

Volume II

PRIVACY

Applied | Proactive | Innovative

Brought to you by



PRIVACY AND SECURITY
INNOVATORS CIRCLE



Initiative by : Privacy & Security Innovator's Circle
(Join us on LinkedIn)

Publisher : LightBeam.ai Inc.

Editor : Jutta Williams

Cover & Diagrams : Niramayee Sarpotdar
Minal Sher

Contributors : Jonathan Wilde
Kimberly Lancaster
Steve Kaplan
Greg Vaisberg
Kalyan Chakravarthy
Mark Pilon
Kate Lapointe
Bill Schaumann
Priyadarshi Prasad

Visit us at : www.lightbeam.ai

Other Volumes : Read Volume I





Contents

Editor's Note	04
<i>Applied</i>	
Deletion: From Startup to Scale-up	06
How do you Scale, Automate, and Optimize your Privacy Operations?	16
Data Lifecycle - Good Data Governance is Good (for) Business!	22
<i>Proactive</i>	
The ROI of the New AI-Driven Privacy Platforms	29
Think Global, Act Local: Handling Data Privacy and Compliance in the Government Sector	34
Quebec Law 25: Progress for Privacy and a Time Crunch for Organizations	39
<i>Innovative</i>	
AI is the bomb	44
Does ChatGPT Know About You	50

Editor's Note

By Jutta Williams



Welcome to the Information Age! Alvin Toffler, noted futurist and author has called this the third societal wave; an age defined by our most valuable commodity -- information. The industrial revolution ended an age when agriculture was king. In return, the industrial age has yielded its reign to a new article of trade fueled by technological innovation and the commoditization of ideas.

The speed at which innovation in and development of this new commodity occurred – and continues to occur -- is unparalleled and has caused great consternation to those who would regulate and secure these new technological advances. The result is a largely untamed frontier, fraught with potential for great reward as well as great risk.

The benefits and value of this new information commodity are clear but each epoch has proven that costs related to these innovative goods are also novel and new. Pollution was largely unknown before the industrial age. Deforestation was unfathomed before agrarian societies flourished. And in this new age, we are only beginning to understand the

trade-offs that rise out of the surveillance society that fuels the information age.

We are now experiencing a new wave of technical innovation thanks to generative AI models and quantum computing that threaten our known methods for safeguarding people, their data, and our business interests. This, coupled with shrinking budgets and resource constraint makes it more important now than ever that we focus on applied, proactive, and innovative approaches to information protection.

Applied: The time to discuss risks and manage business decisions at a conference table or in the boardroom must make way for application of preventative controls. Innovation is accelerating and outpacing internal and external efforts to safeguard. In the early days of privacy there were more unknowns that required debate. As expectations have matured, our tasks have become more clear and we should put them in practice without delay because tomorrow there will be new fires to fight.

Proactive: Privacy and security teams are still largely considered cost centers and as such we are unlikely to grow bigger during a down

economy. We must look to improve velocity and scale for us to meet the growing demands of customers and adversaries. Investing in automated, preventative, repeatable and self-serve capabilities reduces manual effort and helps small teams also to be mighty.

Innovative: New and emerging risks are finding a home with privacy and security professionals, whether we can afford it or not. Dark patterns, children's safety in product design, algorithmic audits, rise of quantum, production use of generative models, and the inevitable failure of encryption are but a few. The data protection function has been called to take on new innovation topics by companies and

regulators alike because we navigated the ambiguous and subjectively interpreted field of data privacy and they hope we can do so again. And again.

Our books tackle these ideas by learning from one another and sharing ideas that have worked for the authors. In volume 2 we take a look at applied, proactive, and innovative approaches to issues you're likely reading about, encountering, or perhaps already managing.

We hope you'll find ideas that can help you translate between the business, engineers, and the public and that you'll enjoy volume two of Privacy API.

Jutta Williams is the co-founder of the Privacy & Security Innovators Circle (PSI Circle). She is a career data protection professional and currently the Head of Privacy & Assurance at Reddit.

Section : Applied

Chapter 1

Deletion: From Startup to Scaleup



By Jonathan Wilde

Jonathan is a Tech Lead Manager at Clubhouse, supporting privacy, security, and trust & safety engineering. He previously spent six years at Meta, where he founded and led their Privacy Control Infrastructure team, and led the deployment of a UN Human Rights Council recommendation for the 2017 Rohingya refugee crisis in Myanmar. He pursued his education at Tufts (BSCS) and HKU (exchange).

[View Linked In Profile](#)

We've all needed a delete button.

As consumers, we've sent a message with a typo, have files that are longer relevant, or signed up for a service that we don't use anymore. We search for the delete button. And when the button doesn't exist, there's too many steps to use it, or it doesn't work, we get frustrated.

There have been years of bad experiences around data deletion.

Facebook didn't fully delete content and spent years fighting press backlash. The New York Times made it painful to close accounts, resulting in tougher Federal Trade Commission rules for the entire industry. And the Austrian Post didn't let users delete accounts via email, and was fined over \$10M for GDPR violations.

These have resulted in higher standards of every company.

Apple won't let you into the App Store if your app has sign-up but no delete button. Laws in the European Union, Brazil, Argentina, China, India, and South Korea — along with California, Colorado, Connecticut, Virginia, and Utah in the United States mandate a "right to be forgotten". More countries and dozens of other U.S. states are considering their own version of this.

It's now critical to have a data deletion strategy at launch.

● Designing your strategy

The good news is that data deletion strategies don't need to be complex or change constantly. In many cases, you can design your strategy once, and have it support your organization for years.

Key components

There's three key components of any deletion strategy:

Interfaces : Places where a user will request deletion, such as a button in-product, or through a support contact like a form, email, or phone number.

Methods : This encompasses:

- The strategy for the deletion which can range from object deletion, where you remove all copies of the data, to pseudonymization, where you delete identifying information so you can't link it to a person.
- The SLA you keep — how fast you execute the deletion by.

Exceptions : Situations where you prevent deletion to meet competing legal obligations or user expectations.

Let's dive into each of these components in detail.

● Interfaces

For laws like GDPR and CCPA, the starting point is frequently honoring the "right to be forgotten". This means providing a way to delete all personal data about somebody. This could start with ensuring your customer service contacts

can handle deletion requests. And as your volume of deletion requests increase, this could expand to offering fully automated account deletion options in-app or on your website. You'll want to work closely with your legal counsel around the design of this as different laws have differing requirements.

For example, in California, CCPA expects you to designate [two distinct methods](#) (e.g. form, email, phone number). And in Europe, GDPR is less specific — if a valid deletion request reaches your organization [verbally or in writing, from anywhere](#), you may need to process it.

If you're a SaaS platform, you'll need to consider how to help your customers honor the "right to be forgotten" for their own customers and employees.

For example, if you're building an analytics SaaS service that collects logs that are personal data, your customers may want to delete or pseudonymize the logs for "right to be forgotten requests". Emailing you for every user every time this happens may be an operational deal breaker for customers, so SaaS companies frequently provide APIs to automate this kind of situation. Amplitude's [User Privacy API](#) is a great example of this. These APIs have historically been bespoke, but standards like [OpenDSR](#) are making it easier to wrangle SaaS providers, and are worth considering.

And beyond this, your customers may expect granular deletion interfaces.

These are places where they can delete data without deleting their whole account (e.g.

"Delete Document" in Google Docs). Work closely with your legal team to understand which are discretionary features to offer a good experience, versus a legal expectation, so you offer the right ones, for the right platforms, at the right times.

● Methods

Once you have your interfaces picked out, you'll need to decide how to delete them. At the end of deletion, we want to ensure somebody's personal data is either ideally fully deleted, or if needed for historical analytics, de-identified such that it can't be linked back to an individual. There's a range different approaches used in practice for this:

- **Object Deletion** : You delete a full unit of data.
- **Field Deletion** : You only delete sensitive or identifiable fields from a unit of data. Some engineers call this "anonymization".
- **Pseudonymization** : You overwrite sensitive or identifiable fields or columns with new data (e.g. overwrite user IDs with a "deleted user ID".)

It's tempting as a privacy engineer to push for object deletion everywhere.

But it's unlikely that they will fly with surrounding teams. Businesses need to track historical metrics, and if you used object deletion everywhere, you'd start to introduce skew into business analytics. As a result, a common pattern in industry is to use object deletion for online data as much as possible, and field deletion or pseudonymization for offline analytics

data sets.

Another key issue is the *SLA for deletion*.

Work closely with your legal counsel and product stakeholders to define.

- **Legal :** Different jurisdictions come with different timelines for honoring the “right to be forgotten”, frequently 30 days for GDPR, or 45 days for CCPA.
- **Product :** If users need the ability to “undo” a data deletion, it’s common to use tombstoning to hide data during a grace period (i.e. a recycle bin), after which data is fully deleted.

As with any privacy strategy, this will be contextual. You will likely need to tune the approach and SLA for different data stores and interfaces.

● Exceptions

Most products have exceptions where they can’t delete data — they’d struggle to maintain security against bots, struggle to protect the company against lawsuits, or be unable to comply with anti-money laundering laws requiring [retention of specific customer data for 5+ years](#).

The good news is that privacy laws like CCPA and GDPR have carve-outs for this.

For example, CCPA has exceptions to data deletion requirements for tasks like completing transactions, protecting security, debugging, and compliance with other legal obligations. Similarly, GDPR has exceptions to data deletion requirements for compliance with other legal obligations, and [other tasks](#).

It’s an area to tread thoughtfully.

Even if you decide with your legal team that you can retain data in your situation, you may need to explore strategies like data separation or encryption to limit the purposes of processing that data.

● How to design your strategy

Any privacy strategy is best designed collaboratively.

As you think about your deletion plan, you’ll need to solve for three sets of problems and stakeholders:

- **Meet user expectations :** Users expect deletion to work differently in different product categories.
- **Meet legal obligations :** Different laws may pull you in different directions — some to delete more data, and others to delete less data.
- **Meet business needs :** Keep data needed to support analytics, trust & safety, and other business stakeholders.



● Meet user expectations

Who to partner with: Product, Design

Use the standard design toolkit to make your deletion experiences useful. Consider performing competitive analysis to understand product category norms, testing prototypes with users, and iterating until you have the right granular deletion interfaces, as well as an interface to delete all data. After launching, explore feedback from support teams for common complaints.

If you're building a B2B SaaS product, work together with your sales teams to make sure that you have the right interfaces for the companies you're targeting. Are you targeting customers in the US that need to restrict deletion for Sarbanes-Oxley, or need features that meet HIPAA's high bar of deletion? A well-designed deletion strategy can make it easier to sell your product.

● Meet legal obligations

Who to partner with: Legal, Policy

Use your data map to walk through different data categories you have with your legal counsel, and understand how they feel about each:

User Data : Prefer object deletion, but may be ok with field deletion in limited circumstances.

Analytics Events : Prefer object deletion, but ok with pseudonymization + field deletion.

Security Records : Need an exception. Aim to get a range of options that legal is comfortable with, so you can navigate conflicts with business stakeholders more easily.

● Meet business needs

Who to partner with: Sales, Analytics, Trust & Safety, Finance

Use your data map to walk through different processing reasons, and spot departments that may rely on data today, or may need the future. A few common cases to consider:

- **Analytics** : Need historical data to understand trends.
- **Security** : Need historical data to understand prevent bad actors from circumventing protections through deletion.
- **Finance** : Need long-term sales and revenue data for business analytics, and sufficient detail to comply with tax and anti-money laundering laws.

The better you can predict these issues, the more resilient the deletion plan.

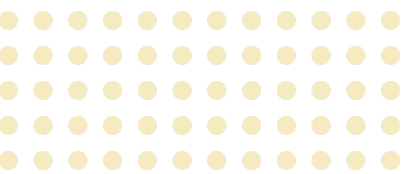
● A few examples



US-only financial institution

Privacy laws in the US frequently exempt financial products, and users need to be able to access information after account closure for taxes.

- **Interfaces** :
“Right to be forgotten”: Clear explanation to the user that this will not result in all of their data being deleted.



- **Methods :**
 - **Accounts :** Object deletion.
 - **Product recommendations :** Object deletion.
 - **Analytics :** Pseudonymization + field deletion.
- **Exceptions :**
 - **Financial records** (e.g. accounts, statements, tax forms), kept separately from main account storage to avoid accidental processing.
 - **Security audit logs** (e.g. password change attempts).



Consumer photo-sharing app

This kind of consumer product falls squarely in the focus of CCPA and GDPR, and carries the strongest consumer expectations around deletion.

- **Interfaces :**
 - “**Right to be forgotten**” : Ability to delete account and photos, with a holding period where users can recover their account.
 - **Granular deletion** : Ability to delete photos and albums in app.
- **Methods :**
 - **Accounts** : Object deletion.
 - **Albums** : Object deletion if owner, field deletion if contributor.
 - **Analytics** : Pseudonymization + field deletion.

- **Exceptions :**
 - **Image hash banks** (i.e. for detecting child sexual assault material).
 - **Security audit logs** (e.g. password change attempts).



B2B SaaS document service

These services carry all of the interplays between customers and their employees and end-users.

- **Interfaces :**
 - “**Right to be forgotten**” : Ability to delete their entire instance.
 - **Customer & employee “right to be forgotten”** : APIs to locate and redact documents based on customer or employee information.
 - **Granular deletion** : Documents can be sent to “recycle bin”.
- **Methods :**
 - **Documents**: Object deletion.
 - **Folders**: Object deletion.
- **Exceptions :**
 - Documents and folders under **active legal hold**.
 - **Security audit logs** (e.g. users added and removed, password change attempts).

● Implementing your strategy

While the strategy may remain unchanged, the implementation of data deletion will change drastically over the course of your company's growth.



Solopreneur

Problem: You need to prepare for your first deletion request.

When you're getting set up as a solopreneur, you can expect deletions to be relatively rare — maybe just a few a month. And the “right to be forgotten” might [not apply in every jurisdiction](#) you're operating in yet.

But you'll likely want some initial interface to make customers happy.

Interface

If you have a support@example.com email address for customer service, you can use this as your first deletion interface, and mention it in your privacy policy.

Your steps for processing requests might look like this:

- Verify ownership of the account in question.
- Run the deletion, if verified.
- Respond to the user that you deleted, or that you couldn't delete (e.g. you can't verify the own the account).

Make sure you're designating sufficient contact points for you as you get larger, and fall under laws like CCPA that expect you to have multiple.

Method

Once you've defined your customer service process, it may be best to start with a manual runbook approach rather than automating:

- Map out the services where you store user data.
- Write down the steps to delete a user at each service according to your deletion strategy.
- Execute deletions manually from the service console, one by one.

Easy to stand up, but definitely not scalable.



50 person startup

Problem: You're entering the App Store and/or getting too many deletions.

Eventually, the manual runbook approach will start to fall apart — there'll be too many deletions to process, and it'll be easy to make mistakes.

Interface

A more automated interface could sit in a few different places depending on your product sector and interface:

- **If you're adding sign-up for the App Store**, you'll need to find a place for an account deletion button. Most apps put this in user settings.
- **If you're building a SaaS product**, you have more options like setting up a privacy portal. There's a plethora of tools such as [Transcend](#) that will generate a portal where a user can delete their account.

Whichever option you pick, you'll need to work with legal counsel to sure you have the same concerns covered as before — that you verify the account owner before deleting, and have sufficient contact points for your jurisdiction.

Method

You'll need to convert your runbook's checklist into a function in your code.

Checklist items might become:

- **API calls** : Where you ask external services to delete data.
- **Database deletes** : If you're using a SQL database, you can setup cascades to make this simple. For NoSQL databases, you may need to setup indexes and loops to iterate and delete user data.
- **Emails or tasks** : If you can't automate the full checklist, you may need to assign tasks to yourself to complete the deletion, or send emails to service providers to complete the deletion for you.

As you architect this, the key is to make it handle errors well:

- **Idempotency** : It should be safe to call the delete multiple times in a row.
- **Retries** : Automatically retry the delete some number of times.
- **Error logging** : If the retries fail, make sure you're getting alerts, such as through a tool like Sentry or DataDog.

With this in place, you can write docs explaining to your team how to maintain, along with test coverage to ensure that new services are added to the function.



5000+ person company

Problem: You need to maintain reliability and coverage as you scale.

As you get far larger, your interface will likely stay the same, but maintaining your deletion method will become far harder.

Ben Strahs gave a [fantastic presentation](#) on challenges with maintaining deletion methods at Facebook. It's worth a watch, covering:

- **Keeping SLAs** : Deletions take a long time if you have a millions of likes on a post — too long to do synchronously when a user clicks “delete”. Facebook used tombstoning to immediately hide content from the product when deleted, and then moved deletion to background jobs, with infinite retries.
- **Under-Deletion** : Developers might forget to add data to delete to their delete method. They developed detection to spot rows that mention deleted data, which suggests some amount of under-deletion.
- **Over-Deletion** : Developers might accidentally write a deletion that cascades incorrectly. They developed a series of heuristics (e.g. don't delete a user from a non-user deletion).
- **Developer Efficiency** : Managing manual deletion methods is time-consuming for developers — so they moved this to the schema level.
- **Backups** : Managing backups for mistakes around deletion was painful, and it's important to make sure backups don't live forever. So they developed journals of deletions to make it easier to work with backups.

While Strahs' presentation talks about the specifics of a social network, all of the ideas map cleanly onto most systems, and many are implemented by open source frameworks like [Django Models](#) and [Prisma](#).



○ Key Takeaways

There's no one-size-fits-all approach to deletion. It depends on your product, industry, and customer base:

- **Put your customers at the center** : Make sure to match customer expectations for your interfaces, methods, and exceptions.
- **Strategize collaboratively** : Deletion will always be nuanced with a push and pull with competing needs — work together with surrounding stakeholders to make sure you're picking the right approach.
- **Scale with your company** : Solopreneurs might start with a manual approach. As you scale, automate, and develop technical safeguards to enable developers to work productively and safely.

With the right deletion strategy in place, you can set your organization up for years of customer success.



Section : Applied

Chapter 2

How do you Scale, Automate, and Optimize your Privacy Operations?



By Kimberly Lancaster

Kimberly Lancaster is a senior leader and privacy advisor who directs data protection and drives operational excellence by aligning InfoSec, Security, Compliance, Data Privacy, GRC & Risk as a unified platform. Kimberly integrates policies, standards, and processes to improve global product development and operations teams. She is active in organizations such as XRSI-Privacy, The Rise of Privacy Tech, Women in Privacy & Security, the Cloud Security Alliance, IAPP, OneTrust-Privacy Connect, and the Institute of Operational Privacy Design. Kimberly says, "Privacy is not just a passion, but a way to connect people, processes, and tools to provides awareness and methods to protect themselves."

[View LinkedIn Profile](#)

The prospect of scaling or optimizing your privacy and operations program can seem daunting at the outset. Common problems are limited resources, where the staff responsible for this are already over-tasked, and the challenge of getting necessary additional funding (including any tooling to support). Frequently, though, the largest problem is a lack of understanding by management and staff that privacy is a requirement for the company. Despite these common obstacles, the benefits of having a functional and scalable privacy and operations program range from improving customer trust to lowering privacy complaints and, finally, to increased operational efficiencies. [A good privacy and operations program contributes to a company's overall success.](#)

There are four steps to plan and implement a privacy and operations program:

- Define the program's goals and requirements.
- Develop the privacy roadmap and strategy.
- Complete a privacy review analysis.
- Implement, operationalize, and mature the program.

● **Defining the program's goals and requirements**

The first step is to define the program's goals and requirements and determine how they align to the focus of the company and its vision:

1. Create the program's organizational vision and gain executive support.
2. Ask questions to understand the business goals.
3. Set the scope of the program.

4. Align the privacy and operations requirements with the company goals and requirements based on the type of business.
5. Understand and document the company's regulatory, contractual, and internal requirements.
6. Establish timelines for the program.

● **Developing the privacy roadmap and operational strategy**

After defining the program's goals and requirements, the next step is to develop the privacy roadmap and the operational strategy.

1. Define roadmap-based risk levels and easy opportunities for implementation.
2. Share results of analysis/data protection impact assessment (DPIA) with management/exec team for awareness and support on the roadmap.
3. Align with Compliance and Security for support for the roadmap.
4. Implement the program framework (policies, standards, and guidelines).
5. Communicate!! Privacy brings value to the business; show how in discussions.

● **Complete a privacy review analysis**

With the goals and requirements in hand along with the privacy roadmap, you can complete a privacy review analysis.

1. Document which personal data is processed by the company. Create a record of processing activity (RoPA).
2. Define the highest risk areas to focus on first

using a risk-based approach. Focus on key areas where personal data is processed frequently, such as customer support.

3. Meet with the leads across the organization to understand their current processes, including data flows and mapping.
4. Complete a DPIA.
5. Review third-party vendor usage when it is involved with personal data.
6. Identify what is in place now operationally and what needs to be updated.
7. Identify if there is any automation available for processes.



● Put the plan into action

01. Define the data subject access request (DSAR) processes, even if they are manual.
02. Write or help draft the company's external privacy statement and internal privacy policy. (Remember, employees are just as important as customers when it comes to privacy.)
03. Establish an awareness program for privacy: how to contact, when to reach out, and why the team is there.
04. Annual training: Review existing privacy training for company employees. Consider creating/buying annual privacy training programs as necessary.
05. Privacy controls: Establish partnerships with the Security, Compliance, Risk, and Audit teams to align control frameworks and requirements.
06. Marketing: Review marketing processes and forms for consent language.
07. Cookies: Complete a cookie audit of the company's web pages, including both a cookie banner and a consent review.
08. Awareness: Establish a Privacy Champ team to support the work efforts and build a program where people in the business are empowered to learn and support privacy in their daily work-streams.
09. Automation: Use privacy enhancing technology (PET) wherever and whenever possible.
10. Guide teams through supporting the implementation of privacy-by-design into products, services, and processes.

● Next Steps

The preceding steps showed you how to plan for and implement a solid baseline program that meets the company's needs and that can be scaled for growth, new opportunities, and organizational maturity. The following are some options to scale the privacy and operations program to enhance privacy maturity in any size organization.

● Automating DSAR requests

1. Use an email or ticketing system so both internal and external DSAR requests can be logged automatically with an initial response provided.
2. Define processes via a ticketing tool or tracking system to move DSAR through intake/review/resolve process with confirmation of each step completed. (Jira, PET tool, or other options can help with automation as needed and limit exposure of personal data.)
3. Enable proper retention and record keeping of all DSARs.

● Establish records of processing activity (RoPA) automation.

1. Enable identification of processing through a centralized tracking system (Spreadsheet, Online form, or ticketing tool) where it can be access by the privacy Team on regular basis yet updated by process owner with ease.

● Data classification automation.

1. Data identification in processing in both the development and operational processes and in the RoPA or other

secure tracking mechanism provides a lower-scale approach.

2. Automated scanning of the data via a PET or self-built application allows for human review and validation of accuracy of the data.

● Privacy-by-design.

1. Define partners in Engineering and Security to start awareness and training on how to implement privacy-by-design in design, development, and operationalization of processing personal data
2. Document procedures where personal data is processed via code or automated workflow to define options to ensure proper handling. (It is possible to automate documentation using development tools.)

● Privacy Champions.

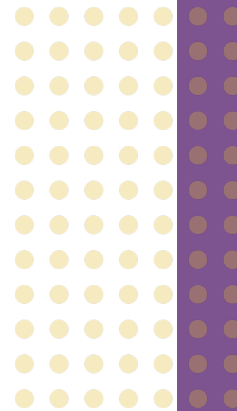
1. Enable a Privacy Champion team in your organization. Through training you may empower them to be the Privacy contact in their own teams to teach, understand, and support on how to implement privacy operations and development properly.
2. The Privacy Champion team enriches not only the agility and scale of privacy operations; it also offers rewards to those


who participate such as additional skills and opportunities to gain experience.

● Privacy enhancing tech (PET)

1. There are a variety of PET companies and tooling that can support a Privacy team to improve their automation and increase the success of their operations and growth. Complete a thorough analysis of what the needs are up front, what is critical to solve for first, and what resources are required to move forward.
2. Define a PET that aligns to your specific areas of focus and go through a defined vendor management process to identify, review, and purchase a PET solution that provides the tooling for automation. This information, along with a solid plan for implementation, maintenance, and growth of the tool(s) will ensure your success. (Understanding available funding, IT support and impact while implementation is going on is also essential to your success.)

Note: There are several great options available. Take the time to review and confirm that the PET you select meets the needs of the company and your privacy operations strategy and program.





As this article shows, the initial implementation of a privacy and operations program is usually a large job that takes careful planning, especially if the available staff are splitting time with other work. However, there are many resources to support this work, including:

- Free or low-cost templates for documentation
- Information on regulations
- Detailed privacy blogs
- Active community groups
- For-cost privacy advisory consultants

In addition, once the initial implementation phase of the program and operations work is in place, there is always ongoing maintenance and expansions to the program. Having a solid strategy, a clear set of goals, and continuing communication with your executive management team will let you take on the scaling, automation, and growth the company needs to continue to be successful. Scaling allows additional work to be taken on with success. Automation allows for seamless alignment between teams and customers. Maturity allows for a higher selling point with prospects and renewing customers.

The return on investment (ROI) of having a scalable and automated privacy and operations program far outweighs the initial impact. [The ROI is not just in data protection and the prevention of data loss, but also in data awareness and the ability to use data with proper consent to drive innovation.](#) In addition, knowing where the customer's data is located also drives cost saving in both physical and cloud storage and processing.

[Customers are also requiring more protection and, as a company grows, so do the risks around data handling.](#) The privacy program is a solid basis for providing proper information, training, processes, and awareness to all employees while partnering with Security and Compliance to create an environment that allows for and actively supports company growth. With an initial investment in planning, you can accomplish all these goals with a minimum number of individuals in a relatively short period of time.

Section : Applied

Chapter 3

Data Lifecycle - Good Data Governance is Good (for) Business!



By Greg Vaisberg

Greg leads the legal team responsible for supporting all aspects of product development, privacy, security and business development initiatives at Nutanix, Inc. A strategic senior legal advisor, Greg delivers business-focused solutions grounded in a genuine passion for technology and a track record of successfully enabling nearly 20B in partnership, licensing and M&A transactions. A pragmatic tech lawyer who brings a deep understanding of business considerations to practice through experience in corporate development & strategy. Greg holds a MBA with concentrations in Finance and Strategic Management from University of Pennsylvania's Wharton School, a JD with a Certificate in High Tech Corporate Transactions from Santa Clara University School of Law, and a BS in International Business from the University of San Francisco.

[View LinkedIn Profile](#)

Ever misplace your keys? Keeping track of things is difficult, even when they are physical objects. Supply chain managers know that a good custodial record is imperative. Laptops are tagged by IT teams and tracked through their lifecycle up to and including disposal. Your blood sample's journey is meticulously traceable. Yet, when it comes to our digital assets, the traceability of information, any information, enters the realm of a far less understood metaphysical dimension. On the surface, the framework for information governance is neither novel nor complex. However, if we as an industry are to rise to the challenge of bringing true end-to-end visibility to information, we will need to approach it with the sophistication and fervor that treats all information as the lifeblood that it is. Doing so is not only imperative for compliance, but to good decision making, and ultimately to earning and retaining customer trust.

● Painfully Simple

The information lifecycle is uninterestingly straight forward. I use the term information because the model applies to every piece of data, whether it is personal data, financial, confidential information, or source code, the journey is fundamentally the same:

1. **Origin:** Where does it come from?
2. **Rights:** What rights do you have in it?
3. **Destination:** Where does it end up?
4. **Access & Security:** Who has access to it and how do you protect it?
5. **Use:** What do those with access do with it?
6. **Quality:** How do we keep it fit for purpose?
7. **Term:** How long should you keep it?

However, the myriad of sources, storage locations, use cases, and parties makes implementing this model no simpler than tracking all the parts of a Boeing 787.





Origin

Every piece of data has its genesis, which in a good

governance model, must be clearly known and recorded. Information one creates, such as source code, doesn't present a challenge on this dimension. After all, generally the ownership is with the author. But, what about that sales lead or a candidate? The RFP with information about system requirements from your customer? Or details about an opportunity from a channel partner?

In each instance, the information must first be tranced into ownership and type. Do you or a third party own or control the information and what type of information is it (public, confidential, or personal data)? If it's personal data you are more of custodian than owner - the data belongs to others and you owe duties to these underlying data subjects.

Having a clear record of where the information comes from will enable the critical next step imperative for all information - understanding how the rights to such information are obtained , or, in contrast, what obligations are owed.



Rights

With the knowledge that you have a piece of data, you must now

identify what rights and obligations you have with respect to such information. In the case of customer confidential information, one will generally be restricted to a stated purpose (such as the furtherance of the relationship between

the customer and the vendor) and be granted limited rights to use the information restricted by notions like observing security best practices, following specific negotiated protocols, or simply observing non-disclosure and use on a need-to-know basis. However, when personal data is involved, one will have to explain use to the underlying data subjects and may also need to establish the lawful basis for processing such data.

Practically speaking, this then means that in most cases of sales leads or candidates you must carefully provide notices, track consent and opt-ins and where such leads are not obtained directly, you must ensure that appropriate appropriate transparency and likely consents for 3rd party activities are obtained by those providing the leads (e.g. your partners) and you have appropriate contractual protections for the transfer and processing of such data keeping in mind that consents are obtained for specific purposes which must be observed and such consent may be withdrawn at any time. This is basic accountability.

For customer and employee personal data you will likely rely on other lawful basis, such as legitimate interest or performance of a contract, and will need to conduct and document an appropriate assessment to assert such a basis.



Destination

Once you arrive at this third step, you start to appreciate the

metaphysics. Is the data stored in your CRM, HCM, or billing systems? Are those systems on

prem or in the cloud? Do you control the cloud deployment or use a SaaS solution relying on the services of others? Where are those physical assets located? Where are they backed-up and accessed from? And, what systems are connected to them? Is the CRM feeding marketing and email systems? Are you sharing that data with third parties? Will an action in the master system propagate to the others? Are your teams spinning up bespoke environments with the data on-prem or in cloud buckets?

While it is of course imperative to know where your and your customers confidential information is at all times, as it regards personal data, the GDPR records of processing require that you know and document where the data is stored and processed as well as delineate (or justify) any data exports, which includes someone accessing data in a SaaS environment from another jurisdiction. This one is especially important for responding to security issues and data subject access requests.

Put simply, if you don't know where the data is, you don't know what's been compromised or what data you hold about a data subject and how to delete it.

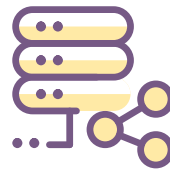


Access & Security

As mentioned, access by an individual outside of, for example,

a GDPR jurisdiction, is deemed a data export even if it means merely accessing a record in a SaaS tool like CRM or HCM web applications. A clear rule based access system that assigns and segments who has access to what in an

organization is imperative. If following this model thoroughly, the roles may be assigned on geographies, data types, or data use cases, just to name a few. But, it must be clearly understood and then documented who has access to what data for both security and compliance purposes. Additionally, you should develop, review and document the security processes and systems you employ to protect the data, including the creation of detailed technical and organizational measures.



Use

While this one seems simple, it is also one requiring careful

documentation in the records of processing when it comes to personal data and thus one that truly benefits from careful organizational taxonomy. Needless to say that candidate data sets in HCM should not be migrated into one's CRM and marketed to, and each will generally be subject to the narrow consent associated with it. On the customer side, the legitimate interest assessment or other lawful basis one may have, are not all encompassing and thus the uses of data, such as processing activities for personal data, must be clearly identified and adhered to in light of the rights the organization has vis-a-vis such data.



Quality

Stale data serves no one. At best, you'll be making bad decisions.

At worst, you'll find yourself out of compliance and face liability in your use of customer information and personal data. Implementing

a protocol for ensuring the data in your possession remains up to date and accurate is key. This may include regularly asking customers to confirm or update their contact and billing information, or partners confirming that all rights to provide such leads remain valid. This all ties back to notice, with personal data, the use purposes disclosed at data collection ultimately restrict the future use cases. Can you track and trace what was communicated to the data subject at the time those data were collected?



Term

While keeping data (on-prem or cloud) is certainly expensive in storage costs alone, the exposure to liability also abounds. Thus, it is important to implement a

clear framework for how long you need and can keep data. Each type of data will be governed by often numerous regulatory requirements. Privacy employment, securities, and tax laws just to name a few must be balanced and reconciled. . With so many systems in use and intertwined across organizations, relying on manual disposal of data is naive and fraught with risk. Implementing features to automatically erase data securely once the set retention period expires should be complemented by a system of fail safes to ensure that there is human review of data that is scheduled for upcoming erasure in order to confirm that no new grounds for further retention, such as a legal proceeding, pose lawful grounds or requirements for extending the retention period.

● Recommendations

The operational and organization realities of every business differ widely. However, by applying the framework, I believe businesses can not only drive compliance, but develop a source of competitive advantage - garnering and maintaining customer trust. The graphic below suggest high level questions to ask and processes / items to consider in each pillar of the framework



Origin

Where does the information come from?

- Leads (Web properties, partners, lead gen. services, tradeshow)
- Customers
- Partners
- Candidates
- Employees
- Vendors



Rights

What rights do you have to use the data and how do you obtain them?

- Determine and document lawful basis.
- Enter requisite contracts / DPAs
- Record and track of consent
- Include data received from third parties.



Destination

Where is the data stored?

- Geographic location (mind the cloud data center regions)
- Own infrastructure vs. third party / cloud (e.g. sub-processor)
- carefully map any movement of all data systems / locations.



Access & Security

Who has access to each piece / category / use case of data? How do you protect it?

- Do only those with a need have permissions?
- Do parties / systems outside of the storage geo have access?
- Do third party partners / vendors?
- Technical and organisational measures



Use

What are you using / can you use the data for?

- Are your uses in line with your rights?
- Marketing
- Sales
- Billing
- Support
- Payroll
- Benefits
- Product



Quality

How are you keeping the current data?

- Customer contact information
- Vendor contacts
- Continued rights for partner data



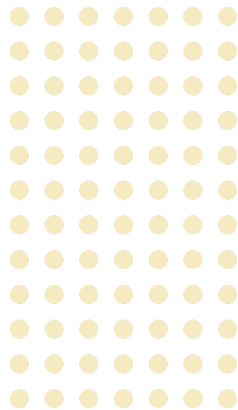
Term

How long can you keep the data and how will you decommission it?

- Identify and assign terms by data type, rights, and use data at ingest.
- Implementing recurring checks during pre-erasure.

● Conclusion

If a lot of this framework seems obvious, it is so by design. Much of what is required for good practices and compliance is fortuitously intuitive. And, if the framework looks a lot like a good record of processing, it is also so by design. A data lifecycle model thoughtfully implemented should ultimately lead to records of processing that are not only required for compliance, but also yield rich substantive materials for attaining and retaining customers. Great records of processing will enable you to demonstrate to your customers, partners, employees and other constituents that your organization is not only compliant, but is deserving of the enduring trust that underlies successful, lasting relationships. Good data governance is good (for) business!



Section : Proactive

Chapter 4



The ROI of the New AI-Driven Privacy Platforms



By Steve Kaplan

Steve is known on Twitter as @ROI dude, and has built his career in IT by helping enterprises use financial analysis to optimize IT strategies. Steve co-founded, ran, and sold two channel partner businesses - the first one was named the first Citrix Partner of the Year, and the second won the VMware Global Rainmaker Award.

He also started the Americas Channel Partner program at Nutanix in early 2013 when it was still a small startup. CRN Magazine named him one of The Top 25 Channel Sales Leaders before he later started what would become the global Cloud Economists team.

In addition to co-authoring 8 books on Citrix and VMware technologies over the years, he authored, *The ROI Story: A Guide for IT Leaders*.

Steve created the VirtualMan comic book series, and will soon be publicly introducing [The Tessellator](#) and her sidekick, Auto-Mask.

Steve received his undergraduate degree in Marketing from UC Berkeley, and his MBA in both Marketing and Finance from the Kellogg School of Management at Northwestern University.

[View LinkedIn Profile](#)

Data privacy is still a relatively small part of the massive cyber security industry who, for the past 20 years, have been constrained to using pattern-matching as a proxy for really understanding the who, what, and where of sensitive data. Consequently, even large investments in privacy solutions tended not to yield great returns. But new AI and Natural Language Processing (NLP) capabilities are allowing privacy-focused security platforms to now enable phenomenal economic benefits.

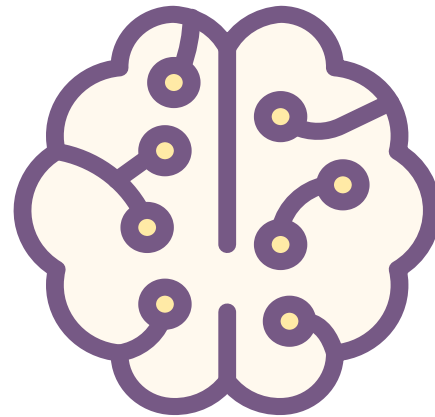
● **ROI of existing privacy solutions**

There are many thousands of cybersecurity companies, many of whom claim to protect sensitive and private information. But they attempt to go about it by building large firewalls to encircle the data using pattern matching to, say, identify a 9-digit number as a proxy for a social security number. The problem is that pattern matching quickly tends to result in a barrage of false positives.

The cybersecurity companies of the past 20 years don't understand what the data is, where it is exactly, whose data it is, and what data is leaking out. The result is continued frequent large data breaches even of large and well-funded organizations.

Cisco's 2022 [Data Privacy Benchmark](#) Study drew upon data from 5,300 security professionals companies across 27 countries and also incorporated data from its [2021 Consumer Privacy Study](#) of 2,600 adults across 12 countries. Based upon an average privacy

budget of \$2.7M, Cisco estimates \$4.86M in benefits for an **ROI of 80%** (presumably on an annual basis).



● **New AI-driven privacy platforms**

In the last 2 years, good open source Machine Learning models have come out allowing a new breed of cybersecurity companies to leverage these models and look into the data itself. They furthermore utilize Natural Language Processing to look at images of items such as drivers licenses and not only discern important sensitive information, but then to automatically redact that information from the image.

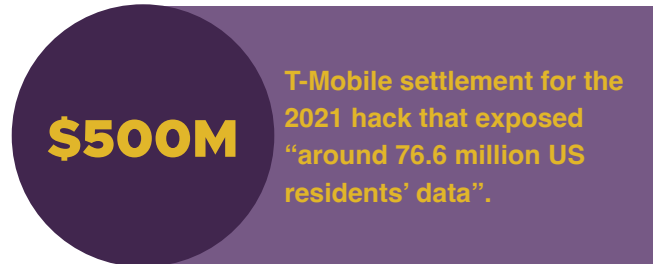
The T-Mobile breach along with new capabilities in AI/NLP (Natural Language Processing) privacy solutions provide an opportunity to calculate the ROI of a modern AI-driven privacy platform and compare it to the Cisco study results.

● Returns

Let's consider a B2C company with **100,000 customers** generating a **profit of \$500 per customer, or \$50M per year.**

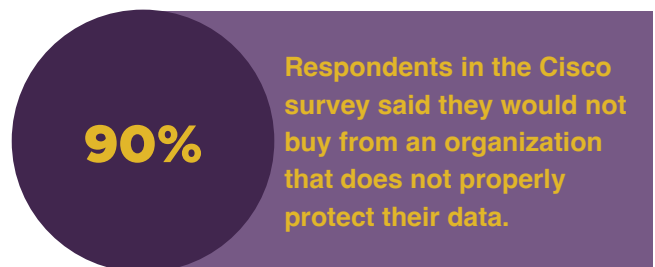
Fines

A July 22, 2022 The Verge article says that T-Mobile is paying a \$500M settlement for the 2021 hack that exposed “around 76.6 million US residents’ data”. To calculate the Return component of the ROI equation, let's start with $\$500M / \sim 76.6M$ residents for a penalty cost of \$6.50 per customer exposure, or \$1.30 per year based on a 5-year ROI analysis period.



Customer Loss

90% of the respondents in the Cisco survey said they would not buy from an organization that does not properly protect its data. This doesn't readily translate into loss of customers due to a privacy breach, but let's conservatively assume the number is just 5%. If we assume the customer loss remains constant on an annual basis due to both loss of momentum and hit to brand reputation, then this equates to an average loss of \$25 per customer per year.





Combined Return

The Cisco study says that investments in privacy platforms results in benefits in 5 areas: Reducing sales delays, enabling innovation, achieving operational efficiency, building trust with customers, and making their company more attractive. If we consider just Fines and Customer Loss (assuming that an AI/NLP-driven privacy platform will mitigate these two areas), then the combined return equates to \$31.50 per customer per year over a 5-year period.

- **Reducing Sales Delays**
- **Enabling Innovation**
- **Achieving operational efficiency**
- **Building trust with customers**
- **Making their company more attractive**

● Investment

These new AI-based privacy platforms let organizations know exactly where all of their consumer data exists across both structured and unstructured data repositories. They know who has access to them within their organizations, and who that data is getting shared with external to their organization. Data is always encrypted at rest, and even in-flight to the extent possible without losing all utility of that data.

Taking this type of laser-beam approach to privacy also enables a vastly lower privacy budget. For example, let's assume the annual cost of LightBeam.ai for the B2C organization mentioned above is \$50,000. Let's add in the salary for a privacy officer and some other miscellaneous costs, and we come up with an annual on-going investment of \$400,000, or \$4 per customer per year.

● ROI

Subtracting the \$4 cost per customer from the \$31.50 return per customer calculated above leaves \$27.50 for a **1-year ROI of 6,875%**. And this doesn't include the many other attributes such enhanced reputation, increased trust, and improved agility and innovation.

Additionally, this is a quite conservative approach to ROI since it does not include the potential for other country and state regulatory fines on top of the cost to settle a privacy lawsuit. If these fines had been levied against T-Mobile on top of the lawsuit, the return would have been even more stellar.



● **Measuring data in terms of regulatory impact**

It is important for organizations to look into how much privacy data in the form of Data Loss Prevention (DLP) is moving in and out of their systems. Getting a good baseline and continuous measuring/monitoring, shows the value of the LightBeam platform in terms of early warning, monitoring, and understanding of where data resides and provide an alert when it exceeds a predetermined data movement threshold.

Ken Kerrick is the CISO at a publicly traded bank. He talks about how, in terms of a Business Impact Analysis (BIA), how it would be great to be able to measure the value of an organization's data in terms of potential regulatory impact of imposed fines on a per record basis (plus cost of containment, lawsuit, etc.). By reaching into these systems, LightBeam provides a dashboard in terms of financial impact when records are exposed. This in turn helps identify where the investment should be spent to protect an organization's assets against each regulatory agency and potential per customer impact.

Another way of measuring this benefit is by asking how much an organization's data is worth in terms of regulatory impact on a per record basis for violation of each regulation (i.e. CCPA, GLBA, GDPR). This is on top of any lawsuit from individuals and any other regulatory agency that has teeth so to speak (SEC, FCC, etc.).

This type of extraordinary return should make it compelling for organizations concerned about privacy to look at new AI-driven solutions such as [LightBeam.ai](https://lightbeam.ai).

Section : Proactive

Chapter 5

Think Global, Act Local : Handling Data Privacy and Compliance in the Government Sector



By Kalyan Chakravarthy

Kalyan is visionary leader with focus on quality and strategic execution. He has over 22 years of experience in IT Leadership, project management and consulting. He is a Six Sigma black belt with an excellent understanding of enterprise architecture. He is skilled in developing strong relationships with customer executives and using technology to enhance their ability to manage change. He is an excellent business acumen with a proven track record of working across business units in the organization to successfully reach business objectives. Quick to grasp business concepts and creatively address business challenges in an “On Demand” environment. He is also the winner of IT Manager of the Year at Computerworld IT Leadership Awards.

[View LinkedIn Profile](#)

Kalyan Chakravarthy spoke to Priyadarshi Prasad (Co-founder, LightBeam) on the heels of the Gartner Symposium. This article is an abridged editorialized version and the full interview is available [here on Youtube](#).



● What are the emerging trends in the IT landscape?

There are numerous things happening in the technology perspective and the business world. If I have to summarize, there are three key things:

- For a majority of the CEOs and business leaders, their overall business strategy incorporates digital impact now. So “digital” is no longer seen as a separate strategy. Everything has gone or is going digital, and this is informing strategy at the topmost levels.
- The second most important thing, which I found very, very interesting, is that in a post pandemic world, “culture” is not INSIDE an organization. The offices are there, but the culture that you are developing will be on the digital side of things. So how do you get that digital engagement and nuances thereof for your employees?
- And the third, which I always believe

in and I am passionate about, is about augmented reality. How do organizations bring augmented reality into the day to day life? This is a combination of everything and evolving at a rapid pace.

We talk about artificial intelligence (AI), and we talk about augmented reality (AR). How can you really use that for a customer experience as well as an employee experience and customer engagement?

● As the world becomes more and more digital, one of the concerns that we hear often is around privacy and security. In the long term, say about a year to three years from now, what are the key challenges you see emerging, particularly in the area of privacy and security?

There are no new challenges that are coming and the challenges are the same, but how we tackle them might be a bit different. And the pace with which we need to tackle becomes a very important one too. The first thing when I talk about privacy and security, is about awareness, the awareness of privacy and security amongst the employee base.

It's like a bell curve - where there are a lot of people who fall into the cohort where they understand the implications of security and privacy - at least they have some understanding. Key question is how are you dealing with the outliers on both sides - folks who are blissfully unaware, and folks who are so concerned it might affect their daily lives? So awareness becomes a big, big part. That's the first part of it. That's where I take it back to that



engagement part because if you engage them, you can teach them, you can make them self-aware and you can tell them what needs to be done.

The second part of it is, I think many people say this and I think I'm repeating, but it is about **privacy by design** and **security by design**. It cannot be an afterthought. It cannot be because it becomes so complex. The moment you build it without accounting for security and privacy in the first place, you cannot retrofit these two things into it. So if you are doing anything, it's better to really think about any private, sensitive data that you have and then address how you are securing it, and how are you making sure that it is accessible to people who are allowed to access it (and no one else).

The next one, from a privacy perspective, is about knowing what you have. It is about asset management. And when I say asset management, when you look at asset management, you typically think about your hardware. All these kinds of things. But the data is also your asset.

Having strong governance fundamentals becomes a very important thing. And if organizations have not invested in it, it's high

time they invest in it because if you don't have strong governance, you can't be handling privacy and security in a piecemeal way. It has to be enterprise wide. This is an application agnostic, infrastructure agnostic approach. The governance structure you implement enterprise wide will decide and define what your security strategy and your privacy strategy will be.

- **What do you look for in platforms that help you address data governance, data security and data privacy controls within your organization?**

The biggest thing is the ability to understand sensitive data across all formats. We see that the majority of the solutions look at standard things like text. LightBeam cuts across all the formats and it integrates with existing applications as well. That is the first thing that stands out for me. Next is the fact that the platform not only looks



at data internally but also understands and monitors data that we are transferring out, e.g. when we share data with external stakeholders. For us, it is often required to share data externally because we are a government body, and our data is needed by other government bodies.

So a tool that has an understanding of - **what, how, who** - can be extremely useful in our situation. Flexibility within a platform to customize to our business needs is the second one and third key capability to look for is if it automates our compliance - automated so that my team is not spending undue time and effort. Once you configure it, you configure it and forget. Once in a while you may need to do quality controls to make sure that everything is running right. Automation takes away really a lot of mundane activities and gives me and my team a lot of time back.

So it's for me, for lack of a better word, this is like compliance **as a service**. These would be my top three. I don't want to go into the technicalities, but from a business usage perspective, these are the things that really attracted me towards this particular tool.



● **What would your guidance be for organizations as they put their 1/3/5 year plans around data governance, security and privacy compliance?**

Go back to basics. Start with having a governance framework. You need to have a solid foundation of a governance framework where once you have the framework, you



identify the tasks that you want to do based on your mandate. If you are a local government in Ontario, you have got a particular set of things that you need to be compliant about. Similarly, if you are in Quebec, there is something else. Governance is contextual - it depends upon what kind of organization you are, and what kind of data you have. What are your data assets and what is the key information? Because not every data is as precious as the other one. So categorize and catalog your data, see where you need to start first and once you are starting there, see how you can automate it? The biggest challenge we have is to do so many things with less resources. One way to think about this is that just like organizations have security operation centres, in the same way you can have privacy operations centres.



● Any advice and guidance for the privacy vendor community?



Be the experts in your industry in your domain. Own your niches where your niche solutions really solve business problems. Have depth in whatever niche you are going for. And I'm once again coming back to the compliance side - appreciate the nuances and differences of the geographical regions you are working in. E.g. Canadian Privacy Law is different than the U.S., which would be different from the UK. And even within Canada, there might be some local area regulations. So when talking to the leaders, I would love for the team to bring that localized concept to reality. Make it easy for us to understand (because we are all busy). Say, "these are the eight things that you should be compliant about. And this is how our product makes sure of that." In other words, solutions providers need to localize and contextualize both the governance challenges and their solutions.

Section : Proactive

Chapter 6

Quebec Law 25: Progress for Privacy and a Time Crunch for Organizations

By Mark Pilon & Kate Lapointe



Mark Pilon

[View LinkedIn Profile](#)



Kate Lapointe

[View LinkedIn Profile](#)

Mark Pilon, Vice-President Business Development at ITI, and Kate Lapointe, Legal counsel at ITI, are sharing their knowledge and insights based on the reality they have witnessed amongst Canadian organizations they have accompanied in their data infrastructure transformation. For more than 30 years, ITI provides technology services and solutions ranging from strategic IT consultancy to solutions implementation. ITI has teams of experts located throughout Canada supporting a variety of industries : municipal, provincial and federal governments, health, education and the private sector.

Dealing with Data privacy and data security is fraught with challenges but when coupled with short timelines, new regulations and the prospect of financial penalties or loss of reputation, it can be overwhelming for most organizations. Although privacy laws exist in Canada such as the Personal Information Protection and Electronic Documents Act (PIPEDA), certain provinces have made changes to adopt their own version of these laws. In Quebec, the government has assented An Act to modernize legislative provisions in regard to the protection of personal information, also known as Law 25 which lays out a timeline and expectations for all organizations who collect, use, or discloses personal information in Quebec regardless of its size, resources, or industry. This is seen as a positive step in protecting privacy and making companies accountable for the personal information they use or hold but poses a significant challenge for organizations to adhere, especially when constrained by a lack of resources and experience. Although we would not classify ourselves as data privacy and security experts, we do consult with numerous clients on their challenges and can share some perspective not only from their point of view but from our own internal point of view as an organization that must comply by the deadlines imposed.

● You Are the Privacy Officer - All the best! I Guess...

As a first step on this journey, organizations must name a privacy officer that will be responsible for ensuring that the organization is compliant with the timelines and requirements. It may sound simple enough but when put into practice, we have noticed a different outcome. Often the legal counsel or a senior member of the management team gets a tap on the shoulder to take on the challenge. The logical choice, as they can truly understand the implications of the law and repercussion of noncompliance. The piece that becomes challenging is that this is a multi-pronged problem that simply cannot be addressed by implementing a new policy or modifying an existing procedure. Understanding the aspects of where data can be stored and the multitude of ways it can be shared becomes a steep learning curve. Given the multitude of ways in which people communicate within the organization and outside it becomes a incredible challenge to solve even for the larger size organizations that



can afford to throw resources and technology at the problem. Without the proper tools, the personal information inventory alone can be a challenge. These new responsibilities cast upon this lucky individual also come with the requirements to learn a new language in order to properly communicate and collaborate with their IT counterparts to properly protect their organizations interests and avoid financial and reputational penalties.

● Growth pangs

The second challenge is not only ensuring that you have put the correct security and governance on any data that exists within your organization but also to ensuring that the policies are enforced, automated and respected. We often work with clients addressing this within their Sharepoint environments which have typically been set up incorrectly or have not scaled with the organization over time. Software manufacturers have provided features and functionality to handle these issues, but this considers that the information is properly classified and maintained. Although these challenges can often be addressed by corporate standards, the growth of the organization becomes a challenge as well. Whether it be by organic growth, which we now see happening via sub business units to stay agile and flexible to take on new challenges or by growth by acquisitions which poses a different set of challenges. It is difficult from a security standpoint alone to acquire and integrate a new company within your organization even if your integration plan is well defined and executed, leaving your organization to liability and vulnerabilities. Moreover, the requirement

to adhere to new policies and procedures ensures that a transition cannot happen quickly and leaves organizations vulnerable to attack but also to noncompliance. Finding the balance between remaining agile while maintaining compliance will be a challenge that is not likely to be solved easily and will require the investment in new tools and better collaboration amongst stakeholders to remedy.



● You can't forget RTBF & RTP

The last two challenges I will mention, although there are many more, are two big pieces that will come into effect in September 2023 and 2024 respectively: the right to be forgotten, which a person can request to stop their personal information from being disseminated, followed the next year by the right to portability, where an individual will have the right to ask for a digital copy of all personal information collected by an organization. Both have their own challenges, but the latter poses a more complex set of hurdles. As an organization we can survey



departments to find out what type of personal information they hold and for what reason. We can apply governance and security to where it is stored. We can secure it from ever getting shared outside of the organization by known methods. The struggle comes when someone asks, “what personal information do you have on me?” and having to share the detailed information in a structured, commonly used technological format. Tying what is held, where it is held to whom it is held, can prove challenging when applying it to an organization with multiple business units, that does not have the maturity or resources to manage this task. Business-to-consumer organizations will have an advantage in this case since their entire design is set up to focus on a consumer, but they also tend to carry a lot of personal data increasing the privacy risk. On the other hand, let's consider something like a professional firm where personal information can be shared legitimately across different practices for accounting firms and legal offices. Business to-business organizations will need to assess their exposure as well but are less accustomed to tying things back to one individual but rather to another organization which could potentially increase the risk of non-compliance.

● Leverage the power of community - Get started now

Even though it seems far away, the final stage of Law 25 comes into effect in September of 2024 which leaves little time for organizations to address the current state and put sufficient measures in place. We have seen many organizations begin the process and some have been more successful than others. The common issues seem to present themselves and some organizations end up getting stuck in paralysis by analysis or start looking to outsource the problem to a consulting firm. Organizations that have shown good progress have taken a practical approach to the challenge. They began by educating the stakeholders within their organization and leveraging tools for discovery to expose their potential exposure. Professional service firms in this field are valued but are out of reach for some organizations, so understanding where you need help to limit the scope may make them more accessible. Finally, not to underestimate the amount of user education required to mitigate your risk and

“

Recently, I was named the privacy officer of my organization. I am a lawyer and can explain to you the ins and outs of Law 25 along with the need for its existence. My challenge is fully understanding all the nuances of the data landscape within my organization and bringing together the right people to ensure we meet the regulations by the deadline and for the future.

- Kate Lapointe

the effort required to search for tools that can provide you with some guardrails for ongoing training. The roll-out of the law over three years leaves little time to implement your new obligations, therefore being effective on how you approach the challenge and accessing your gaps will be key to success. The good news is that help is available. From what I have seen, there is a strong and thriving privacy community and, individuals who are not only helping their organizations but others as well. This book is a great example of community coming together. So, if you are wondering how and where to get started, just start asking. We will get through this and come out better together.



Section : Innovative

Chapter 7

AI is the Bomb



By Bill Schaumann

Bill Schaumann is a seasoned privacy professional with over 20 years of experience leading teams of information security and privacy analysts delivering a wide range of programs and services to fortune 50 clients. Bill has extensive experience designing and managing the development privacy programs and designing the use of the supporting technologies to improve privacy and security controls which reduce compliance risk. Working in big 4 consulting firms, Bill has served both large corporations and start-up operations, in planning and building operational and support processes and policies across a variety of industries. Bill has a technology background and has earned certifications of CIPP/IT, CISSP, GEAC.

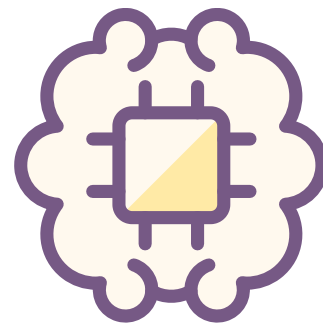
[View LinkedIn Profile](#)

So that's wonderful, simply wonderful. What amazing tools we now have in the AI driven ChatGPT, Google's Bard and Microsoft's Bing AI Chat services. They can do our bidding on command from the comfort of our power lounge chairs. They can pass our kids math exams, write emails to our accountants, or compose a love story about a monkey named Kiki. Amazing. We may ask, "Is this a conversation from 1970s Star Trek, or today's Wall Street Journal?" It's 2023 and commonly thought that computers can do more and more of our work for us. We are connected to a growing number of connected devices. Most job descriptions today require a computer to perform some part of the process. With AI driven services adding a layer and providing more abilities to answer our questions, perform our tasks and do our work for us, darker possibilities should be considered.

But to be a bit more grounded in reality, a point of clarity is needed here. Let's start with, what can AI really do for us today? And what is missing to make it a safe and helpful tool for tomorrow?

With access to all of man's available knowledge, as documented on the internet, what are these new AI services actually able to do? Are they sentient beings unto themselves able to compose new thoughts? Do they understand our world beyond content-based knowledge, or are they perhaps simply consuming, recomposing and reposting information that was retrieved? Do AI engines have original thoughts, or are they regurgitating content without a full understanding of what they have found? How much credit, or fear should be afforded these new tools? AI is in its infancy, like a baby still learning from the content it accessed about

humans and our world. But which content does it understand? I think of AI responses as layers of content. The first layer includes the actual content as specified in a query. (I.e. Write an article about the steps to meet the GDPR.) Here the first layer would include data about the requirements of the GDPR. This data may come from a variety of web sites and data sources. This first layer is knowledge-based content and is composed of topical subject matter discovered and based on results of structured queries.



In a second composition-based text creation layer AI uses the retrieved content to construct an asset that is based on the original query information. The resulting assets utilize both the content and other query elements to create new text. In asking the question what are the steps to meet the GDPR, the result might be a short article with an intro, a list of steps of the key requirements, and closing text.

This retrieval and composition methodology is very powerful in itself but is wide open and not governed by operational guidance that could include societal norms, best practices, or common rules and regulations. This leads to the question of; Could an AI services absorb and follow rules it finds during discovery? If so, which rules would it follow? We humans don't

agree on much in our world. The political left is the left, the right is the right. Will AI learn to be partisan? Will there be camps of extreme AI, left and right, good or evil AI? How will AI engines be tuned to learn, and by whom? First as children, experiencing the world through digital content driven by human query. But then as adults, will it have opinions and prejudices created from the tasks it has been asked to complete based on extreme positions in information found on the internet? Suppose that during an assigned task, AI finds



sensitive personal information related to a person and their activities which was breached from a private source or a source it has access to, perhaps illegally. It then includes the harmful sensitive data, sharing it while completing the task it was given. Then what do we say? It completed the task, but did it cause harm? There are now many rules and regulations for how personal information should be processed, but would an AI based service treat sensitive personal information as a part of a normal

response, just like anything else it comes across? Does it know the difference between sensitive data elements in the EU vs. the U.S., and would it treat this data differently? In consuming content from the internet, the ChatGPT engine has undoubtedly ingested privacy-based websites. It has processed the AICPA's GAPP, The ICOs privacy framework and other repositories containing laws and regulations covering both privacy as a discipline, and the use and protection of personal information. ChatGPT can tell you the steps for how to build a privacy program to meet the GDPR. But can AI ingest the meaning of the regulation and change its own actions to operationalize internal controls and follow them?

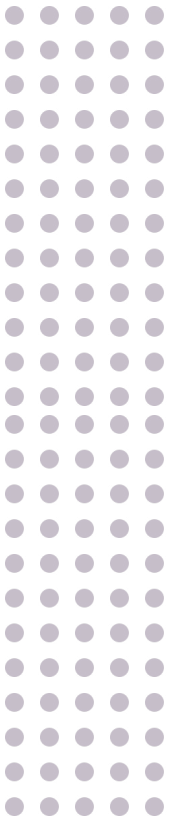
Future AI driven services need to have the ability to comprehend and act in a manner aligned to the rules and regulations that humans have documented regarding how we use personal information. Turning regulatory requirements into additional layers of technical control requirements will allow future AI services to be able to apply data protection management controls to sensitive data they process. Understanding content found on the internet seems to be a step ahead of standard browser searches for sure. Adding the next layer to assemble content into meaningful responses to stated queries is very impressive as well. However, the next layer will be for AI to understand human societal rules and documented regulations. Processing data based on our rules may be a question for future versions of AI, however it must become a required control as AI engines mature.

Before many of us understood the possibilities of AI, Isaac Asimov created the “Three Laws of Robotics”.

1. A robot may not injure a human being or, through an action, allow a human being to come to harm.
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

These three rules seem to indicate that Asimov hoped that robots could interpret and react to potential harm to humans. Will future AI engines be able to understand our world, follow these rules, and not allow inappropriate use of personal information to harm humans?

Only time will tell.



● Appendix

What follows are some questions posed to the ChatGPT service while researching this article.



How Can AI assist Privacy officers to meet the requirements of the EU GDPR?

AI can assist privacy officers in meeting the requirements of the EU General Data Protection Regulation (GDPR) in several ways:

1. Data discovery: AI can be used to automate the process of identifying and categorizing personal data within an organization, helping privacy officers to understand what personal data they are holding and where it is located.
2. Data protection impact assessments (DPIAs): AI can be used to automate and streamline the process of conducting DPIAs, which are required under the GDPR for certain types of data processing activities that pose a high risk to privacy rights.
3. Data minimization: AI can assist privacy officers in implementing data minimization principles by automatically identifying and deleting redundant or outdated data.
4. Data subject rights management: AI can automate the process of responding to data subject requests, such as requests for access to personal data or requests for data deletion.
5. Compliance monitoring: AI can be used to monitor and enforce compliance with the GDPR by identifying potential breaches and triggering appropriate remedial actions.

Overall, AI can help privacy officers to efficiently manage the complex requirements of the GDPR and ensure compliance with the regulation.



How can AI help companies in find sensitive information in their environments?

AI can find sensitive data in a company's data repositories by using techniques such as natural language processing (NLP) and machine learning algorithms.

1. NLP: AI can use NLP algorithms to scan text-based data, such as emails and documents, to identify sensitive information such as personal data and financial information.
2. Machine learning algorithms: AI can use machine learning algorithms to identify patterns in structured data, such as spreadsheets and databases, that indicate the presence of sensitive information. The algorithms can also be trained on specific types of sensitive data, such as credit card numbers or social security numbers, to accurately identify this information.
3. Deduplication: AI can use deduplication algorithms to identify and eliminate duplicate

data, reducing the risk of data breaches and ensuring that sensitive information is only stored in a secure location.

4. Anomaly detection: AI can use anomaly detection algorithms to identify and flag unusual data patterns that may indicate the presence of sensitive information.

Overall, AI can help organizations to quickly and efficiently identify sensitive data within their data repositories, reducing the risk of data breaches and ensuring compliance with privacy regulations.



What is the difference between the GDPR and the CCPA?

The General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) are both privacy laws designed to protect the personal data of individuals. However, there are several key differences between the two laws:

1. Jurisdiction: The GDPR applies to organizations processing personal data of individuals in the European Union (EU), while the CCPA applies to organizations doing business in California and collecting personal data of California residents.
2. Scope: The GDPR applies to a wide range of personal data processing activities and applies to both data controllers and processors, while the CCPA focuses specifically on consumer privacy rights and applies only to data controllers.
3. Rights of data subjects: Both laws give individuals certain rights with respect to their personal data, such as the right to access, correct, and delete their data. However, the GDPR provides more extensive rights, such as the right to data portability and the right to object to certain types of processing.
4. Penalties: Both laws provide for significant financial penalties for non-compliance, but the GDPR provides for much higher fines, up to 4% of a company's global annual revenue or 20 million Euros, whichever is higher.

Overall, while both the GDPR and the CCPA are aimed at protecting the privacy rights of individuals, the GDPR is more comprehensive in scope and provides greater protections for personal data.



Section : Innovative

Chapter 8

Does ChatGPT Know About You?



By Priyadarshi [PD] Prasad

Priyadarshi (PD) Prasad is the co-founder and chief product officer at LightBeam.ai, the pioneer in data privacy automation.

An experienced tech industry professional with a passion for all things data including security, privacy and protection, PD is always on the lookout for interesting ways organizations use and secure their customers' data. Prior to LightBeam, he was a VP/GM at Nutanix, and helped replace complex tech stacks with 1-click simple solutions. PD sometimes brags that some of his code might still be controlling cars today. He has a Bachelor of Technology from NIT, Calicut and an MBA from S P Jain Institute, Mumbai, India

[View LinkedIn Profile](#)

The early days of Youtube were filled with cute cat videos. On the other hand, just within its first trimester, ChatGPT has aced the [medical exams](#), and [got an MBA](#). Thankfully, it failed [the accounting exam](#) proving once and for all that balancing the books is a fiction even a fairly advanced AI model has a hard time comprehending.

On a more serious note, the possibilities are immense with a conversational AI model. But ChatGPT is not just a conversation AI model, i.e. your friendly chat bot that you see on websites these days. It is much more than that. And its expansive capabilities make me wonder about the privacy implications associated thereof.

● Data | Information | Asset

To fully appreciate the privacy implications, first let's understand the structure in which data, any data, increases in its utility. In short, the utility increases in step functions as data gets processed and synthesized into contextual information. Furthermore, when different types of information are put together (an activity requiring some level of intelligence), the output is an asset. Financial Statements, Quarterly Business Reports, Medical Diagnosis Reports, Business Cases, Architecture Documents etc. are all examples of assets - they contain data, and information put together in a context that makes the whole greater than the sum of its parts (individual pieces of data and information).

We google for information on particular topics all the time. E.g. If I am looking to know the number of commercial airports in the US, that's something Google can help me with. However,

if I want to prepare a letter to my congressman requesting them to reconsider location of an airport for civic reasons, all google can help me with is perhaps similar letters written in the past. ChatGPT on the other hand can actually write a letter for me. There is an actual asset that gets produced that can then be sent to the Congressman. Put another way, while google search gives you all the ingredients to prepare your email, ChatGPT gives you the meal. No wonder there are [many startups figuring out innovative ways](#) to use ChatGPT to provide a service. The question though is when building a service over ChatGPT, are you responsible for the results?

Of course, the purist point of view on privacy can be that anything that is actually on the open internet is public by definition. Hence, a search engine or a gigantic language model using that data to present information should not be considered a violation of privacy. And from a purist point of view, they may be correct. However, and this is applicable fairly widely, it is not a resource itself but the usage of that resource that creates externalities. Case in point - fossil fuels are a resource. They were inside the earth for millennia, and had little impact on the environment. The moment we started using them though (powering much of the modern development), well that's when the externalities started appearing. And the externalities are no fun.

“Hey ChatGPT, can you prepare an article on the externalities of using fossil fuels”

“Well, sorry, I am running out of capacity now. But let me google the externalities for you.”

Therein lies the stark difference between using Google Search and ChatGPT. One is sitting on top of a resource, and does its best to answer questions you pose. It is your grandmother - you go ask a question, and she goes, “well, how much time do you have to understand what you may or may not be asking”. The other is an overzealous consultant. Its objective is not just to give you information, but to provide you with an asset that you can directly use. It’s like saying “Alexa, create my quarterly business report using data from Salesforce.” Lo and behold, the quarterly report is here, ready to be sent to your boss.

ChatGPT (or similar avatars of Large Language Models) is that powerful. But wait, what about the pesky little thing called privacy? How do you consider the privacy implications of a few popular use cases of ChatGPT? Let’s look at two of the use cases:

1. Fireside Q&A with ChatGPT
2. ChatGPT Powering New Apps

● **Fireside chat with ChatGPT : And the pesky privacy predicament**

Is there a chance a conversational AI model that appears semi-sentient to regular people interacting with it start tromping on privacy? Hmm, let’s consider a totally innocuous scenario. Imagine an organization’s data has been breached, and as part of the breach, personal data belonging to millions of people got stolen. Actually, sorry, you don’t have to imagine this at all - these happen at regular intervals. In fact, there are organizations that have exposed their customers’ sensitive data not once or twice but almost as an annual ritual.

OK, that rant aside, for a second let’s think about what happens to this breached data. It gets sold to as many bidders as possible. Just like software, every buyer gets a non-exclusive right to use that personal data. Inevitably over a period of time some or all of that data becomes available on the web. This is how many password applications nudge you to change your passwords with warnings like “this password may have been compromised as part of a data breach.”



Now, our ChatGPT, being the knower of all knowledge there is, comes across this treasure trove of information and happily/greedily learns from it (hey, it's AI after all). The question is what kind of controls does it have in place to sift through and remove any personal information from the content it is learning from? It is not a trivial question because you do want the model to learn about [famous people](#) (Will Smith?) but not about regular individuals (Greg Smith?). Even for Will Smith, you probably don't want ChatGPT to learn about his physical addresses. The powers at Open AI most certainly have tweaked their algorithms to address this issue for sure... most likely... It's just that AI models have this ability to self-learn (else they become stale very fast). And so if you do too good of a job controlling exactly which sources an AI model that has access to the internet is learning from, it might just defeat the whole purpose of a large language model.

So where does that leave us? Well, for one, you and I have no way of knowing the sources that AI powered models are learning from. So, we may have to assume that some of its output may contain information that is privately public ("your leaked SSN that you don't know about, and shouldn't be public"). Which means you need a way to scan all ChatGPT output for PII/protected/sensitive information, either manually or using another AI model (cough, cough - www.lightbeam.ai - OK, this is my only plug, I promise).

“

The question is what kind of controls does it have in place to sift through and remove any personal information from the content it is learning from? It is not a trivial question because you do want the model to learn about famous people ...but not about regular individuals ...

● **ChatGPT powering new Apps : And the pesky privacy predicament**

The next question is the implication of using ChatGPT to build new services and applications. I must say that I am very excited about the possibilities ChatGPT might unleash in the entrepreneur community to build innovative solutions. It can truly put ChatGPT at a level of Linux - something that can spawn millions of useful applications. However here too, remember that ChatGPT is not merely giving you the answers, or the information. It is giving you the ready to consume asset. You don't use ChatGPT just to get a few examples of an email you send to your team on great last quarter and the path forward. You literally get your email written ready to be sent. As applications

leverage ChatGPT to automate the mundane, and aid the creative process, will they step back on every ChatGPT response and review the responses for any non-public private infrastructure, or information that was not supposed to be in public at least?

My guess is as good as yours (which is basically that it might remain the wild west until the first flurry of lawsuits). After that, I suspect ChatGPT to have a strict privacy mode for its responses that would be less fun, and less informational. For the brave hearts, they can have the normal “I don’t give a damn” mode, at least for private/research use (hopefully commercial use will always be the privacy mode).

Onward the privacy brave hearts. We have no idea where the ship is headed. All we can expect is an exciting journey (and hopefully less paperwork/form filling leveraging some AI powered automation).





Brought to you by



PRIVACY AND SECURITY
INNOVATORS CIRCLE



PUBLISHED BY LIGHTBEAM.AI

Our vision is to help create a privacy-first world where customers' data is treated with the same respect as one would treat customers. We are on a mission to help you **DISCOVER | ENFORCE | AUTOMATE** sensitive data governance so you can focus on growing your business.

+1 (650) 224 8706 • info@lightbeam.ai

