Infosys | CONSULTING

VIEWPOINT



SOVEREIGN AI: HOW MUCH CONTROL DOES YOUR ORGANIZATION NEED?

Advanced AI depends on a small number of governments and businesses that could stop access at any moment. Businesses need a fail-safe sovereignty posture that's managed as an architectural and risk-management discipline.

Infosys | Knowledge Institute

In June 2026, the US government issued an export-control directive requiring Anthropic to suspend access to its Fable 5 and Mythos 5 models for foreign nationals, including those physically located in the US. The episode prompted criticism from AI policy and security observers and demonstrated how government action could interrupt access to an important AI capability.

China is considering similar actions. In July 2026, Reuters reported that Chinese authorities were discussing restrictions with leading Chinese technology companies on overseas access to China's most advanced AI models. In the same month as the Anthropic plug-pull, the European Commission presented its Technological Sovereignty Package, spanning semiconductors, cloud, AI and open-source AI models.

The proposal aims to reduce dependencies on the US and China and to strengthen European autonomy and resilience. Ursula von der Leyen, president of the European Commission, said that Europe could not afford to depend on others for technologies that support essential services and infrastructure.



External Document © 2026 Infosys Limited

A matter of control

It is clear then that who owns AI and the technologies that support it is now an important item on the agendas of governments and large corporations. It will only get more important as organizations move beyond proofs of concept and embed advanced AI into core processes, products, and decision-making.

> This has driven the rise of sovereign AI, which we define as the degree of autonomy and control an organization retains over the capabilities on which it depends, including its ability to maintain and change them.

This can cover everything from energy and compute to cloud, data, models, operations, applications, and talent.

That AI stack is unusually concentrated. At the infrastructure level, only a handful of companies dominate. For instance, ASML, a Dutch company that makes the machines used to manufacture the most advanced computer chips; TSMC, a Taiwanese chip foundry; and NVIDIA, the Silicon Valley darling that many feel is responsible for the AI boom, occupy critical positions in advanced AI hardware. Meanwhile, the hyperscale cloud infrastructure is dominated by a small handful of US providers, including Google, Amazon, and Microsoft. At the same time, the most capable frontier models are in the hands of US developers like OpenAI and Anthropic. These models increasingly exist alongside a more geographically diverse open-weight ecosystem across the US, Europe, and China, including Llama, Qwen, and DeepSeek. Even so, the pool of companies and countries from which an organization can select its tools remains vanishingly small.

This concentration risk expands the scope of concern for the average IT governance function. Questions about where data is held and who can access it are now further complicated with questions about who controls the compute, cloud, models and other AI capabilities on which many businesses are betting their future growth and productivity.

To get ahead, organizations need to assess where data, models, processing and supporting infrastructure reside, along with which legal regimes can exercise authority over them. They must also ascertain which operational and governance decisions remain with the organization and how readily a component could be replaced if circumstances changed. This assessment then determines how exposed the organization is to concentration risk and helps devise the target sovereignty posture by building resilience into different layers of the organization's AI stack.

Not all plain sailing

However, being able to control levels of autonomy is only one half of the equation. The other is ensuring that you don't spend too much, and that AI capability isn't lost in the calculus.

For example, an organization may try to reduce concentration risk by using AI tools from a wider or more diverse set of vendors and geographies. Or they may use more open-source alternatives that they can control. But this approach could be more expensive to operate since the organization would need to provision its own servers and graphics processing unit (GPU) clusters, spend time and money on fine-tuning models, and patch vulnerabilities. It would also forgo the benefits of scale that frontier models possess, and this could reduce the capability of the AI tools.

The EU's Technological Sovereignty Package may help here by improving the environment for Europe's semiconductor industry. Specifically, the Chips Act 2.0 targets the investment and competitiveness, while strengthening semiconductor supply chains and boosting domestic demand. The Cloud and AI Development Act addresses AI capacity for industry, while the EU Open Source Strategy aims to make portable, lower-cost open-weight options easier to adopt.

At the same time, there are also ingrained dependencies between energy, compute, cloud, data, models, operations and applications to think about, especially as agentic AI processes get more embedded in the organization. A disruption that previously affected an isolated analytical tool might now interrupt customer service across the enterprise, or even affect operational decision-making and transactions. Full-stack control is unlikely to be worth the cost outside a limited number of strategic or sensitive use cases.

External Document © 2026 Infosys Limited

Applications and AI agents become optimized around particular model behaviors, application programming interfaces, tool-calling conventions, platform services, security controls and operating processes. Replacing a provider can require substantial re-engineering even where alternative providers exist.

> Because of these dependencies, even with cheaper and easier government-mandated options, moving AI infrastructure in-house is a challenging engineering issue.

Because building or owning AI can increase cost and complexity while reducing capability, organizations must decide where switching layers of the AI stack for something more controllable represents a reasonable commercial trade-off, and where it creates a strategic vulnerability.

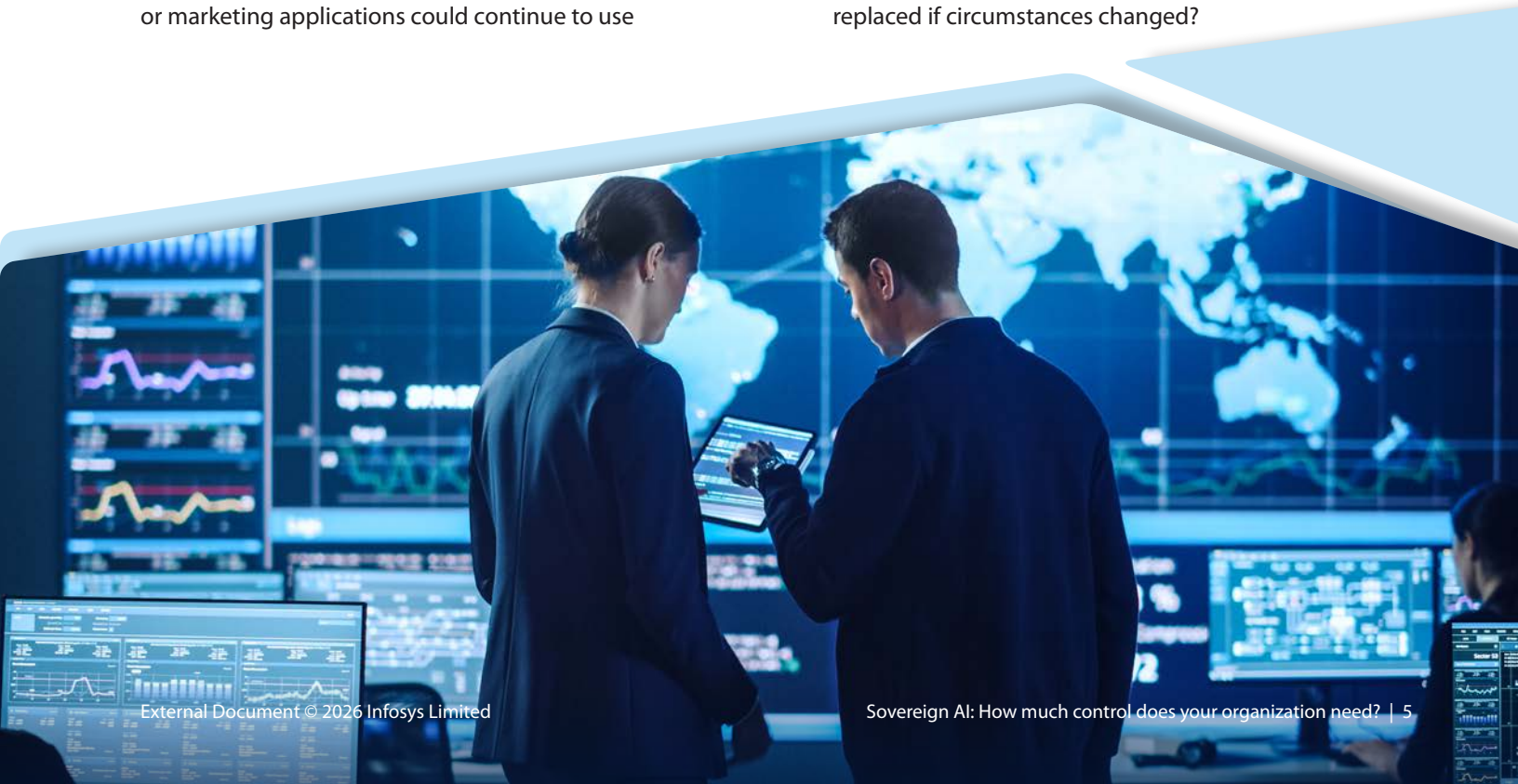
How to get sovereignty right

The answer to the control, cost, and capability conundrum is to treat sovereignty as both an architectural and risk-management discipline, even as EU policy widens the options.

In practice, achieving the target sovereignty posture means creating a tiered architecture in which control increases with the sensitivity and criticality of the workload. For instance, generic productivity, coding, or marketing applications could continue to use global frontier models and software as a service. Mainstream enterprise AI can also often operate on regional hyperscaler infrastructure with appropriate residency, encryption and access controls. More sensitive customer, employee, industrial, or regulated workloads might justify sovereign cloud, customer-controlled encryption keys, and local inference or portable open-weight models. Selected defense, intelligence, and critical-infrastructure workloads might require nationally controlled infrastructure and highly isolated environments.

To achieve the correct balance of control, cost, and capability, organizations need to first assess each layer of their AI stack, including talent; applications and agents; operations; models and inference; data; cloud; compute; and energy, against a consistent set of questions, including:

- **Location:** Where is the capability physically or logically located?
- **Jurisdiction:** Which governments and legal regimes can exercise authority over it?
- **Access:** Who can access the capability, data or underlying systems?
- **Control:** Which decisions remain under the organization's direct control?
- **Substitutability:** How readily could the component be replaced if circumstances changed?



External Document © 2026 Infosys Limited

In Infosys Consulting's work with clients, each area is scored on a common 1-5 maturity scale. A score of 1 represents an exposed position with limited visibility or meaningful control; 3 means controls are defined, implemented and owned; and 5 means controls are demonstrably effective, independently verifiable and regularly tested. Importantly, a 5 does not automatically mean full domestic ownership or even that AI systems are isolated from undue risk. Rather, the required level depends on the workload's business criticality, regulatory obligations and risk exposure.

The scoring is evidence-based. Contractual commitments provide less assurance than technical configurations, audit logs, control testing or independent assurance, so the current score is adjusted for evidence quality. That evidence-adjusted score is then compared with a required target score for the workload, and the resulting gap shows where sovereignty is insufficient. Those gaps are considered alongside criticality, external exposure and specific red flags, such as the need for more advanced capabilities, unknown jurisdiction, unverifiable privileged access, inability to intervene, or lack of a credible exit route. It is then determined whether the risk should be accepted, mitigated or escalated (Figure 1).

Figure 1. Complete sovereignty is rarely the answer

Assess sovereignty layer by layer: location, jurisdiction, access, control and portability.

The AI stack	What must remain controllable
1 Talent and know-how	People who can run, evolve, and rebuild the capability
2 Applications and agents	Workflows, integrations, stopping rules, and tool boundaries
3 Operations	Monitoring, security, recovery, rollback, SLAs and evaluations
4 Model inference	Runtime priority, availability, quotas, and behavior changes
5 Data	Residency, ownership, movement, lineage, retention and deletion
6 Cloud services	Orchestration, keys, legal framework, and control-plane access
7 Compute and data centers	Hardware supply, software ecosystem, and facility capacity
8 Power and energy	Grid access, pricing, backup, water and cooling

Choose control where it matters

1 **LOCATE** Where is risk embedded?

2 **DECIDE** Accept it or mitigate it?

3 **TRADE** What is control worth?

Source: Infosys Consulting

External Document © 2026 Infosys Limited



Organizations can then decide the sovereignty required for different workloads across the AI stack and establish a target sovereignty posture, roadmap, and business case.

The target sovereignty posture determined by the assessment is a continuum rather than a binary choice. Where greater control is required, clients can deploy open-weight models on infrastructure they control. For example, in the Middle East, Infosys Consulting supported a client in building a meeting intelligence capability using Qwen 3.5, an open-weight model deployed in-country. This enabled the client to retain control over both sensitive data processing and model deployment, while also addressing local language and cultural requirements.

For particularly sensitive government workloads, the posture can go further. For instance, in work with another client, Infosys Consulting built an enterprise intelligence platform for confidential documents using secure, in-batch processing, minimizing the movement and external exposure of sensitive information. Even an organization running an open-weight model on client-controlled infrastructure may retain dependencies further down the stack – for example on GPU suppliers, firmware, software components, and potentially energy and infrastructure providers.

> The practical test of sovereignty is the extent to which an organization retains credible alternatives when things go wrong, and which external dependencies the organization is prepared to accept.

Resilience thus depends on designing critical components so they can be changed before individual dependencies become strategic constraints. This is the poly-AI approach, a platform-native architecture where models, compute, and other capabilities can be switched in and out as business requirements change. Sovereign AI takes this one step further, adding risk-adjusted cost factors, along with portability, leverage, and flexibility.

What to do now if you're a leader

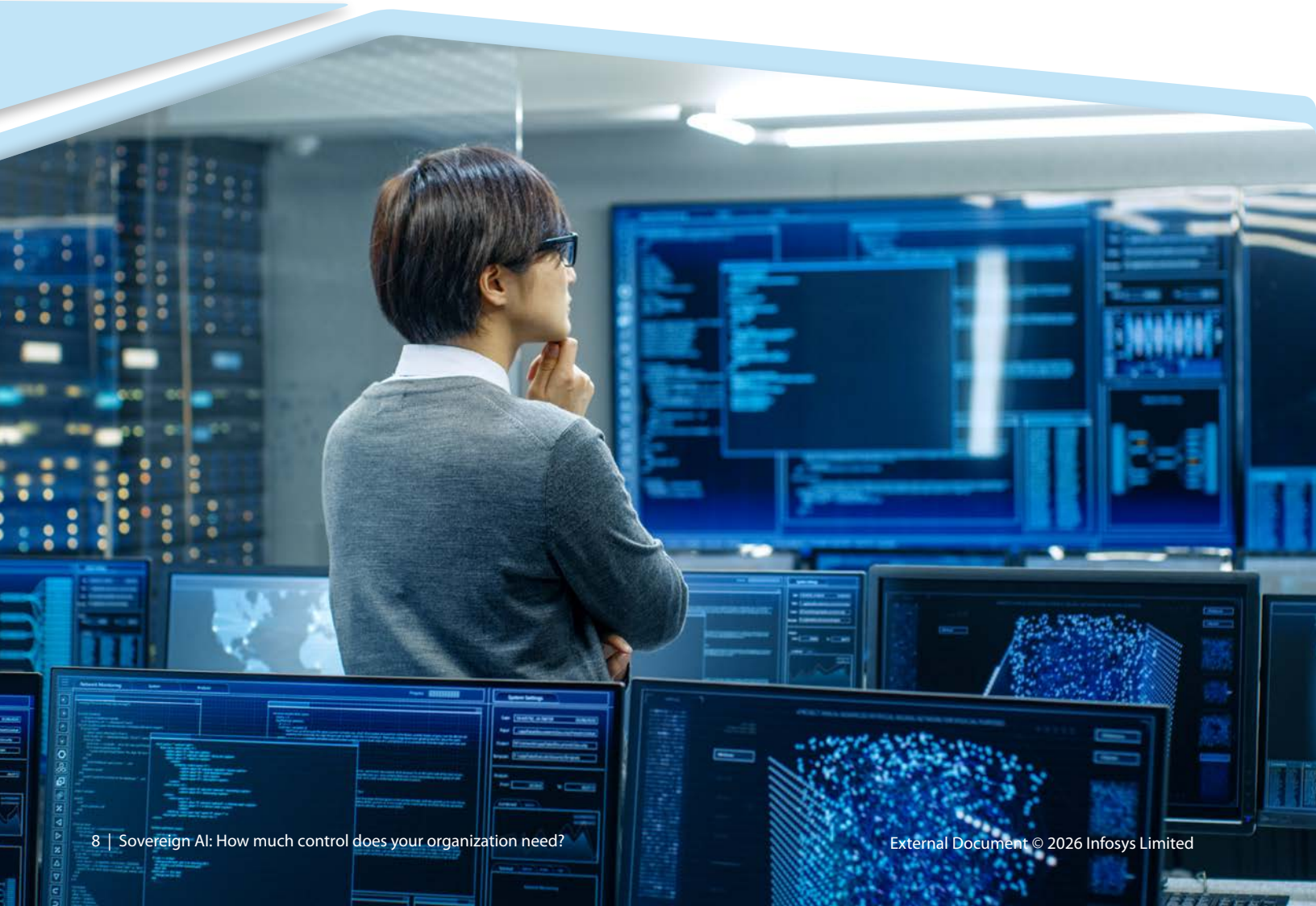
Organizations should establish a baseline for their most important AI capabilities, and then consider five things:

1. **Identify critical workloads:** Following the right-fit assessment, critical workloads can be identified.

2. **Map dependencies:** The next step is to identify their dependencies across the AI stack. For each dependency, organizations should understand the concentration risk, jurisdictional exposure, switching costs, and the operational consequences of withdrawal or disruption.

3. **Define a target posture:** Leaders should then define a target sovereignty posture based on business criticality, regulatory obligations, risk appetite, and the strategic importance of the capability. Different layers and workloads will warrant different levels of control. As mentioned, the posture is a continuum – where at one end organizations host everything within their own private cloud for control over data residency, network isolation, access, and security, and at the other they could consume a managed model through a hyperscaler. Investments should concentrate

> The question is no longer whether to adopt AI, but how much control organizations need over the capabilities their business cannot afford to lose. Sovereignty is becoming a strategic resilience decision, not just a technology choice
>
> **Neil Thomas**
>
> *EMEA head, Infosys Consulting*



External Document © 2026 Infosys Limited



on AI dependencies whose interruption, withdrawal, or change in jurisdiction would create significant strategic, operational, or regulatory exposure.

The transition doesn't need to come all at once and can be iterative. Actions could include taking greater control of selected elements of the AI stack, while at the same time increasing portability between providers, diversifying suppliers, retaining alternative deployment options, and reducing dependencies that would be particularly difficult to unwind, such as data pipelines with one cloud provider or embedded agentic workflows with vendor-specific tool integrations. Architecture and procurement decisions should explicitly consider substitutability and exit alongside cost, capability, and time to market.

4. **Test how quickly alternatives can be activated:** An architecture diagram showing that a second model, cloud, or infrastructure provider could theoretically be used provides limited protection unless the alternative can be activated within an acceptable timeframe. Periodic migration or substitution exercises can reveal dependencies that would otherwise remain hidden until a disruption occurs.

5. **Evaluate options on a risk-adjusted basis:** Finally, the economics of sovereign AI should be assessed on a risk-adjusted basis. For example, a cheaper AI provider might look like the obvious choice on cost alone, but if it carries higher concentration risk, weaker governance, or greater geopolitical exposure, it could be the worse choice on a risk-adjusted basis. In the same way, a more portable or diversified architecture could carry a higher business-as-usual cost, but the relevant comparison also includes the potential cost of service interruption, forced migration, regulatory remediation, and excessive supplier leverage. Selective sovereignty investment should therefore be treated as part of organizational resilience, where organizations maintain sufficient control and credible alternatives for risks that may be relatively unlikely, but potentially high impact.

How to move ahead

As AI becomes more critical for enterprise operations, the strategic question for leaders is clear. Where is dependence acceptable, and where does the organization need enough control, portability, and alternative capacity to remain resilient when technology, regulation, or geopolitics change? As the Economist has argued, reliance on just a few providers of AI hardware and software isn't tenable given the criticality of the processes involved in industries such as energy, supply chain, and telecommunications.

For organizations, it also means revisiting substitutability scores as the repercussions of the European Commission's Package are felt and options mature significantly, including these changes in periodic substitution tests, and reflecting them in risk-adjusted business cases.

However, as MIT's State of AI in Business 2025 found, trying to do too much themselves is also a risk: the report found that external partnerships had twice the success rate of internal builds.

Organizations should know when to switch and when to stand still. That requires a sovereignty posture that reduces risk and ensures competitiveness as they go agentic AI-first.

> The true measure of AI readiness is not how quickly organizations can deploy new capabilities, but how resilient those capabilities remain when conditions change. Leaders must build AI foundations that combine innovation, control, and long-term adaptability.
>
> **Neil Thomas**
>
> *EMEA head, Infosys Consulting*

Authors

Neil Thomas
EMEA Head, Infosys Consulting

Phil Benton
Partner, Infosys Consulting

Harry Keir Hughes
Infosys Knowledge Institute

Analysis and production

Kate Bevan
Infosys Knowledge Institute, London

Pragya Rai
Infosys Knowledge Institute, Bengaluru

External Document © 2026 Infosys Limited

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling throught leadership. Our researchers and subject matter experts provide a fact base that aids decision making on critical business and technology issues.

To view our research, visit Infosys Knowledge Institute at infosys.com/IKI or email us at iki@infosys.com.

About Infosys Consulting

Infosys Consulting is global management consulting firm helping some of the world's most recognizable brands transform and innovate. Our consultants are industry experts that lead complex change agendas driven by disruptive technology. With offices in 20 countries and backed by the power of the global Infosys brand, our teams help the C-suite navigate today's digital landscape to win market share and create shareholder value of lasting competitive advantage.
www.infosysconsulting.com

For more information, contact askus@infosys.com

Infosys
Navigate your next

© 2026 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.