
Hey 👋

I'm Oliver Patel, author and creator of *Enterprise AI Governance*.

This week's newsletter presents, for the first time, my **PROTECT Framework for Managing Data Risks in the AI Era** © 2025.

Last week I published Part 1 of my 2-part series on the Top 10 Challenges for AI Governance Leaders in 2025. It focused on how the democratisation of AI, coupled with the sheer volume and rapid velocity of AI initiatives, is putting serious strain on the enterprise AI governance function. It outlined how, in response to these challenges, AI governance leaders must refine and update their risk-based approach and narrow the focus of their teams' work, to avoid being overwhelmed, distracted, or neglecting the most serious risks.

Although I promised Part 2 this week, I ended up writing a full article on the fourth challenge: **protecting confidential business data.** This is a hugely important topic and there was simply too much to say. The beauty of having your own Substack is that you can follow whichever creative or intellectual direction most appeals to you, rather than rigidly sticking to prior plans. I hope that you will indulge my partial detour and that you find value in this article for your work. The "real" Part 2, covering challenges 5-10, will have to wait another week.

## Challenge 4. Protecting Confidential Business Data

*How can enterprises protect confidential business data when there is immense hunger to*

*experiment with and use the latest AI applications that are released on the market?*

The generative AI boom has amplified and exacerbated a plethora of data risks that all enterprises are exposed to. It has never been more important to ensure that your AI governance, cybersecurity, and information security frameworks are designed to, and capable of, mitigating these risks.

However, designing, implementing, and scaling robust controls that actually protect your organisation's data and intellectual property requires precise understanding of what these risks are and how they are impacted by AI. That's where my PROTECT Framework comes in.

The PROTECT Framework empowers you to understand, map, and mitigate the most pertinent data risks that are fuelled by widespread adoption of generative AI. Below is a high-level summary of the framework, followed by a detailed breakdown of each of the 7 themes.

**PROTECT: Managing Data Risks in the AI Era**

The PROTECT Framework focuses primarily on protecting (*no surprises there*) confidential business data from exposure, disclosure, and misuse—as well as associated data privacy and security risks fuelled by AI. It also outlines how

organisations can use data in a compliant way, in the context of AI development, deployment, and use.

**P - Public AI Tool Usage**

**R - Rogue Internal AI Projects**

**O - Opportunistic Vendors**

**T - Technical Attacks and Vulnerabilities**

**E - Embedded Assistants and Agents**

**C - Compliance, Copyright, and Contractual Breaches**

**T - Transfer Violations**

The rest of the article breaks down each of the seven themes, highlighting both the core risks that enterprises are exposed to, as well as practical mitigations that can be implemented to manage and control these risks.

My forthcoming book, *Fundamentals of AI* Governance, provides a comprehensive visual overview of the PROTECT Framework, as well as a deep-dive on the Top 10 AI

Risks impacting the modern enterprise. To secure a 25% discount during the pre-launch period, sign up at the link below.

## P - Public AI Tool Usage

**Enterprise risks:** Use of publicly available AI tools is arguably the most severe data risk, because of how easy it is to do and how difficult it is to prevent. Simply put, there are thousands of publicly available AI tools that anyone can access via the internet, most of which are free or cheap. As well as mainstream generative AI chatbots like Claude, Gemini and ChatGPT, there are countless AI tools for creating presentations, managing email inboxes, transcribing meetings, and generating videos.

No matter how mature your enterprise AI capabilities are—and even if you are a frontier AI company— it is not going to be feasible to keep up with the latest and greatest AI tools and AI models that are released on the market each day, whilst also performing robust due diligence on AI vendors. Even with enterprise-grade licenses to mainstream generative AI services, you will not always get immediate access to the latest features included in the consumer version. This fuels immense hunger for employees to experiment with and use the most cutting-edge AI tools, irrespective of whether they are “internal” and approved or “publicly available” and unapproved.

This inability to keep pace with the market, coupled with a lack of awareness regarding the risks of using publicly available AI tools and the pressure that employees and teams are under to become “AI-first”, exacerbates the risks. Many employees may not understand the difference, from a data risk perspective, between using internal AI tools and public AI tools. But the risk is real. For example, when enterprise data is shared with publicly available AI tools, the organisation no longer has any control over what happens to it. This confidential business data could be used to train the AI models that power publicly available AI tools and in turn be disclosed, via future AI outputs, to competitors or malicious actors. It may even end up on the public internet—as we saw when various shared AI chat logs were indexed and publicly accessible online—or be retained indefinitely, as a result of court orders like the one OpenAI faced from the New York Court. Simply put, if you want to be in control of how your data and intellectual property is used, who has access to it, and for how long it is retained, your employees should avoid using publicly available AI tools.

**Practical mitigations:** Although shadow AI use is ubiquitous, there are various controls you can implement to mitigate this risk. My **3 Gs for Governing AI Democratisation** offers a useful starting point:

**Guidance:** educate and train the workforce on the risks of using publicly available AI tools and the importance of protecting confidential business data.

**Greenlight:** provide access to secure, best-in-class AI tools, platforms, and capabilities that are approved for internal use and can process confidential business data.

**Guardrails:** implement guardrails—including technical and legal mitigations—to mitigate outstanding data risks that the use of internally approved AI tools entails. This can include scanning and blocking certain types of data from being uploaded as an input or generated as an output.



## R - Rogue Internal AI Projects

**Enterprise risks:** Bypassing governance, especially when it happens at scale, creates compliance blind spots. Various risks emerge, and are difficult to manage, when teams develop and deploy AI systems, or procure AI solutions from vendors, without adhering to the mandatory AI governance, privacy, and cyber security processes.

In such scenarios, there is unlikely to have been any legal or compliance review, privacy assessment, or security evaluation. In turn, this means that genuine risks are

unlikely to be understood, required documentation and artefacts may not have been produced, and robust mitigations will not be in place to address any important risks.

This increases organisational technical debt, which can lead to costly and burdensome efforts to retrospectively re-engineer non-compliant AI systems that are already deployed in production. Finally, it increases the likelihood that data is used without authorisation, and in a manner that constitutes a contractual breach or potential compliance violation.

In most cases, this does not happen because of malicious internal actors or intentional rule-breaking. Rather, it is more likely due to enthusiasm, competitive internal pressures and competing priorities, or a lack of awareness of internal governance processes and how to navigate them.

**Practical mitigations:** To mitigate the risk of rogue internal AI projects bypassing compliance checks, a proportionate degree of oversight must be applied to all AI projects. The level of governance scrutiny and oversight should flex in relation to the type of data being used. The use of sensitive personal data, confidential business data, or copyright protected material should entail more rigorous oversight, both at the project outset and throughout the AI lifecycle. Oversight should also be more rigorous for scenarios that involve sharing data with external vendors and the applications they provide.

The most important thing you can do is to make your AI and digital governance processes as easy to navigate as possible, by integrating different processes where possible, providing accessible guidance and support, and using automation to streamline processes and improve the user experience.



## O - Opportunistic Vendors

**Enterprise risks:** Almost all enterprises must work with, and procure from, external organisations to progress with their AI ambitions. In the generative AI era, the trend is from build to buy. Whether it is leveraging pre-trained foundation models and generative AI chatbots, or working with vendors that provide bespoke AI products, exposure to third-party AI risk is unavoidable. In some cases, AI vendors and service providers may seek to use your confidential business data to train, develop, and improve their AI models and services—potentially without your explicit knowledge or consent.

The terms of service for many AI platforms and products are ambiguous or difficult to understand, and grant vendors broad rights over customer data. The key risk is whether your organisation's data is used to train AI models that other customers can access and use. If so, competitors (or any other organisation) using that same vendor's products and services may benefit from insights derived from your data. The risk is lower—or potentially fully mitigated—if your data is only used to train AI models and services that only your organisation has access to. This can enable you to benefit from feature improvements and customisation whilst mitigating data exposure and leakage risks.

**Practical mitigations:** Consider contractually prohibiting AI vendors from using your data (e.g., prompt, input, log, and output data), to train AI models and improve services that are accessible to other customers. This requires robust vendor due diligence, and a specific AI governance process pathway for AI procurement. It also requires clear guidance and training on third-party AI risks, acceptable data use terms, as well as template contracts and addendums that can be used across the business. Although the demand for AI vendors (and the products they provide) is high, the presence of opportunistic or shady operators in the market—and the immense value of the data they can obtain from enterprise customers—makes rigorous due diligence and contractual safeguards essential.

## T - Technical Attacks and Vulnerabilities

**Enterprise risks:** AI systems introduce novel attack vectors that are linked to the distinct vulnerabilities of these systems. In particular, generative AI and agentic AI systems can be compromised and exploited, leading to confidential data—that was part of the AI model's training, input, or output data—being extracted and stolen. Such data exfiltration is a known and widely documented AI vulnerability and can be caused by various attack methods. Prompt injection attacks, for example, are when an AI model is provided with malicious inputs (during inference) that are designed to manipulate and steer the outputs it generates, by jailbreaking model guardrails.

Simply put, the goal of prompt injection is to make the AI model do something that it is not supposed to. This includes, but is not limited to, data exfiltration, as well as reconstruction of training data. This risk of prompt injection is amplified with agentic AI, given the ability of AI agents to use tools and execute actions that can have a material impact (rather than "just" generate outputs for consumption).

**Practical Mitigations:** Although there are no foolproof mitigations against prompt injection attacks, there are nonetheless important steps you can take. Consider implementing AI system-level security controls and guardrails including input

validation, prompt sanitisation, output filtering, and incident monitoring and detection. Also, for high risk applications, conduct red-teaming and adversarial testing. Cybersecurity best practice emphasises the importance of multiple overlapping security layers. However, no technical controls can fully prevent prompt injection or data exfiltration. Therefore, carefully control who has access to sensitive AI systems, what data they can access, and what actions agentic AI can execute.

## E - Embedded Assistants and Agents

**Enterprise risks:** The AI assistants and agents that are increasingly embedded in the core workplace software we all use pose novel data risks. In particular, these embedded AI tools can inappropriately disclose and disseminate data to people and groups that were not supposed to have access to it. For example, a personalised AI assistant that summarises your emails and daily tasks can analyse wider organisational data that you have access to, such as document libraries and shared calendars. If that data has not been protected with appropriate file sharing permissions—and moreover has erroneously been made available to the entire organisation—then elements of it may be surfaced to you via your handy AI assistant.

Although the root cause of this is often inappropriate or inadequate file sharing

permissions, AI significantly increases the likelihood of such data being shared with the wrong person. It also provides malicious internal actors with a powerful tool for mischief.

AI meeting assistants and note-taking applications are of particular concern. It is commonplace to join calls with external organisations, only to find that a random AI bot is also on the call, recording and transcribing everything that is said. From an enterprise perspective, this is akin to uploading all of this information into a publicly available AI tool (which poses similar risks to those outlined in 'Public AI Tool Usage '), unless you have assurances from the external organisation regarding the technology they are using and how it processes your data. Furthermore, be wary of individuals having access to AI meeting recordings and transcripts of parts of the discussion they were not part of.

Increasingly autonomous agentic AI systems exacerbate these risks. In order to effectively determine the best course of action and use tools to execute tasks, LLM-based AI agents will need to mine, retrieve from, and synthesise myriad enterprise data sources. Establishing appropriate access controls, and maintaining AI agent audit trails, will become increasingly complex yet important.

**Practical mitigations:** Robust data governance and data risk management is critical to ensuring your increasingly autonomous AI tools do not cause havoc. Ensuring

appropriate file sharing permissions for sensitive and critical data sources is paramount, given the ways in which AI agents can mine through your document libraries and other repositories, as well as the obvious value this capability provides. Also, when deploying agentic AI, start with lower risk use cases, applications, and data sources. Furthermore, apply the principle of least privilege, to ensure that AI agents only have access to data that is necessary for their tasks.

## C - Compliance, Copyright, and Contractual Breaches

**Enterprise risks:** Data science, AI, and business teams are under significant pressure to leverage AI to deliver value for the business. To do so, they require seamless access to vast amounts of high-quality, business-critical data. However, the increasingly stringent data and AI regulatory landscape—particularly in the EU—creates numerous compliance risks when using data for AI activities. It is therefore crucial to have robust controls in place to prevent unauthorised or non-compliant use of data. The most important regulatory and legal domains to consider are:

**Privacy and data protection:** Privacy and data protection laws restrict the way in which personal data—in particular sensitive personal data—can be used. For example, under the EU's GDPR, you must have a lawful basis to process personal data. Personal

data processing is therefore only lawful if at least one of the following lawful bases apply: i) consent, ii) performance of a contract, iii) legal compliance, iv) protection of vital interests, v) performance of a task in the public interest, or vi) legitimate interests. Therefore, just because you have access to personal data, does not mean you are permitted by default to use it to develop or deploy AI.

**Copyright and intellectual property:** Organisations must be cautious when using external data for AI development and deployment, as it is often copyright protected. Different data sources come with different licenses, terms, and conditions. This cautiousness must extend to "everyday" employee use of generative AI tools—in particular their prompts and document uploads.

**Contracts:** Your organisation may have access to data that another organisation provided in the course of an engagement, such as the use of a product or service you provide, that is governed by a bespoke legal agreement. Therefore, you must protect and handle that data in accordance with the applicable legal agreement.

**AI-specific laws:** Finally, AI regulations like the EU AI Act typically include provisions that stipulate how data should be used in the context of AI activities. For example, the AI Act requires providers of high-risk AI systems to use high quality, accurate, and "representative" training, validation, and testing data sets, in order to mitigate bias risks and promote reliable AI performance.

**Practical mitigations:** The legal and regulatory domains outlined above are vast; comprehensive risk mitigation across all of them is beyond the scope of the PROTECT Framework. However, the overarching principle is that you must embed proportionate governance and oversight throughout the AI lifecycle, to prevent non-compliant or unauthorised data use. This means legal and compliance review must occur at critical stages, including before data is sourced, before AI models are trained, and before AI systems are deployed in production or released on the market. As ever, this governance must be complemented and reinforced by company-wide and role-specific training. When these governance checkpoints are bypassed—as discussed in the 'Rogue Internal AI Projects' theme—the aforementioned data-related compliance and legal risks materialise.

## T - Transfer Violations

**Enterprise risks:** Leveraging cloud-based AI services, such as platforms for accessing and using foundation models, almost always involves international data transfers. Given that AI processing is rarely confined to one jurisdiction, navigating international data transfer compliance is an important part of AI governance.

Privacy and data protection regimes worldwide restrict the way in which personal data can be transferred internationally. Under the GDPR, organisations can transfer personal data freely from the EU to entities in a non-EU country if there is an EU adequacy decision in place. An adequacy decision is the EU's way of "*protecting the rights of its citizens by insisting upon a high standard of data protection in foreign countries where their data is processed*". 15 jurisdictions are recognised as "adequate" by the EU. This includes the U.S.—but it only applies to commercial organisations participating in and certified under the EU-U.S. Data Privacy Framework. If there is no EU adequacy decision, alternative legal safeguards must be put in place (e.g., Standard Contractual Clauses), before personal data can be transferred from the EU to the non-EU jurisdiction.

On a separate note, the U.S. "Bulk Data Transfer Rule" prohibits or restricts organisations from transferring "U.S. sensitive personal data" and "government-related data" to "countries of concern", including China, Cuba, Iran, North Korea, Russia, and Venezuela. The Rule was issued by the Department of Justice in January 2025.

**Practical mitigations:** Although the above was far from an exhaustive overview of international data transfer regulations, the key point is that you must implement specific mitigations—as required by the applicable law—prior to transferring personal data across borders. For example, before onboarding U.S. AI vendors and service