dathena

Solution Brief

An Introduction to **Dathena Privacy**

A Pioneering, AI-Powered Solution that is Transforming How Organizations Protect Personal and Sensitive Data

March 2020

Contents

Introduction

All organizations across the globe are struggling with how to keep personal and sensitive data private. Primarily driven by ever-evolving government and industry regulations, organizations that manage consumer data or maintain any form of intellectual property (IP) have many reasons to protect personal and private data. In addition to incurring significant fines for non-compliance, failure to protect personal and sensitive data can also result in loss of customer trust, brand damage, competitive disadvantage, and lost revenues.

This solution brief introduces **Dathena Privacy**, an **Artificial Intelligence (AI)-powered** solution that addresses the critical problem of data privacy management and fast tracks data privacy compliance at a fraction of the cost of other solutions and manual processes.

This brief is appropriate reading for any line-of-business or IT professional looking to learn more about data privacy and how Dathena Privacy addresses the challenges issues associated with protecting personal and sensitive data. By reading this document, you will discover how Dathena Privacy is revolutionizing the data protection with a next-generation AI-powered solution.

Data Privacy Defined

Data privacy is an individual consumer's right to ensure that their personally identifiable information (PII) or Sensitive Personal Data (SPD) is kept private. For any organization that collects, stores, and manages consumer data, data privacy must be engrained in the organization's culture and is built on a foundation of trust, transparency, and accountability to individual customers. From an execution perspective, data privacy is an important element within Information Technology (IT) that addresses how PII is collected, used, accessed, and managed to ensure it is kept private.

PII can range from a person's name, social security number, driver's license number, bank account number, passport number, email address, financial information, even information about an individual's health, gender, family, religion, and criminal history.

Different regulations and industries refer to personal data by different names, including PII, PCI (card information), SPD, and PHI (Protected Health Information). In addition to personal data, organizations also need to protect sensitive business data as well. Regardless of what it is called, Dathena Privacy can protect it.

Five Reasons
Why Data Privacy is Important

Meet Regulatory Requirements and **Reduce Compliance Fines** 1

Capture Your Customers' Trust 2

Protect Your Intellectual Property 3

Eliminate Manual Processes and **Reduce Costs** 4

Be the First to Market (or Close to it) 5

The Importance of Data Privacy

1 Meet Regulatory Requirements and **Reduce Compliance Fines**

Privacy is essential to the autonomy and protection of human dignity, serving as the foundation upon which many other human rights are built[1]. In fact, many countries consider privacy a basic human right. Organizations that handle personal data have, at a minimum, an ethical obligation to keep it private and, in many countries, data privacy regulations require – under threat of fine – that PII, and PHI be private and never shared with unauthorized individuals.

The enactment of the EU General Data Protection Regulation (GDPR) has compelled European countries to act quickly to protect personal data as it is enforceable by fines of up to €20 million or 4 percent of global revenue — whichever is higher. As of January 2020, "European data protection regulators have imposed EUR114 million in fines under the GDPR regime."[2]

If your organization isn't impacted by GDPR today, you don't want to get too comfortable. According to an April 2019 Forrester report, "many national and local governments around the globe have indicated they will model their own laws on the GDPR, such as the California Consumer Privacy Act (CCPA)."[3] While there is still confusion as of this writing on the effective date of the CCPA[4], CCPA has been described as "almost GDPR in the U.S." and is "far and away, the strongest privacy legislation enacted by any <U.S.> state at the moment.[5] [6] With California hosting many global technology companies, such as Google® and Facebook®, the CCPA will have a major impact on how these organizations do business going forward.

Other countries that have enacted data privacy regulations include Singapore, which implemented the Personal Data Protection Act, 2012 (PDPA). Brazil just passed their version of GDPR called Brazil's Lei Geral de Proteção de Dados (LGPD), which became effective in February 2020. India is currently legislating a Personal Data Protect Bill (DPB). But even in countries that are not currently governed by personal data protection regulations, many organizations must comply with industry-specific regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S.

2 Capture Your Customers' Trust

According to a recent Forrester report, "44 percent of U.S. online adults, 52 percent of German online adults, 53 percent of French online adults, and 47 percent of UK online adults say that they are likely to cancel an online transaction if they read something they don't like in the privacy policy."[7] Clearly, this indicates that if you don't protect your customers' personal data and give them the right to choose how their personal data is used, you lose their trust, which impacts loyalty, and ultimately your brand reputation, your competitive position in the market place, and revenues.

3 Protect Your Intellectual Property

In addition to personal data, your organization maintains highly sensitive business data that differentiates your company, products, and services, all of which must be protected. These include financial information, your customer database (which some argue is your most important asset), research and development information, brand secrets, trade secrets, patents, formulas, recipes, designs, code and search algorithms.

4 Eliminate Manual Processes and Reduce Costs

If your organization is managing personal and sensitive data using manual processes, you are spending a lot of time and money doing so. Worse still, manual processes result in high error rates. Automating the labeling of data can save your organization thousands of dollars a year and free up your employees to focus on higher-value work.

5 Be the First to Market (or Close to it)

If you know anything about data analytics, you know that managing, working, and synthesizing data can give you a competitive advantage when it comes to product development and marketing. Organizations across the globe are looking for ways to differentiate themselves by focusing on digitally transforming their business and culture and finding ways to improve the customer experience. As Forrester states: "Insight differentiates a digital business. If a firm can analyze data effectively, it can generate new intellectual property (IP) and beat competitors to market."[8] Finding innovative ways to protect personal data can lead to providing your firm with new data or new ways to use your existing data.

Data Privacy:
The Challenges

As every Data Protection Officer (DPO) knows, data privacy will continue to be an unrelenting issue for both IT and business users as the volume of your organizational data grows. Data growth has been phenomenal over the course of the past decade and will grow at an exponential rate over the next decade. In fact, IDC forecasts that the Global Datasphere – the sum of the world's data - will grow at a 61 percent compound annual growth rate (CAGR) from 33 Zettabytes (ZB) in 2018 to 175 ZB by 2025. [9]

Data growth is driven by many things including consumers' increasing use of embedded devices and the Internet of Things (IoT). Couple these drivers with every organization's focus on digital transformation, improving the customer experience, and the predictive analytics required to achieve these goals and we are now seeing data given birth to more and more data. The cycle is endless.

This begs the question: how will your organization ensure the privacy of personal data? To identify personal data, many organizations require that their IT team and business users manually classify data, but this approach is not conducive to an environment of significant data growth. Even at today's data levels, Forrester reports in a study that:

Of respondents have no idea where their most sensitive unstructured data resides

Don't classify data properly

Don't enforce a least privilege model for access to this data

Don't audit use of this data and alert on abuses [10]

“To accommodate future data growth, comply with regulatory requirements, and gain a competitive edge, your organization must embrace modern technology to automate data privacy by using solutions that incorporate purpose-built AI."

Christopher Muffat
Founder and CEO, Dathena

Data privacy is and will continue to be an unrelenting issue for both IT and business users as the volume of organizational data grows. In addition, privacy regulations will continue to evolve to add even more challenges, more work, and more costs. For example, with the advent of GDPR, Microsoft launched a global privacy self-service portal. In the first year, it received 18 million Subject Rights Requests (SRRs), with 37 percent of the requests coming from the U.S. and high volumes in countries with GDPR-like regulatory changes. In terms of population density per country, Gartner forecasts that more than 10 percent of customers will want to make use of privacy rights in some jurisdictions.[11]

According to the same report, Gartner clients indicate they spend, on average, more than $1,400 to respond to a single Subject Right Request (SRR) and that many organizations are not capable of delivering swift and precise results in answer to the SRRs received.[12] Thirty-six percent of organizations indicate it takes them three weeks or more to respond to a single request.[13]

With data volumes and the number of consumer data requests growing, your organization can no longer afford – from a cost and risk perspective – to continue to use manual-based processes.

And industry analysts agree. As recently stated by Gartner,

Organizations can "enhance privacy maturity in digital business by pursuing artificial intelligence use that focuses on compliance reporting and on continuous personal data discovery and management, especially where there are human-only processes."[14]

Gartner

Dathena's Data Privacy Solution

Dathena Privacy is a pioneering, AI-powered solution that is transforming how organizations protect personal and sensitive data. With Dathena Privacy, authorized individuals in your organization have a complete view of organizational data, its locations, and the access rights to every file and folder in the company's file system. Moreover, the solution provides an overview of the personal data in each document, allowing users to detect documents at risk based on the sensitive information they contain. Finally, Dathena Privacy links personal information across documents, enabling you to comply with regulatory requirements, such as GDPR's Subject Access Request requirement.

As an AI-driven solution, Dathena separates itself from competitive solutions with a proprietary approach to employing AI technologies. Whenever traditional approaches to using AI technologies and methodologies do not address the problems associated with scale, performance, and accuracy, Dathena's Research Team develops new methodologies, source code, protocols, APIs, and modules to ensure personal data identification is fully optimized. To that end, Dathena has filed for six patents for proprietary AI technologies, including smart sampling and autolabeling, with 13 patents pending on other AI-proprietary technologies as well.

You can learn more about the AI technologies Dathena uses at www.dathena.io

The Benefits of Dathena Privacy

With the onset of new and ever-evolving data privacy regulations in countries across the globe, Dathena Privacy can reduce the risk of data leakage and the ensuing compliance fines. This can save your organization significant money and protect your brand's reputation. However, there are many other benefits Dathena's customers are realizing.

- Reduce the risk of compliance fines
- Protect your brand's reputation
- Easily respond to compliance requirements, such as the main articles of GDPR and Subject Access Requests (SARs)
- Improve productivity, eliminate operational disruptions, and reduce incident monitoring costs
- Provide consumers with the privilege of choice, improve consumer trust, customer loyalty, and your competitive position
- Generate new intellectual property (IP) and be the first to market
- Protect IP assets

With the ability to identify personal data and act on that data as required by privacy regulations, such as responding to GDPR's SARs, Dathena's customers are realizing improved productivity, less operational disruptions, and reduced incident monitoring costs. Leveraging general and Dathena-developed AI technologies, Dathena Privacy can identify and map personal data with unprecedented accuracy – 80 percent accuracy out-of-the-box and reaching 99 percent accuracy once the model is fully "trained." When you compare the time it takes to manually identify personal data and the low accuracy rate typically achieved – 25 percent – Dathena's customers realize significant time and cost savings and enjoy peace of mind.

In an era where all organizations are focusing on improving the customer experience, Dathena Privacy supplements this initiative by providing consumers with the privilege of choice – the choice to be forgotten, the choice to know when and where their personal data is used. Ultimately, this improves consumer trust, customer loyalty, and your competitive position.

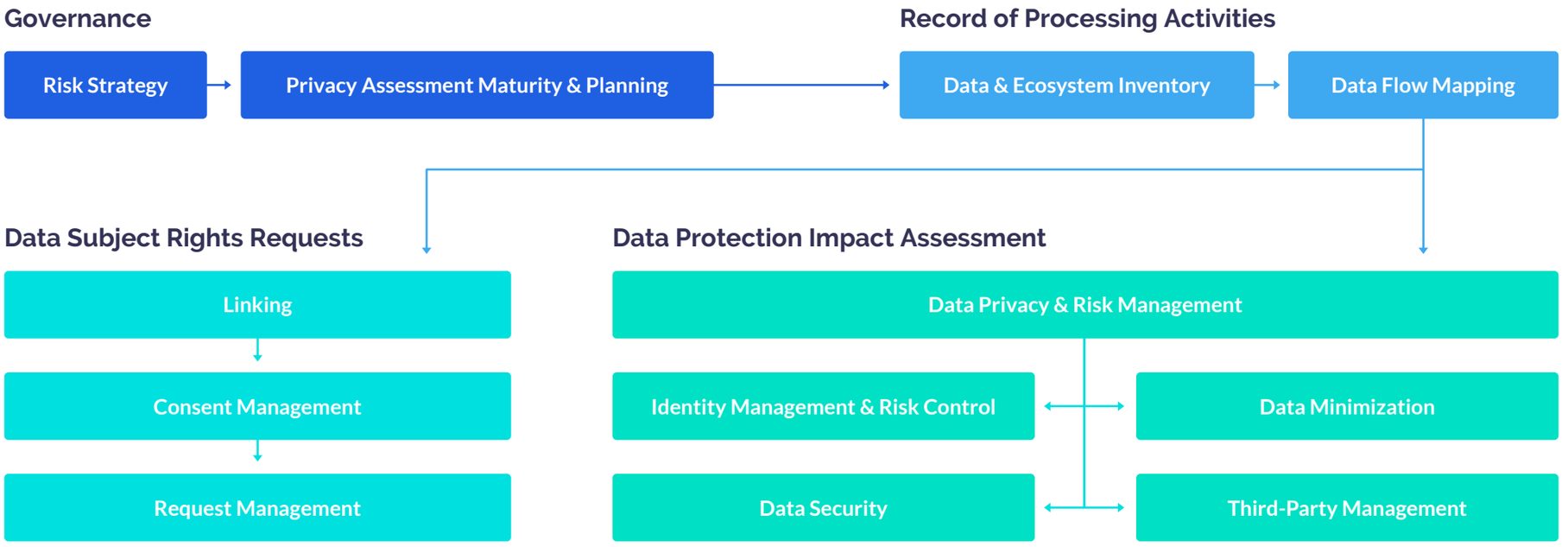
Dathena Privacy lets you analyze data differently, which leads to identifying new data and new IP that can position your organization as a leader when it comes to introducing new products and/or identifying new ways to market to your customers.

Lastly, Dathena Privacy does more than just protect consumer data. It also protects your organization's IP – the assets that power your brand and differentiate your company.

The DPO Journey Drives Dathena's Privacy Framework

Dathena's customers use Dathena Privacy to solve different business problems associated with private data. We refer to these different types of deployments as "use cases." By identifying and combining different use cases, Dathena can map out the DPO Journey and the corresponding product features, data workflows, outcomes, and technologies. Figure below provides a high-level view of the DPO journey to support GDPR. More detail on each use case follows.

Data Privacy DPO Journey & Use Cases



Governance

KEY FEATURE:

- **Develop an initial maturity assessment to create a high-level action plan.**

Governance identifies the level of an organization's compliance for each GDPR requirement in terms of policies, processes, and procedures. An organization will plan its compliance level based on its **Risk Strategy**, that is, the level of risk the organization is willing to assume.

Dathena has formed a strategic alliance with PwC, which combines Dathena's technology and implementation capacities with PwC's multidisciplinary capabilities in digital strategy, data privacy, and cybersecurity. PwC has also created a Centre of Excellence (CoE) for Dathena in Switzerland to advise Dathena's clients globally. Dathena also developed similar partnerships with consulting firms across the globe.

In partnership with a third-party consulting organization, such as PwC, Dathena performs a **Privacy Maturity Assessment and Plan**. The assessment identifies an organization's Current Profile (the "as is" state) with a Target Profile (the "to-be" state). "A Current Profile indicates privacy outcomes that an organization is currently achieving, while a Target Profile indicates the outcomes needed to achieve the desired privacy risk management goals."[15] With a Maturity Assessment, an organization can identify, evaluate, prioritize, and respond to specific privacy risks and develop an action plan for improvement. The action plan includes a task management feature to assign an action to a data controller or processor.

Dathena's customers can also use a questionnaire builder, which allows them to import their own questionnaire and perform the assessment in partnership with any consulting company of choice.

Record of Processing Activities (ROPA)

KEY FEATURES:

- **Automatically detect personal data within unstructured and structured data.**
- **Automatically predict the purpose of processing at the personal data level.**
- **Automatically create a continuously updated data processing register, which is ready to export for national authority review.**
- **Create an action plan based on answers with suggestions.**

Dathena Privacy can provide an automatic, up to date ROPA – which identifies personal data and the purpose of processing activities with high accuracy, high speed and broad coverage (e.g., data type, language, country). It also provides an action plan that is continuously updated and ready to export for compliance purposes.

To accomplish this, Dathena creates a **Data Inventory** by identifying and analyzing the source data across the enterprise whether structured, unstructured, on-premise, or in the cloud. A data inventory improves an organization's ability to identify the locations and designate accountability for the data it manages, reducing security and non-compliance risks.

Dathena also develops an **Ecosystem Inventory**, which includes a data subject inventory, a data processor controller inventory, and a third-party inventory. The data subject inventory identifies a list of data subjects and automatically predicts the data subject category.

The inventories are the key components for **Data Flow Mapping**, which is used to automatically develop the ROPA. For organizations that must comply with GDPR, creating and maintaining this register is mandatory. However, there are other benefits including the ability to identify data redundancies to minimize data and having the information necessary to operationalize DSR responses, plan data retention/destruction activities, and streamline data collection.

Data Subject Requests (DSRs)

KEY FEATURES:

- **Automatically link personal data with a Data Subject and reconcile Data Subject personal data with data processing activities.**
- **Automatically produce user consent information.**
- **Automatically respond to Data Subject Requests.**

Dathena enables **Consent and Request Management** by **linking** personal data with data subjects and reconciling data subject personal data with data processing activities.

To support consent management, Dathena Privacy collects and stores the consent from the data subject. An authorized user can retrieve this information from the database to demonstrate that the Data Subject agreed to the processing of their personal data.

To support Request Management, Dathena captures the request, produces an automatic report, and assigns a task to the data controller or processor who can add, edit, or delete information as required and send a confirmation back to the data subject. Automating responses to DSRs can reduce costs and improve the accuracy of information and response times.

If your organization is subject to GDPR or CCPA regulations, you are required to obtain user consent for collecting their data. You are also required to respond to DSRs within designated timeframes.

Data Protection Impact Assessments (DPIA)

KEY FEATURES:

- **Automatically assess risk for every processing activity.**
- **Automatically suggest remediation action.**
- **Suggest an Access Recertification Program for high-risk groups and users.**
- **Integrate with Identity and Access Management (IAM) and Risk Management Solutions (RMS).**
- **Suggest files/folders to be encrypted.**
- **Suggest files to be archived/deleted.**

Dathena Privacy can automatically perform a **risk assessment** at different levels (e.g., data subject, data processing activity, third-party) to comply with GDPR or CCPA and suggest remediation action. Remediation action includes mitigating risk impact by reducing the personal data a user can access or mitigating risk likelihood by removing access to sensitive data or by protecting data a user can access.

Dathena Privacy can identify user access privileges and suggest an **Access Recertification Program** for high-risk groups and users.

Dathena can automatically identify high-risk files for **encryption** and tagging and **minimize data** by identifying redundant and stale files.

Lastly, Dathena can be used as your platform to ensure third-party organizations are compliant when it comes to the protection of personal and sensitive data.

What Makes Dathena Different?

Dathena Privacy differentiates itself from competitive solutions by providing the following unique features and benefits:

Dathena Privacy quickly and accurately discovers Personally Identifiable Information (PII) at scale, including information that is linked or linkable.

Dathena automatically discovers where your personal and sensitive data resides and can analyze millions of files per hour by leveraging Big Data technologies that **speed** the processing through a parallelization approach.

AI technologies such as Natural Language Processing (NLP), deep learning, and computer vision reduce costs and deliver unmatched relevance and unprecedented **accuracy** – almost 300 percent more accurate when compared to human labeling processes. And when compared to competitive solutions, such as IBM Watson®, Dathena's NLP algorithms are 1,000 times faster.

Scalability is at the centre of Dathena's engineering design choices. Leveraging Big Data technologies, its solutions can seamlessly scale vertically by adding more computational power and horizontally by increasing resources.

Dathena Privacy automatically links personal data with data subjects to build a knowledge graph, which is comprised of the relationships between entities. This knowledge is then leveraged as a training dataset for **Personal Information Linking**, deep-learning models, which will produce a tailored, repository-specific predictive model.

Dathena Privacy can automatically identify the purpose of processing through unsupervised artificial intelligence (AI) and autolabeling, making it the most cost-effective privacy and security solution on the market today.

Dathena modernizes and enhances data privacy with the unique power of purpose-built AI but it is Dathena's development and use of proprietary AI technologies, such as autolabeling, that is its key differentiator and make it the most **cost-effective** data privacy and data security solution on the market. When compared to manual-based processes, Dathena eliminates 10,000 manhours of work for every AI hour worked, providing immediate business value. And if your organization is using traditional data loss prevention tools, where 90 percent false positives are typically reported, Dathena reports less than 20 percent false positives, eliminating operational disruptions and reducing incident monitoring costs.

Dathena is committed to an AI no-magic approach to ensure the results are explainable and deterministic.

According to the English Oxford Living Dictionary, AI is the “theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.”

The use of AI means that Dathena‘s solutions are based on pure mathematics, but many individuals view AI as a "black box" or something "magic." To overcome this perception, Dathena’s goal is to ensure the AI algorithms used in all its solutions are explainable and deterministic.

Explainability refers to the methods and techniques that are used in the AI application so that the results can be understood by human experts.[16] Dathena’s goal is to explain the results of its model to customers. When outcomes are presented to the end user, they will see the document and the personal information and by clicking on the data, they can see the formulas and rules.

GDPR, as an example, provides a "Right to Explanation," which is fully aligned with model explainability. Dathena's goal is to explain the results to our customers so we can leverage their knowledge in an efficient way to update the model with new rules, conditions, and factors.

Deterministic means that the algorithms are based on a series of conditions and rules that represent human logic. They mathematically calculate the probabilities for the output and are precisely determined through known data relationships. There is no room for random variation. This helps the user visualize the logical path and explains why the model came to a specific decision and which conditions, factors, and features are used to make the final prediction.

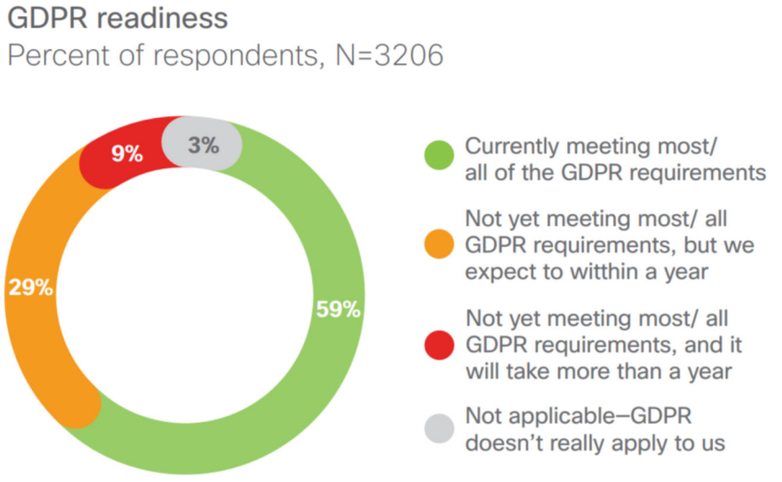
Another key feature of Dathena Privacy is its auditability. The goal of an auditable system is to answer questions such as: Is the algorithm suitably transparent to end users? Do we put any biases into the model to predict the result we want? Is the model fair?" While we can explain and track the model's logic, we can also show which data was used for decision making and prove that no biases were brought into the algorithms. By building the solution in this manner, we enable the examination of the purpose, process, execution, monitoring, and security of all models.

While AI and its subset technologies are the foundation for all Dathena solutions, the technologies are behind the User Interface (UI), which provides a series of easy-to-use interactive dashboards that let you discover and identify PII and SPD.

Final Thoughts

In a recent Data Privacy Benchmark Study sponsored by Cisco, 59 percent of companies reported that they are meeting all of most of GDPR's requirements today, with another 29 percent expecting to get there within a year.

Whether you have achieved compliance or not, consider Dathena, the only AI-powered Data Privacy Solution that helps you quickly comply with GDRP, CCPA, LGPD, and HIPAA regulations. Its easy-to-use dashboards show you where personal data lives, who has created, modified, and accessed it, so you can report, modify, and delete personal data to quickly comply with any consumer access request. Best yet, with Dathena's use of NLP, Deep Learning, and Machine Learning (ML), you will get the most accurate and scalable solution available on the market today, allowing you to achieve significant operational cost savings and immediate business value.



Source: Cisco 2019 Data Privacy Benchmark Study, n=3206

For more information on Dathena Privacy, click here to contact a Dathena sales representative.

References

1. https://rightsinfo.org/the-right-to-privacy-and-why-it-matters/
2. https://www.dlapiper.com/en/uk/news/2020/01/114-million-in-fines-have-been-imposed-by-european-authorities-under-gdpr
3. https://www.forrester.com/report/Protect+Your+Intellectual+Property+And+Customer+Data+From+Theft+And+Abuse/-/E-RES61476#endnote11
4. Although the statute becomes "operative" on January 1, 2020, the only enforcement of the statute as of that date relates to suits involving data security breaches. Effective July, the Attorney General can bring enforcement actions premised on any provision of the CCPA. Source: https://www.jdsupra.com/legalnews/ccpa-privacy-faqs-so-what-is-with-the-55886/
5. https://www.bankinfosecurity.com/californias-new-privacy-law-its-almost-gdpr-in-us-a-11149
6. https://digitalguardian.com/blog/what-california-data-privacy-protection-act
7. Forrester. (April 16, 2019) Protect Your Intellectual Property and Customer Data from Theft and Abuse. Executive Overview: The Data Security and Privacy Playbook
8. Ibid.
9. https://www.datanami.com/2018/11/27/global-datasphere-to-hit-175-zettabytes-by-2025-idc-says/
10. https://www.itsecurityguru.org/2017/01/25/62-companies-dont-know-sensitive-data-resides-report/
11. Gartner. (2019) 5 Areas Where AI Will Turbocharge Privacy Readiness.
12. SSR, DSAR, SAR, VCR are acronyms related to consumer requests regarding their personal data. Different privacy regulations tend to use different terms/acronyms.
13. Gartner, (2019) 5 Areas Where AI Will Turbocharge Privacy Readiness.
14. Ibid.
15. https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework_V1.0.pdf
16. https://analyticsindiamag.com/explainable-ai-fair-privacy-neural-networks/

Additional References

https://blog.seagate.com/business/enormous-growth-in-data-is-coming-how-to-prepare-for-it-and-prosper-from-it/

https://en.wikipedia.org/wiki/National_Institute_of_Standards_and_Technology

https://www.jdsupra.com/legalnews/ccpa-privacy-faqs-so-what-is-with-the-55886/

https://gdpr.eu/gdpr-vs-lgpd/

About Dathena

Dathena is a deep-tech company that brings a new paradigm to data privacy and security solutions. In a world of ever-growing information, regulation, and consumer privacy expectations, enterprises around the globe rely on Dathena to identify, classify, and control sensitive data, reduce risks, and enhance data protection framework.

Leveraging the power of modern AI technologies, Dathena delivers breakthrough, petabyte-scale solutions with unprecedented accuracy, efficiency, and speed that build consumer trust in a digital world and ensure the "privacy and data security protection journey."

Founded in 2016, Dathena continues to grow with its latest round of funding. With offices in Singapore, Bangkok, Geneva, Lausanne, Paris, and New York City, Dathena employs more than 70 people, including the world's top data scientists and information risk experts.

Contact details

www.dathena.io

hello@dathena.io

www.facebook.com/dathenascience

twitter.com/dathenascience

sg.linkedin.com/company/dathena-science