widely considered one of the most influential in the American legal canon and still cited in court opinions to this day. The article began by briefly charting the

development of the concept of privacy up until that point before determining "Recent inventions and business methods call attention to the next step which must be taken for the protection of the person…. Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life…'"[7] After outlining the perceived harms of these intrusions, Warren and Brandeis looked to the existing common law (i.e., law developed over time by judges as they decide cases) for the foundations of a "right to be let alone."[8] Notably, they also delineate limits to this right suggesting that at some point "the dignity and convenience of the individual must yield to the demands of public welfare or of private justice."[9]

Thirty years later, Brandeis went on to erect yet another pillar in privacy history with what became one of the most frequently cited dissenting opinions in U.S. Supreme Court history. In 1928 in *Olmstead v. United States*, the Court determined in a 5–4 decision that federal agents could wiretap a phone without obtaining judicial approval. In a fiery dissent, Brandeis reaffirmed the importance of "the right to be let alone—the most comprehensive of rights and the right most valued by civilized men." Brandeis chided the majority that "time works changes, brings into existence new conditions and purposes" and therefore the Court must be prepared to apply constitutional protections to situations not envisioned by the Framers, which in this case meant applying the Fourth Amendment protections against unreasonable search and seizure beyond the "sanctities of a man's home." He warned that technology would continue to challenge the Court's conception of privacy protection. "The progress of science in furnishing the Government with means of espionage is not likely to stop with wiretapping," he wrote. "Ways may someday be developed by which the Government, without removing papers from secret drawers, can reproduce them in court, and by which it will be enabled to expose to a jury the most intimate occurrences of the home."

However, it would be nearly four decades before the Supreme Court would recognize Brandeis' prescience, and it would be an invention that had at that point already existed for more than 90 years, the telephone, which would inspire the Supreme Court to a new era of privacy protection. Noting that it could not "ignore the vital role that the public telephone has come to play in private communication," the Supreme Court, in *Katz v. United States* (1967), declined to follow the *Olmstead* majority's view that the Fourth Amendment be narrowly construed to apply only to the home, finding instead that it "protects people, not places."[10] It therefore could protect activity conducted in areas accessible to the public from government surveillance (in this case, requiring a warrant for wiretapping a public telephone booth). The near-ubiquitous adoption of technology over the better part of a century had at last dragged the law forward.

By the 1960s, the growth of the post-New Deal government combined with the post-war economic and population boom resulted in an explosion in the number of records kept about people in both the government and private sector.[11] Computerized record-keeping, which had begun as early as the 1890 U.S. Census using Herman Hollerith's mechanical tabulator, was not just convenient—it was becoming essential as a means of managing an ever-increasing volume of data.

As data proliferated, academics and activists became increasingly concerned with how this data was being managed—particularly since much of it was then in the hands of the U.S. government and there was little transparency as to how it was being used. In 1972, the Secretary of the Department of Health, Education, and Welfare (HEW)[12] established the Secretary's Advisory Committee on Automated Personal Data Systems. It was formed to address "growing concern about the harmful

consequences that may result from uncontrolled application of computer and telecommunications technology to the collection, storage, and use of data about individual citizens."[13] In assembling the Committee, Secretary Elliot Lee Richardson specifically cited technological innovation as the driver of this reassessment of privacy. "The use of automated data systems containing information about individuals is growing in both the public and private sectors…," he wrote. "The Department itself use many such systems…. At the same time, there is a growing concern that automated personal data systems present a serious potential for harmful consequences, including infringement of basic liberties."

In response, the Committee produced "Records, Computers and the Rights of Citizens," a (perhaps surprisingly) helpful government report that continues to influence privacy policy to this day. The report was submitted on June 25, 1973, the same day that John Dean, former White House Counsel, testified before the Senate Watergate Committee that President Richard Nixon was involved in the cover-up of the Watergate burglary. Allegations of governmental abuse of power pervaded the zeitgeist when the HEW Committee concluded that, "Under current law, a person's privacy is poorly protected against arbitrary or abusive record-keeping practices."

The Committee went on to propose a set of principles that should apply to the construction and use of automated personal data systems. These principles would eventually come to be known as the Fair Information Practice Principles (FIPPs), and they have been adopted around the world as the basic framework of information-privacy legislation and policy.

The FIPPs have been formulated in a variety of ways, and carry significant weight in the operational and technical frameworks of privacy. They are summarized in the following sidebar.

FAIR INFORMATION PRACTICE PRINCIPLES (FIPPS)

Collection limitation

Do not collect more information than you need.

Data quality

You have a responsibility not to collect, store, and use inaccurate data.

Purpose specification

Tell people why you want their data and get their permission to use it that way.

Use limitation

Before you try to use already-collected data for an unexpected new purpose, explain why and get permission from the appropriate people.

Security

Protect the data you hold.

Openness

Be as transparent as possible to the people who entrust their data to you.

Individual participation

People should be able to see what you know about them and ask you to correct mistakes.

Accountability

You are liable for responsibly handling information.

These principles are now enshrined in such diverse places as the U.S. Privacy Act of 1974, the European Union Data Protection Directive, the Australian Privacy Act's Information Privacy Principles, the Singaporean Personal Data Protection Bill, India's Information Technology Rules (formally, The Reasonable Security Practices and Procedures and Sensitive Personal Data or Information Rules), and a number of other national laws and policies that make up today's privacy landscape.[14]

Today

The HEW report and the legislation that flowed from it over the course of two decades represent arguably the last major watershed moment in informational privacy law development. Since then, the legal infrastructure has been strained to the breaking point as policymakers and judges struggle to apply decades-old law to technology that was barely imaginable when those laws were passed.

The U.S. Privacy Act, for one, has not been substantially amended since its initial enactment in 1974, forcing innovators in data processing technology to figure out how to fit sophisticated new data structures into the filing cabinet-record paradigm that characterizes the Act. The Electronic Communications Privacy Act (ECPA), meanwhile, which governs U.S. federal law enforcement's use of wiretaps, pen registers, trap-and-trace devices, and the interception of electronic communications such as email, was enacted in 1986—long before most Americans had even heard of the Internet, let alone adopted it as one of their primary modes of communication and commerce. Consequently, ECPA has created a set of confusing, inconsistently applied standards, yielding strange results.[15]

U.S. state privacy laws have fared somewhat better, with states creating context-specific privacy requirements for an assortment of data types (e.g., bank records, insurance, educational information). However, each state takes a different approach to privacy. When state laws conflict with federal laws, legislatures and courts are forced to engage in complex legal analysis to determine which system should take precedence. This often leads to confusing outcomes. Such a hodge-podge of privacy rules often leave multistate and multinational businesses scrambling for strategies to build one product or adopt one policy that meets the requirements of every state.

Meanwhile, the European Union's sometimes aggressive enforcement of assorted Member State "data protection" laws has led to stronger global privacy practices as multinational companies hoping to operate in Europe attempt to comply.[17] Yet even these laws are built on the foundation of the FIPPs, and are cracking under the strain of the new paradigm of contemporary data scale and complex analytics. The European Union has proposed an update to its data protection regime, which is discussed in more depth in Chapter 11.

Outside of legislatures, the courts have fared little better in trying to keep pace with technological development. In one of the more significant privacy decisions of the last twenty years, *Kyllo v. United States* (2001), the Supreme Court ruled that police would be required by the Fourth Amendment to obtain a search warrant in order to direct a thermal imaging device at a private residence. Acknowledging that "[i]t would be foolish to contend that the degree of privacy secured to citizens by the Fourth Amendment has been entirely unaffected by the advance of technology," the Court concluded that an unreasonable search has occurred because "here, the Government uses a device that is not in general public use, to explore details of the home that would previously have been unknowable without physical intrusion."

The Court's use of "not in general public use" could be read to suggest that the Court "deliberately adopted a rule that allows the outcome to change along with society," thereby trying to create a standard of privacy protection that adapts with the growth of technology.[18] But since there have been few follow-up cases along this line, it is hard to determine if the Court's rule was actually successful or if it just created more confusion without adding any real protection against intrusions on personal, informational privacy.

Lastly, the United States Federal Trade Commission (FTC) has taken on a lead role in protecting consumer privacy. It is worth noting that they have done so not under the auspices of any of the aforementioned privacy laws but rather pursuant to their authority under Section 5 of the Federal Trade Commission Act (15 USC 45), which prohibits "unfair or deceptive acts or practices in or affecting commerce." The FTC has used this authority to bring legal action against organizations that they argue have deceived consumers by failing to live up to their promises to handle consumers' personal information in a secure way.

The FTC has developed such a reputation that some scholars have claimed that "today FTC privacy jurisprudence is the broadest and most influential regulating force on information privacy in the United States—more so than nearly any privacy statute or common law tort."[19] However, the overall effectiveness of these actions in providing more consumer privacy protection must be measured in light of the fact that it is primarily dependent on the FTC policing the organizations' assertions about their own behavior. This means the level of privacy protection is driven not by government regulation

itself but by the organizations' decisions about the level of privacy protection they'd like to provide their own customers.

While law lurches along haphazardly, technology continues to leap forward. In 2013, the Pew Research Center's Internet & American Life Project reported that more than 90% of Americans own cellular telephones, and some suggested that the adoption of the smartphone was outpacing the sprea of any other technology in human history. The massive amount of transactional and geolocational da generated by these mobile devices contributes to the larger trend of an exponential growth in the amount of stored data in the world, which by one estimate reached around 1,200 exabytes in 2013.[20] Attempting to describe the state of the "big data" world in 2013, economists Viktor Mayer-Schönberger and Kenneth Cukier coined the term "*datafication*" to refer to "taking information abou all things under the sun—including ones we never used to think of as information at all, such as a person's location, the vibrations of an engine, or the stress on a bridge—and transforming it into a data format to make it quantified [allowing] us to use the information in new ways." (See Chapter 11 for more on datafication.) An entire industry of big data analytics has emerged to take advantage of these mountains of information, often developing techniques that can extract unexpected insights (sometimes relating to deeply personal subjects) from seemingly innocuous data.

These vast reservoirs of data—in particular, personal data about individual behavior—have not only been a boon to the commercial sector, they have also provided a treasure trove of information for governments. Police departments, intelligence services, and government agencies of all kinds have harnessed the power of data analytics to do everything from eliminating inefficiencies in housing-code violation investigations to anticipating crime outbreaks to capturing terrorists. Privacy and civil liberties advocates have long expressed concern at the extent to which some of this information is being collected and used by governments, but for the most part they could only speculate as to what was happening behind the veil of secrecy shrouding the clandestine services.

This all changed on June 5, 2013, when *The Guardian* revealed the bulk collection of telephony data by the U.S. National Security Agency on a scale that shocked many observers.[21] Four days after breaking the news, *The Guardian* introduced the world to Edward Snowden, a former NSA contractor who executed one of the largest intelligence leaks in U.S. history in order to reveal "the federation of secret law, unequal pardon and irresistible executive powers that rule the world."[22] The ongoing release of classified materials has triggered one of the largest public discussions about privacy, and one of the most significant reviews of U.S. intelligence activity, since the Church Committee investigated CIA and FBI domestic abuses in the 1970s.

Once more, the law is scrambling to catch up with new technological developments. A declassified opinion of the U.S. Foreign Intelligence Surveillance (FISA) Court, the body charged with judicial oversight of certain intelligence community activities, acknowledged as much when it found that Fourth Amendment protections did not apply to the collection of "non-content telephony metadata." also suggested that this conclusion (which relied on a 1979 Supreme Court decision) would do well t be revisited by the Supreme Court "in the context of twenty-first century communications technology." Other courts have reached similar conclusions, and a robust debate over these issues continues in courtrooms, classrooms, and legislative hearing rooms around the world. While it remains unclear how these issues will be resolved in the coming years, it is clear that technological development will continue to be one of the driving forces in shaping an individual's privacy rights.

“East Coast” Code and “West Coast” Code

Technologists may think themselves helpless in the face of legal developments, resigned to waiting for society to react to a new technology and adapt law and policy to the new technological paradigm. In reality, technologists may have as much influence on the development of the law as the law does c technology. Consequently, the technology described in this book should not be thought of as just a *reaction* to the requirements of law but also as a potential means of *shaping* the ultimate legal outcomes.

As history illustrates, the interaction of privacy law and technological innovation can seem like billiard balls on a table. Often they appear to be largely separate worlds that occasionally collide, sending one or both careening off in a new direction, each one affecting the other in different ways bu never merging. Inventors and engineers solder wires and write computer code, but their understandin of the law tends to be limited to the rules defining what they can and cannot do. Lawyers and policymakers, meanwhile, only become aware of new technology when it reaches a critical mass of usage in popular society, and they often spend years trying to understand how this new technology changes the world around them and then deciding how the law should (or should not) react to those changes.

However, we believe law and technology cannot and should not operate in separate worlds. Ideally they should work together, with technologists understanding and designing technology based on a solid grasp of relevant law and policy, and lawyers and policymakers understanding technological capabilities in order to better inform and even support their policy decisions. This concept derives from Harvard law professor Lawrence Lessig, who, in 1999, sought to explain the state of regulation in the nascent world of cyberspace:

> *“The single most significant change in the politics of cyberspace is the coming of age of this simple idea: The code is law. The architectures of cyberspace are as important as the law in defining and defeating the liberties of the Net. Activists concerned with defending liberty, privacy or access must watch the code coming from the Valley—call it West Coast Code—as much as the code coming from Congress—call it East Coast Code.”*

Lessig later clarified further: “The lesson of *code is law* is not the lesson that we should be regulating code, the lesson of *code is law* is to find the right mix between these modalities of regulation to achieve whatever regulatory objective a government might be seeking.”

The so-called “West Coast” code and “East Coast” code can interact in a variety of ways.[23] In some cases, “West Coast” code defines the physics of the world in which “East Coast” code can operate. The very design of devices and the networks that support them establishes the boundaries of the environment within which policymakers can operate. For example, the creation of biometric authentication technology allows policymakers to require the use of such capabilities to secure sensitive systems. In other cases, “East Coast” code directly limits what “West Coast” code can do. For example, cybercrime laws prohibit the creation of malicious code. It is the complex spectrum between these two extremes that generates the sizeable range of options available to the thoughtful, privacy-minded software engineer.

Consider the development of cellular phone capabilities. Back in ancient times, cell phones were relatively simple devices used to connect two people for a voice conversation. Today, they can conta

(and generate) substantial amounts of information touching almost every aspect of our lives. Cell phones can now store gigabytes of information in the form of documents, pictures, videos, and other types of files. They can also run various applications that allow them to access other troves of information such as server-based email accounts.

While useful and driven by consumer desire for such access, the storage of this data has led to some challenging new issues under U.S. Fourth Amendment "search and seizure" law, and the development of certain cell phone capabilities can have a profound effect on personal privacy and fundamental freedoms. The Fourth Amendment to the U.S. Constitution prohibits agents of the government from conducting "unreasonable searches and seizures" of "persons, houses, papers, and effects" without a judicially issued warrant based on a finding that there is "probable cause" to believe that evidence of crime or contraband will be found. There are several judicially created exceptions to this stricture, including one that has been interpreted to allow law enforcement officers to seize cell phones as part of a search incident to arrest and review the contents of those phones without obtaining a search warrant.

For a long time, courts were split over the validity of these searches. Some have suggested that because the phone is on the arrestee's person and may contain evidence, seizing the phone constitutes little more than reading the contents of a piece of paper found in the arrestee's pockets. Others have argued that the sheer volume of information available on the device changes the analysis, as law enforcement officers would normally only be able to obtain such extensive information via warrants that authorize the search of a computer hard drive or subpoenas requesting access to stored emails from a third-party email provider. Eventually in 2014, the Supreme Court, in *United States v. Riley* settled this question, finding a substantial distinction between the contents of one's pockets and the contents of one's cell phone:

> *"Modern cell phones, as a category, implicate privacy concerns far beyond those implicated by the search of a cigarette pack, a wallet, or a purse. A conclusion that inspecting the contents of an arrestee's pockets works no substantial additional intrusion on privacy beyond the arrest itself may make sense as applied to physical items, but any extension of that reasoning to digital data has to rest on its own bottom."*

Thus, we can see that "West Coast" decisions to create devices with substantial storage capacity has required "East Coast" counterparts to reconsider long-standing legal doctrines.

Another thorny issue surrounds geolocational data generated by cellular phones. Geolocational information can be generated any time a phone call is made, any time a text is sent, any time an application relies on geolocation data (e.g., an application providing information on vehicle traffic), and even any time a device passively "pings" a cellular tower as it moves in and out of coverage area. This information can be stored on the phone, with the cellular provider, and with the maker of the application, thus creating a potentially enormously valuable data source for law enforcement and intelligence agencies.

But this body of information exists *because* of "West Coast" decisions to design systems that generate and store it. "East Coast" law enforcement and intelligence policymakers then responded to the creation of this entirely new set of data by integrating it into their investigatory techniques. Ultimately, the public, courts, and policymakers are left to debate and decide if this is an appropriate

use of the data, and whether there should be legal restrictions on the use of this information both by the public and private sectors. In this case, the "West Coast" code created an entirely new source of information that fundamentally changed the relationship of the individual users to their devices (we now essentially carry tracking devices in our pockets) and it was done in a relative "East Coast" code vacuum, thereby creating a great deal of uncertainty regarding the power of the government and othe to track our every move.

These are just two examples that serve to illustrate the complexity of the technological and legal landscape, and in many ways, even these cases are overly simplified. "West Coast" and "East Coast" are hardly monoliths defined by a single motivation or goal. Instead, they are both composed of constantly shifting coalitions of interests, including individual coders motivated sometimes by profit and sometimes by altruism; businesses with substantial economic stakes in both legal and technical outcomes; policymakers torn between protecting privacy, preventing crime and threats to national security, and promoting economic growth in the tech sector; advocacy organizations looking to foste a free and independent cyber world while at the same time trying to curb the potential for nefarious exploitation of this world; and individual consumers eager to take advantage of useful and fun new technologies while anxiously trying to preserve a seemingly dwindling sphere of private life. Each of these interests can and often do shift from looking to either "West Coast" code or "East Coast" code address any given concern.

Why Privacy Is Important

The historical influence of technology on privacy law raises the question—did technological innovators have privacy in mind when they designed their products? When George Eastman introduced the Kodak camera, how much thought did he give to its ultimate effect on individual privacy? Did he imagine a world of candid, snapshot photography and wonder how it would affect, fo better or for worse, the photographer and the photographed? Did he hesitate for a moment before pulling away the cloth to unveil his invention? Did he consider ways to modify the technology to better protect privacy?

Perhaps the better question to consider is why Eastman, or any technological innovator, would even want to consider these questions in the first place. In today's society, at least, there are a number of potentially significant consequences—both practical and ethical—for businesses that fail to consider the privacy implications of their work.

On the practical side, innovators today face a complex web of privacy law at the state, federal, and international levels. Failure to comply with these laws can open the door to sizeable civil lawsuits, or substantial government fines. Here are just a few recent examples:

- In 2011, Facebook settled a class action lawsuit for $20 million for using the names and pictures o members in "Sponsored Stories" without their consent. Facebook has also agreed to aggressive oversight from the U.S. FTC that could lead to further fines if the company is found to share user information without proper notice and consent.
- Google settled with the FTC in 2012 for $22.5 million for bypassing the privacy settings of the Safari mobile browser. In addition, Google has been fined by a number of European data-protectio

authorities (and is under investigation by several others) for violation of privacy laws.

- Smaller businesses are not immune. In 2013, the makers of a social networking application called Path were fined $800,000 by the FTC for collecting personal information from children without parental consent.
- A four-employee smartphone application developer called W3 Innovations agreed to a $50,000 fi paid to the FTC for similar violations involving the collection and sharing of data from children.

Steep fines like these create incentives to build or buy products that can facilitate the privacy-protective practices demanded by regulators. But aside from financial penalties, companies might als be in the market for such products to help proactively assuage the concerns of a privacy-sensitive customer base. Any customer with sensitive data will likely prefer a product or a service provider tha can keep their information safe from theft or misuse, and otherwise handle data appropriately. Innovators could also favor privacy-protective products to circumvent any bad publicity that might doom a new product before it ever has a chance to flourish.

Government organizations and the businesses selling to them, will face similar pressures. Statutes, regulations, and policy can all require the implementation of complex data-handling procedures. Meanwhile, public opinion can sometimes demand the implementation of privacy-protective measur before data-driven programs can win broad support. The product designers who anticipate these considerations as they build their offerings will often have a business advantage over those who have not incorporated privacy-protective technologies into their core design.

Another practical consideration is the need to hire the best talent. Most companies will only be as good as their engineering talent, and many of those engineers will want to be challenged by their work. Engineers want to work at companies at the vanguard of their respective fields, and innovative data privacy solutions are part of what is considered the cutting edge—this alone may prove attractiv

But there is the ethical component to consider as well. Engineers working for a company that is regularly implicated in privacy violations or that sells its product to companies or countries that mig misuse that technology may not only potentially face the pricking of their own conscience but also th disapproval of their fellow engineers. This latter point should not be taken lightly. In robust online communities in which many play an active part, reputation is paramount. A company that dedicates itself to doing business in a way that enhances privacy protection at best, and at the very least does n harm to individual privacy, may have an easier time appealing to engineering talent.

Finally, technologists may wish to take steps to protect their users' privacy if for no reasons other tha (1) to acknowledge and respect the trust their customers place in them, and (2) to recognize that they too, must live in the same world that their products will shape, and will face the same harm as their fellow citizens would from inadequate privacy protections. Engineers should not divest themselves o responsibility for the societal consequences of the technology they create.

While there may be no absolute "right" answer in terms of how much privacy each of us should have and how that privacy should be preserved, we argue that it is unacceptable for engineers to take an agnostic view—either by choosing to ignore the effects of their technological designs or by simply remaining ill-informed as to the potential political, economic, and social effects of their products. Given their power as agents of change (a subject whose surface is merely scratched by this chapter),