Enterprise-Readiness of OpenAI vs Anthropic vs Microsoft 365 vs Google Workspace

Executive Summary

OpenAI's **ChatGPT** and Anthropic's **Claude** have rapidly added enterprise-oriented features, but they still trail the mature security and compliance ecosystem of Microsoft 365 and Google Workspace. OpenAI now offers **ChatGPT Enterprise** (and a smaller-scale **Team** plan) with strong data ownership commitments – **no use of business data for training by default**, SOC 2 certification, and admin controls. Anthropic's **Claude for Work/Enterprise** similarly guarantees **no training on customer data** and boasts **SOC 2 Type II** and **ISO 27001** certifications. Both AI providers encrypt data in transit and at rest and will sign Data Processing Agreements (DPAs) to support GDPR/NZ Privacy Act compliance. However, limitations remain: **data residency options are currently lacking** (data is generally processed in the US), and **enterprise customers must trust the providers' internal controls** for isolation and deletion of sensitive content.

By contrast, **Microsoft 365 E5** and **Google Workspace Enterprise** are **battle-tested for enterprise security and compliance**. They **never use customer data to train foundation AI models** and offer extensive compliance attestations (ISO 27001, SOC 2, **FedRAMP**, **IRAP** etc.) and customer controls (regional data storage, customer-managed encryption keys, granular audit logs). Microsoft and Google's generative AI features (Copilot and Duet AI/Gemini) explicitly uphold the same data protections – user prompts and outputs **stay within the tenant** and are not used to improve base AI models.

Confidence Assessment: Today, an enterprise can achieve a **moderate to high level of assurance** with OpenAI or Anthropic by leveraging their enterprise plans and contractual commitments. Both have made clear legal promises and obtained key certifications, which inspires **reasonable confidence** for many use cases. However, certain gaps (e.g. **indefinite data retention on free tiers**, lack of local data centres, fewer built-in compliance tools) mean risk-sensitive organisations may rate their assurances as only *moderate* without additional controls. In contrast, Microsoft 365 and Google Workspace offer **very high assurance**, given their long-standing compliance regimes and comprehensive customer controls. In summary:

- **OpenAI (ChatGPT Enterprise/Team):** High data-control guarantees (no training, 30-day deletion) with SOC 2 compliance – **High confidence for general enterprise use**, but **moderate** for highly regulated data due to data residency and evolving certifications.
- **Anthropic (Claude Enterprise):** Strong privacy stance (no training, customisable retention) and broad certifications (SOC 2, ISO 27001) – **High**

confidence, slightly ahead of OpenAI on compliance, but also **moderate for strictly regulated sectors** (newer platform, US-centric hosting).
- **Microsoft 365 E5:** Mature, compliance-aligned across industries (GDPR, Privacy Act, HIPAA, etc.), with rich security features – **Very High confidence**.
- **Google Workspace Enterprise:** Similarly **Very High confidence**, with proven data protections and transparency (verified by independent audits).

Next, we follow with detailed findings across data usage, security controls, legal compliance, risk mitigation, and common misconceptions, followed by a comparative scorecard and practical toolkit

Disclosure

This report was prepared and verified by subject matter experts using traditional research, augmented with AI research, specifically ChatGPT and Perplexity Deep Research.

All information is current as at May 9th 2025.

INDEX

1. Data Usage Commitments

OpenAI – ChatGPT (Free/Plus vs. Enterprise/Team): OpenAI's data handling policies differ significantly between consumer and enterprise services:

- **Free and Plus:** User prompts and chat content *may be used* to train and improve OpenAI's models unless the user opts out. By default, ChatGPT for individuals logs conversations and can leverage them for model training. OpenAI introduced a "history disable" feature in 2023 to let users prevent training use, wherein conversations were retained only 30 days for abuse monitoring. However, as of April 2024, free/Plus users can no longer disable chat history; OpenAI now **retains all prompts indefinitely for non-business users** (though users can still opt out of *training* via a setting). In practice this means personal ChatGPT data is stored unless deleted by the user, and is not used for model training if the user opts out (OpenAI confirms it "won't train on data when opted-out"). *Carve-out:* OpenAI may still analyse and retain data for abuse prevention or legal compliance. For example, content that violates policies may be reviewed and kept longer to improve safety systems.
- **Enterprise/Team/Edu:** By contract and design, OpenAI's business plans **do not use customer inputs/outputs to train models by default**. The customer **"owns and controls"** their data: inputs and outputs remain property of the user (to the extent allowed by law). Data from ChatGPT Enterprise/Team is segregated and **excluded from training datasets** automatically. OpenAI only would use such data if a customer explicitly opts in (e.g. via a feedback-sharing program). In Enterprise, **workspace admins can set a data retention period** for chat history. Conversations can be deleted by users or admins; any deleted content is **purged from OpenAI systems within 30 days** (except backups kept as required by law). By default, ChatGPT Enterprise retains conversations to enable features like history, but an admin can shorten retention or turn off history, giving similar effect to "no retention" beyond 30 days. For **ChatGPT Team** (for SMBs), each end-user can decide whether to save chats; unsaved or deleted chats are also removed within 30 days. Internally, OpenAI restricts access to stored business data – only authorised personnel or contractors can access it for **support, abuse investigation, or compliance** purposes. **Data segregation:** OpenAI's multi-tenant architecture ensures that one customer's fine-tuned models or data are not visible to others; any custom models you train with your data are **"yours alone and not shared"**.
- **API usage:** Similar to enterprise, data sent via the OpenAI API (after March 1, 2023) is not used for training by default. API inputs/outputs are retained for **30 days** for service provision and abuse detection, then deleted. OpenAI offers a **"zero data retention"** option for certain use cases upon request, meaning API data would not be stored at all beyond processing the request. This was a key factor for enterprises like Morgan Stanley, who noted OpenAI's

"zero data retention policy" as critical to keeping proprietary data private. In all cases, OpenAI uses data only to the extent necessary to provide the service and enforce policies, not to improve models (absent opt-in). *Note:* OpenAI does perform automated scanning on inputs (and outputs) for security (malware, abuse content) which could involve processing content through internal classifiers – this falls under legitimate use for abuse prevention.

Anthropic – Claude (Free/Pro vs. Enterprise): Anthropic has taken a similarly strict stance on customer data usage:

- **Free and Claude Pro (consumer)**: By default **Claude does *not* use your conversation data to train Claude's model**. Anthropic's policy explicitly states: *"We will not use your inputs or outputs to train our models, unless you've explicitly reported them to us via feedback or explicitly opted in"*. This means casual Claude users' chat data isn't scraped into future model updates, addressing a major privacy concern. If a user clicks the thumbs-up/down feedback or submits a bug report, that specific conversation may be collected (stored up to **10 years** in a separate feedback database) and could be used to improve Anthropic's systems or model. Even then, Anthropic **de-identifies feedback data** (removing user identifiers) before using it for model training or analysis. Additionally, if a conversation is flagged for violating the usage policy (e.g. hateful or illicit content), Anthropic may review and **retain it to enhance their safety systems** and content filters, including possibly training internal **"safety models"** (not the main Claude model) on such flagged data. Non-flagged conversation data on the consumer Claude platform is typically deleted from active systems after a period. (Anthropic has indicated a **90-day retention** for normal Claude chats on the backend, after which they are deleted, barring the exceptions above. This 90-day window allows the user to retrieve recent chat context and enables abuse monitoring, but data isn't kept long-term or reused for AI training.)
- **Claude for Work / Enterprise:** All **commercial offerings (Claude Enterprise, Claude API)** come with a **"no-training by default" guarantee** – Anthropic **will not use enterprise inputs/outputs to train Claude** absent explicit permission. This matches OpenAI's approach. If enterprise users submit feedback or opt in via a program, that data could be used to improve models, but admins even have the ability to **disable the feedback submission feature** for their organisation to prevent any inadvertent data sharing. Anthropic's standard data retention for Claude Enterprise/API is that inputs and outputs are **automatically deleted after 30 days** from their systems, unless otherwise agreed. (They note that enterprise customers can negotiate **custom retention timelines**, with 30 days as a minimum – e.g. some may choose 7 days or 90 days retention according to policy). Like OpenAI, Anthropic retains flagged **policy-violating content** longer: any prompt flagged by their Trust & Safety classifiers can be kept for up to **2 years** (and

associated safety metadata for 7 years) to improve detection and for legal compliance. Enterprise admins can delete conversation data through the Claude interface, which removes it from user view immediately and ensures backend deletion within 30 days. **Data segregation:** Claude for Work provides an isolated workspace for an organisation. Anthropic does not use one customer's data to assist another, and any **documents or "Projects"** you upload for context (Claude Enterprise allows providing a knowledge base to the AI) remain accessible only within your organisation's tenant. They advertise that *"Anthropic does not train our models on your Claude for Work data"* clearly as a value proposition.

- **Claude API:** The Claude API, like OpenAI's, by default retains data for 30 days for abuse monitoring, then deletes it. Anthropic offers a **Zero Data Retention agreement** for eligible customers/use-cases, meaning no conversation content is stored beyond processing. They also will sign **Data Processing Addenda** making them a data processor under GDPR/NZ law, which contractually binds them to delete or return data as instructed. Anthropic's privacy centre confirms that for commercial users, Anthropic acts as a **Data Processor** for customer-provided personal data (with the customer as Controller), whereas for Claude Free, Anthropic would be a Controller for any personal data individuals input.

Microsoft 365 (E5) and Google Workspace (Enterprise): Both Microsoft and Google have very explicit data use commitments, which set the high bar for cloud services:

- **No training on customer data:** Microsoft has **publicly affirmed** that *"Microsoft does not use customer data from Microsoft 365 (commercial or consumer) to train foundational large language models"*. Rumours in 2024 about Office data being used for AI were refuted – the data from Word, Excel, Outlook, etc. is only used to deliver the service to that customer, not to feed into GPT-4 or other Microsoft AI models. Similarly, Google states unequivocally that **Workspace content is not used to train** Google's broad AI models (like Bard or Gemini) *"without your permission"*. In fact, Google's generative AI features for Workspace (e.g. Duet AI writing assistance) run on models that are not updated with your specific inputs. Your Google Docs, Gmail, or other content stays within your tenant's boundary. Both companies make these privacy commitments part of their terms: the **Workspace AI Privacy Commitments** emphasise *"Your data stays in Workspace"* and is not shared or used for improving AI outside your own use, and Microsoft's Online Services Terms similarly restrict use of Customer Data strictly to providing the services (and troubleshooting, etc.), with no advertising or secondary use.
- **Data retention & ownership:** In Microsoft 365 and Google Workspace, **customers retain ownership of their content** (emails, files, chat logs). Data is retained as long as the customer account exists or as governed by admin

policies. Enterprises can configure retention policies (e.g. auto-delete emails after X years, or retain files in litigation hold) – in other words, **retention is under the customer's control**. If a user deletes data, it's removed from the active system and then from backups per a known schedule (Microsoft and Google typically have transparent data deletion timelines in their DPAs). **Data residency:** Microsoft and Google offer data centre location options – e.g. Google Workspace allows an admin to choose **data region** for certain data at rest (such as EU or US), and Microsoft 365 has regional tenancy (with **New Zealand data centres now available** for NZ customers, or fallback to Australia if selected). None of that data leaves the agreed regions for storage and processing, except as needed for redundancy and as permitted by contract. This means an NZ enterprise can ensure Office 365 data stays in-country (or in-region) to meet sovereignty needs, an option not yet offered natively by OpenAI/Anthropic.

- **Carve-outs (abuse, support):** Microsoft and Google do perform automated scanning on customer content for malware, spam, or policy violations (to protect users). For example, Gmail scans attachments for viruses, and Exchange Online does the same. These scans could be considered "using" the data, but solely for security – not to profile the user or improve general AI. If illicit content (e.g. CSAM) is detected, providers may have legal obligations to retain/report it. Additionally, if a customer raises a support ticket, support engineers may access content with strict controls and only with permission. These are standard practices aligned with privacy laws (and typically covered in the DPA and trust documentation). Both Microsoft and Google also allow customers to opt out of even service improvement telemetry in many cases.

In summary, **OpenAI and Anthropic's enterprise offerings have converged toward the industry standard set by Microsoft/Google**: data is **owned and controlled by the customer**, not used to train models or for any external purpose by default, and deletions are honoured. The main differences are in implementation details (e.g. default retention windows) and maturity of controls for the customer to self-manage data. Enterprises should **sign the DPA** with these vendors to cement these commitments and should use enterprise-specific plans (not free accounts) for any sensitive data, because **free consumer AI services do not provide the same level of data control or deletion guarantees**. Misconceptions that ChatGPT or Claude will indiscriminately learn your secrets are **no longer accurate** when using the proper enterprise services – the contracts clearly forbid such use. That said, due diligence (as discussed in section 5) is still crucial to verify these commitments meet your specific regulatory needs.

---

2. Security Controls and Certifications

Encryption and Data Security: All four platforms enforce strong encryption for data in transit and at rest. **OpenAI** confirms that all data is encrypted **at rest (AES-256)** and **in transit (TLS 1.2+)** between the user and their servers. Anthropic similarly uses industry-standard encryption (given they have ISO 27001 certification, encryption of sensitive data at rest is a requirement). While Anthropic's documentation doesn't explicitly call out the ciphers in marketing material, one can infer **AES-256 at rest and TLS 1.2/1.3 in transit** are in place, as these are table stakes for SOC 2 and ISO certification. **Microsoft 365** and **Google Workspace** also use robust encryption: TLS for data in motion, and encryption at rest in their data centres (e.g. Google uses AES-256 by default for storage, and Microsoft uses BitLocker and Azure SSE across services). Both also offer optional **customer-managed encryption keys** for certain data at rest (for instance, Microsoft's Customer Key feature in M365 E5 allows an organisation to supply its own root encryption keys for Exchange Online and SharePoint content, adding an extra layer of control). Google Workspace similarly has **Client-side encryption (CSE)**, where the customer controls the keys via a third-party key service, ensuring Google's servers never see decrypted content (useful for very sensitive documents). Neither OpenAI nor Anthropic currently offer customer-managed keys for encryption – data is encrypted, but the provider manages the keys. Enterprise customers must trust OpenAI/Anthropic's internal key management and access controls.

Identity and Access Management: OpenAI Enterprise and Anthropic Enterprise integrate with corporate identity systems for authentication. **ChatGPT Enterprise/Team** supports **SAML Single Sign-On**, enabling integration with Azure AD, Okta, etc., so that only authorised employees can access the AI, and their accounts can be centrally managed. Anthropic's Claude Enterprise likewise offers **SSO and domain-based access** (they mention "single sign-on (SSO) and domain capture" to manage user access centrally). Role-based access control is emerging – Anthropic Enterprise has **role-based permissioning** to designate workspace owners, admins, etc.. OpenAI's Enterprise admin console allows setting which users or groups have access and possibly controlling features (for example, an admin could disable plug-ins or browsing if those pose risk). In contrast, Microsoft and Google have fully-fledged identity and access management: Azure AD/Entra ID and Google Identity let admins enforce multi-factor auth, conditional access policies (e.g. geolocation or device trust requirements), and integrate with on-prem directories. **All four platforms support Multi-Factor Authentication (MFA)** either via their own account system or through SSO.

Audit Logging: This is an area where Microsoft and Google are very mature, and OpenAI/Anthropic are making progress. **Anthropic Enterprise** is rolling out **Audit Logs** for Claude – allowing organisations to track usage and actions within their

Claude workspace. This would include logging prompts, responses, or at least metadata (which user used the assistant and when), aiding in security monitoring and compliance. **OpenAI ChatGPT Enterprise** likewise promises enhanced logging: OpenAI states that business products support "enhanced visibility and fine-grained controls". In practice, this likely includes an admin dashboard to review organisation-wide usage, and possibly an API or SIEM integration to export logs of prompts. (OpenAI's documentation doesn't explicitly detail the logs, but one can expect at least timestamps, user IDs, and possibly prompt metadata are logged for admin review.) Microsoft 365 and Google Workspace provide **extensive audit capabilities** – admins can see user activities like file accesses, email sends, login attempts, etc., in unified audit logs. For example, M365's audit log (compliant with FINRA, etc.) can record every time a user accessed a SharePoint file or sent an email, and these logs can be retained for many years and exported. For their AI features specifically, Microsoft's Copilot will have auditing such that you can trace what content it accessed to generate an output. Google's Workspace admin audit logs likewise track activity, and Google offers **Access Transparency** logs for certain admin actions (so a customer can see if Google support engineers accessed their data). In comparison, OpenAI and Anthropic might not reach that granularity yet (e.g. you might not get a log of every single prompt and answer in plaintext via an API without requesting it), but they do limit their own staff's access to data (no one at OpenAI should be viewing your prompts unless needed for abuse investigations, and any such access would be by authorised personnel only). Both OpenAI and Anthropic maintain internal **audit trails** of who in their company accesses customer data. They have 24/7 security teams that monitor for any unauthorised access attempts.

Incident Response and Monitoring: OpenAI reports having an on-call security incident response team available 24/7/365. They also run a public **bug bounty program** to catch vulnerabilities. Anthropic, through its compliance, likely has a formal incident response plan as well (e.g. they mention regular security assessments and employee security training). Microsoft and Google of course have dedicated incident response teams and will also notify customers of incidents per regulatory requirements in their contracts (e.g. GDPR 72-hour personal data breach notifications via the DPA). None of OpenAI/Anthropic's known materials mention a history of breaches to date. One early issue for OpenAI was a **March 2023 bug** where a caching issue exposed some users' chat titles and possibly parts of others' chat content to unrelated users – that was a security incident that OpenAI disclosed and fixed promptly (and it led them to implement more robust data isolation in retrieval). Such incidents show that OpenAI is still maturing, but they responded with fixes and even paused service to investigate. Enterprises should obtain the vendor's **Security Whitepaper or SOC 2 report** (under NDA) to review details of their incident management and infrastructure security.

Certifications and Standards:

- **OpenAI** has completed **SOC 2 Type II** audits for ChatGPT Enterprise, Team, and API. This audit by a third party validates that OpenAI's security controls (access control, change management, etc.) meet the AICPA Trust Services Criteria for Security and Confidentiality. OpenAI is also listed in the **Cloud Security Alliance (CSA) STAR** registry at Level 1 (self-assessment) or Level 2 (third-party audited). (Their site shows CSA STAR Level 1 achieved). They claim alignment with **GDPR** and CCPA internally, and they offer a DPA. There is mention that OpenAI "supports customers' compliance with privacy laws… and offers a DPA" – but note that GDPR compliance for OpenAI is still under scrutiny in the EU; they have improved transparency to users and allowed data controls, which are steps in the right direction. OpenAI has not announced ISO 27001 certification publicly as of early 2025 (though some reports suggest they are working towards it, possibly aligning with **NIST 800-53** and FedRAMP moderate controls). They have not yet achieved **FedRAMP** authorisation on their own. For US government use, OpenAI relies on Microsoft's Azure OpenAI Service, which **is FedRAMP compliant via Azure** infrastructure. (If a US govt agency wants GPT-4, they would use it through Azure's Gov cloud, not directly via ChatGPT.) For **IRAP (Australia) or NZ ISM**: OpenAI hasn't specifically listed those, again leaving it to partners (Microsoft has IRAP assessment for Azure/OpenAI).
- **Anthropic** shines in certifications: They have **SOC 2 Type I & II** reports, **ISO 27001:2022** (information security management) and even the new **ISO/IEC 42001:2023** AI Management System certification – indicating they have a governance system for AI in place (one of the first to get ISO 42001). They also label Claude as **HIPAA-ready ("HIPAA configurable")**, meaning they can support HIPAA compliance and will sign BAAs for healthcare use (with the customer maintaining responsibility to input only allowable PHI). Anthropic likely complies with PCI-DSS standards for their own corporate systems where applicable, but using Claude to handle credit card data is not an intended use case (PCI certification is not listed among their achieved compliance items). Anthropic has a **Trust Centre** where enterprise customers can request their audit reports.
- **Microsoft 365 E5** is covered by a **plethora of certifications**. To list a few: ISO/IEC 27001, 27018 (cloud privacy), SOC 1/2/3, **FedRAMP High** (for M365 U.S. Government cloud and FedRAMP Moderate for commercial), **DoD IL4/5** (for defence use), **IRAP Protected** in Australia, **ENS High** in Spain, **CCSL** in NZ/Australia, and many more. Microsoft publishes audited compliance reports and has a continuous internal compliance program. In short, M365 compliance meets or exceeds requirements of most international standards – giving enterprises confidence that security controls are independently vetted. They also comply with **NZISM** guidelines for SaaS as evidenced by NZ government use (there's even a NZ Government Azure and 365 agreement). Microsoft's

own AI copilots (like Office 365 Copilot) inherit these controls. Azure OpenAI (the service underlying many copilot features) achieved FedRAMP **Moderate** authorisation in 2023 and is on track for higher certifications.

- **Google Workspace Enterprise** similarly holds **ISO 27001, ISO 27701 (privacy), SOC 2/3, FedRAMP Moderate** (Google Cloud FedRAMP authorisation includes many Workspace services), as well as **IRAP** assessed and various local certifications. Google's trust documentation states that independent auditors regularly verify their controls. Google Cloud is pursuing FedRAMP High for certain services as well. Google has also adopted **Binding Corporate Rules (BCRs)** for data protection as a data processor, which is an EU-approved mechanism indicating strong privacy safeguards. In terms of **NZ ISM**, Google Workspace isn't formally "certified" by NZ, but it's used by NZ govt agencies that have assessed it against NZISM. Google provides a **Compliance Resource Centre** mapping how its controls meet regulations including GDPR and NZ Privacy Act.

Network Security & Segmentation: OpenAI and Anthropic host their services on robust cloud infrastructure (OpenAI uses a mix of their own servers and Azure; Anthropic is partnered with Google Cloud and AWS). These environments feature multi-tenant security isolation. For higher security needs, **private network connectivity** is a consideration. Microsoft and Google allow enterprises to access services via private links or VPN: e.g. Microsoft's ExpressRoute can connect on-premises networks to M365/Azure directly, bypassing the public internet (with added encryption), and Google has Cloud Interconnect or identity-aware proxy setups. OpenAI's SaaS (chat.openai.com) is only via the public internet; there is no out-of-the-box private network peering for the ChatGPT service. However, if using OpenAI's API through Azure, one can deploy in a Virtual Network with private endpoints. Anthropic's Claude API on AWS could similarly be invoked from a VPC. So **indirectly, through cloud integrations, private network options exist** – e.g. Anthropic models are accessible in AWS Bedrock (which can be in a customer's VPC), and Google Cloud Vertex AI (Claude 2 is available there) can be used within a Google VPC. But using the providers' own managed apps (ChatGPT UI or Claude web interface) will involve standard TLS over the internet. All providers implement strict **firewalling, DDoS protection**, and routine **penetration testing**. Microsoft even stated that Office 365 undergoes red-team exercises and pen-tests as part of FedRAMP audits. OpenAI notes that external firms conduct regular penetration testing on their API and ChatGPT business services, catching issues proactively.

Summary of Security Posture: OpenAI and Anthropic have made impressive strides in a short time – achieving SOC 2, encryption, SSO, and other essentials – indicating **serious commitment to enterprise security**. Still, Microsoft and Google remain the gold standard with decades of security development, a huge array of controls, and compliance certifications covering virtually every requirement. One notable gap:

customer-side controls for encryption and data isolation are limited with OpenAI/Anthropic (you cannot yet host a private instance of ChatGPT or Claude on-premise or insist it run in a specific country's data centre – you rely on their cloud setup and general controls). In contrast, Microsoft and Google cloud environments are very configurable to meet specific security architectures. As we will discuss in section 4 (Risk Mitigation), enterprises that need to compensate for these gaps can implement supplementary controls (like API gateways, monitoring, etc.) when using OpenAI or Anthropic.

On the **physical and operational security** side, Microsoft and Google run their own data centres with state-of-the-art physical security and redundancy. OpenAI, by leveraging Azure, indirectly benefits from Microsoft's physical security. Anthropic on GCP/AWS does too. Each of these providers offers **99.9%+ uptime SLAs** for enterprise customers (OpenAI's Enterprise likely has an uptime SLA in the contract; M365/Workspace have published SLAs). All perform data backups and have disaster recovery plans – an enterprise should inquire and ensure the vendor's **business continuity** meets their needs (Microsoft/Google have multiple data centres in-region for failover; OpenAI's redundancy details are lesser-known publicly, but presumably Azure's geo-redundancy applies).

Certifications Snapshot: *OpenAI:* SOC 2 Type II; pursuing ISO 27001 (not yet claimed); no FedRAMP ATO on its own. *Anthropic:* SOC 2 II; ISO 27001; ISO 42001; HIPAA-ready. *Microsoft 365:* SOC 2; ISO 27001; ISO 27018; FedRAMP High (Gov); IRAP; GDPR/CCPA compliance; PCI (some components); etc.. *Google Workspace:* SOC 2/3; ISO 27001/27701; FedRAMP Mod; BCR; GDPR compliant; etc. Each vendor's trust centre provides full lists.

---

3. Legal and Compliance Posture

Ensuring compliance with privacy and data protection laws is a major part of "enterprise-readiness." Here we compare how OpenAI, Anthropic, Microsoft, and Google address key legal frameworks and contractual requirements:

New Zealand Privacy Act 2020: NZ's Privacy Act, with its 13 Information Privacy Principles (IPPs), governs personal information handling by agencies. A core concern is cross-border disclosure (IPP12) and overall **data sovereignty**. Neither OpenAI nor Anthropic has data centres in NZ as of 2025, so using their services entails transferring data to the United States (or wherever their processing occurs – likely US regions). NZ agencies must ensure the transfer is to a provider with **comparable safeguards to NZ law**. Both OpenAI and Anthropic, through their DPAs and adherence to GDPR, arguably provide comparable protection. However, the NZ Privacy Commissioner expects agencies to **conduct Privacy Impact Assessments**

(PIAs) and due diligence when using generative AI. Important considerations include: What personal data (if any) will be input? Is it necessary and proportional? Is the individual aware? The OPC's June 2023 guidance cautions agencies about potential risks like the AI provider retaining or disclosing personal information and using it for training. OpenAI Enterprise and Anthropic Enterprise directly mitigate that specific risk by contract (no training use), which aligns with OPC's expectation to **"caution against using sensitive data for training purposes"**. Agencies must also consider IPP5 (storage & security safeguards) – here the onus is to ensure the vendor has adequate security (SOC 2 reports, etc., as discussed) – and IPP11/12 (disclosure and international transfer). Using these AI services likely counts as a "disclosure" of personal info overseas, so either the individual consents or the agency ensures the vendor is subject to a law or contract that upholds similar safeguards (this is typically achieved via a DPA with standard contractual clauses, akin to GDPR's mechanism). Both OpenAI and Anthropic will sign DPAs that include **EU Standard Contractual Clauses (SCCs)** for data export, which would satisfy NZ's requirements for comparable safeguards. Microsoft and Google, being longstanding suppliers, already have agreements in place with NZ Government that cover Privacy Act obligations. In fact, the NZ government cloud framework and guidance (e.g. the **NZ Govt Chief Digital Officer's guidance**) currently lean towards caution with GenAI, requiring case-by-case approval. For private NZ businesses, the Privacy Act still applies if personal data is involved – so those businesses using OpenAI/Anthropic should similarly sign a DPA and possibly **notify individuals** if their personal info might be processed by an AI (transparency is key under IPP3). None of the AI vendors explicitly say "complies with NZ Privacy Act" but by complying with GDPR they meet most similar principles.

GDPR (EU General Data Protection Regulation): GDPR is stricter than NZ law in some aspects and includes hefty fines for non-compliance. OpenAI had a brush with GDPR when Italy's regulator temporarily banned ChatGPT in April 2023 over transparency and legal basis concerns. OpenAI responded by adding privacy disclosures and an ability for users to opt-out of data use – steps to align with GDPR. As of 2025, OpenAI's products can be used in the EU, but ongoing compliance is under watch (there have been complaints filed in some EU countries regarding lawful basis for processing personal data in training data). For enterprise usage, the key is that OpenAI offers a **Data Processing Addendum (DPA)** whereby OpenAI is the **processor** acting on the controller's instructions. This DPA includes SCCs for data transfer from the EU to US and commitments to assist with data subject rights requests, breach notifications, etc. Anthropic likewise provides a DPA (their privacy centre explicitly references a "Data Processing Addendum" for commercial customers). Under GDPR, things like **data minimisation** and **purpose limitation** are critical – enterprise should ensure they only send data to the AI that is necessary for the task (and ideally pseudonymise it). On the vendors' side, **purpose limitation** is respected by not using the data for anything beyond providing the service (no

secondary training use without consent). GDPR also requires robust security (which SOC 2 and ISO cover) and data subject rights. If an EU user wanted to delete their personal data from ChatGPT, OpenAI now has a process for users to delete accounts or specific conversations. Enterprise customers can also request deletion of data (and as noted, OpenAI deletes data within 30 days when you remove it). **Lawful basis:** In enterprise context, the lawful basis is usually "legitimate interests" or "performance of a contract" for processing employees' queries through the AI. The enterprise must inform employees (or customers) that their data may be processed by an AI service in the US. Microsoft and Google have very advanced GDPR compliance programs – including ability to choose EU-only processing (Google offers an EU data boundary for some services, Microsoft has EU Data Boundary commitments by 2023/2024 for Azure/M365). They have designated EU reps and so on. OpenAI and Anthropic, as newer entrants, likely rely on SCCs and don't yet offer EU-local processing or storage (though OpenAI has hinted at exploring data residency in future).

CCPA/CPRA (California) and similar laws: OpenAI's privacy policy states they do not "sell" personal information or use it for cross-context behavioural advertising, which is meant to address CCPA definitions. Enterprise use of these AI would likely fall under the B2B exemption or be covered by a service provider agreement. Both OpenAI and Anthropic would be considered "service providers" under CCPA when under a DPA, meaning they only use the data for the customer's purposes, not their own – thus exempt from "sale/share" classifications. Microsoft and Google already include CCPA addenda in their contracts, affirming no sale of customer data.

HIPAA (US Health Privacy): While not directly applicable to NZ, HIPAA compliance indicates an ability to handle health data securely. OpenAI has stated it can sign **Business Associate Agreements (BAAs)** for its API and enterprise services to support HIPAA compliance. That means OpenAI agrees to the specific data protection and breach reporting rules required for Protected Health Information (PHI). Anthropic, being "HIPAA configurable," likewise will sign a BAA for Claude for Work. That's important for any healthcare or insurance enterprises considering these tools – it shows the vendors are willing to take on legal liability for PHI security. Microsoft and Google have long offered BAAs for their services, and many hospitals use M365/G Workspace in HIPAA-compliant ways (with proper configurations).

PCI-DSS (Payment Card Industry Data Security Standard): This standard applies if credit card data is stored/processed. None of these AI services are designed to handle raw credit card numbers or payment processing. In fact, their usage policies likely forbid inputting sensitive financial identifiers. Microsoft 365 and Google Workspace have some aspects of PCI compliance (for instance, their cloud infrastructure is PCI-certified for the parts of services like if you store a credit card in a spreadsheet, the environment itself is secure, but the **organisation using it** is responsible for compliance). OpenAI and Anthropic are not known to be PCI certified.

If an enterprise wanted to use an LLM with cardholder data, it would need a very controlled scenario – likely not advisable with the current SaaS. It's best to keep payment data out of prompts entirely. In practice, PCI compliance is not a relevant use case for ChatGPT/Claude (and indeed the Digital Marketplace listing for M365 shows "PCI certification: No" for Office 365). This is one area where Microsoft/Google don't fully cover either, as Office docs or emails containing credit cards are generally not recommended. Instead, specialised payment systems are used.

Data Processing Agreements & Standard Contractual Clauses: Both OpenAI and Anthropic make DPAs available. Anthropic's privacy centre points to a Data Processing Addendum and clarifies their role as processor vs controller. OpenAI's Enterprise terms incorporate a DPA (often these are accessible via their trust portals or by request). It's crucial that enterprises executing contracts with OpenAI/Anthropic include those DPAs, which will contain SCCs for international transfer (to satisfy GDPR and NZ Principle 12). Microsoft and Google automatically include DPAs in their enterprise agreements (Google's is built into their Online Acceptable Use Policy/Terms, Microsoft's is in the Online Services Data Protection Addendum). These documents ensure that **customers have enforceable contractual rights** regarding their data. For example, all four providers commit to cooperate with audits or questionnaires to verify compliance (to a reasonable extent) and to flow these requirements down to any sub-processors.

Local Data Residency and Sovereignty: As touched on, Microsoft and Google can meet strict data residency demands by using local datacentres or dedicated sovereign cloud instances (e.g. Microsoft 365 has a Germany sovereign instance, a China instance operated by 21Vianet, etc., and Azure has a public sector NZ region for government). Google is opening a New Zealand cloud region in 2024, which will further help NZ customers keep data local for Google Cloud services (though it's unclear if Workspace will offer NZ-specific data location, it likely will route NZ org data to Australia or the nearest at least). **OpenAI and Anthropic currently do not offer customer-selectable data regions** – data is processed in the US and perhaps backup in other jurisdictions (OpenAI might use some Azure EU resources for certain customers if negotiated, but that's speculative). This is potentially problematic for sectors that require data to remain in-country (some government or finance data classifications). However, some regulatory flexibility can be obtained via contract and encryption – e.g. if no plaintext personal data is stored and all data is encrypted strongly, an overseas processing might be acceptable. Enterprises should closely watch if OpenAI sets up EU or other regional hosting in the future. (Notably, Microsoft Azure OpenAI allows you to choose an Azure region like West Europe or East US – that is one way to get regional control but it's a different product packaging.)

Compliance with Sectoral Regulations: Aside from privacy, consider industry-specific rules. For example, **financial services** in many countries have outsourcing guidelines (like APRA in Australia or RBI in India, or FMA guidelines in NZ) that require risk assessments and often that customer data remains accessible for audit and perhaps within certain jurisdictions. Microsoft and Google have years of experience aligning to finance and government standards (including providing audit artifacts, localised services, etc.). OpenAI/Anthropic are in early stages of engaging regulators. To illustrate, **France's banking regulator** reportedly asked banks to ensure any use of ChatGPT doesn't violate bank secrecy laws. A major mitigation here is that ChatGPT Enterprise provides the contractual confidentiality needed. Another area: **EU AI Act** is upcoming – it will impose transparency and risk management obligations on providers of AI. OpenAI and Anthropic will need to comply once that is in effect (likely 2025/26), possibly requiring things like detailed documentation of training data, etc. Enterprise customers should keep an eye on how the evolving AI regulations (EU AI Act, US AI executive order, etc.) might affect vendor offerings and their own use.

Liability and Indemnity: A legal consideration: do these providers indemnify customers for breaches or AI outputs? Microsoft and Google enterprise contracts typically have robust indemnities for IP infringement, data breach liability (often capped), and compliance failures. OpenAI's standard terms for API disclaim much of liability for content generated (since the user ultimately uses the output), and likely their enterprise agreement limits liability significantly. Enterprises might negotiate liability for data breaches – e.g. if OpenAI's negligence leads to a breach of customer data, the customer wants recourse. A DPA usually includes that the processor is liable for GDPR fines that are due to the processor's actions. It's worth noting that with new AI, many enterprise customers are negotiating custom contract terms (for instance, ensuring **OpenAI will bear responsibility if it were to train on their data contrary to contract** or if a data leak occurs).

Intellectual Property and Output Use: Legally, who owns the AI-generated output? OpenAI's terms for ChatGPT Enterprise clarify that **the customer owns the outputs** it receives, to the extent allowed by law. This means if your employee uses ChatGPT Enterprise to generate code or text, your company can use that output freely – OpenAI won't claim copyright. Anthropic's terms similarly don't claim ownership of output; the user is free to use what Claude produces. Microsoft and Google explicitly state that customer data and outputs remain the customer's. However, one should be aware of IP risk: generative models can accidentally produce someone else's copyrighted text (rare but possible). Microsoft addresses this by offering an **indemnity for Copilot outputs** – if Copilot produces, say, code that infringes copyright, Microsoft will defend the customer (as announced for GitHub Copilot and likely extended to M365 Copilot). OpenAI/Anthropic haven't publicly offered such indemnities, so enterprises might use at own risk or rely on fair use. This isn't a

compliance issue per se, but an emerging legal area. NZ copyright law doesn't yet account fully for AI outputs, but if the output included personal data or sensitive info, privacy and confidentiality laws would kick in – again highlighting that internal review of outputs is important.

Real-world compliance incidents: To date, neither OpenAI nor Anthropic have had known regulatory sanctions beyond the early privacy questions. Microsoft and Google have had many audits (and occasionally fines, e.g. Google got GDPR fines for other services unrelated to Workspace). NZ's Privacy Commissioner hasn't taken action against any agency for using ChatGPT yet, but did set expectations (as cited above). In Australia, a recent development saw the **Victorian Privacy Commissioner ban ChatGPT use in a government department** for failing a privacy impact assessment – illustrating that if a risk assessment finds non-compliance (in that case likely because OpenAI was a controller of data with uncertain deletion), the conservative approach was to ban it. With enterprise versions, these concerns are addressed, so regulators may be more amenable. The **bottom line** is enterprises must ensure **legal paperwork is in place (DPA, SCCs)** and that they configure and use these services in a manner that complies with applicable laws (e.g. do not input sensitive personal info without proper basis/consent, provide transparency to users, and protect outputs that contain personal data). Microsoft and Google's environments make it easier to comply due to features like content classification, retention policies, etc., whereas with OpenAI/Anthropic, more reliance is on policy and training of users.

---

4. Enterprise Risk Mitigation Features

Using generative AI in an enterprise setting introduces new risks (data leaks, misuse, inaccurate output). Beyond trust in vendor commitments, enterprises look for **technical and contractual controls** to mitigate these risks. We compare features provided by the platforms and additional controls enterprises should implement:

Built-in Enterprise Controls (OpenAI & Anthropic): Both ChatGPT Enterprise and Claude Enterprise have introduced controls for the admin to manage usage:

- **Access Management:** As mentioned, SSO integration is supported (Azure AD, Okta etc.), so you can centrally provision or deprovision users. Anthropic Enterprise has **SCIM support** for automated user provisioning, which is helpful for large organisations to manage accounts at scale. Domain allow-listing ensures only users with company email can join the workspace.
- **User Roles and Permissions:** Anthropic allows setting a **Primary Owner, Owners, and Members** in a Claude workspace. This means some users can have elevated rights to manage settings while regular users just chat. OpenAI Team/Enterprise likely has a similar concept (an admin who can invite users,

set policies). These permissions help with governance – e.g. restricting who can enable certain features.

- **Data Controls:** OpenAI Enterprise gives the admin control over **data retention duration**. An admin could choose, for example, that chat history is only retained for 7 days or 30 days, balancing utility vs. risk. Shorter retention means if a credential or personal data is accidentally typed, it won't live on their servers for long. Anthropic Enterprise goes further by letting you choose retention period (min 30 days) or even zero-retention by special agreement. Also, both allow disabling of data sharing mechanisms: e.g. Anthropic Enterprise admin can **disable the thumbs-up/down feedback** for all users (to prevent anyone from accidentally sending conversation data as "feedback" that would be stored longer for model training purposes). OpenAI's business plans by default do not share data, and ChatGPT Enterprise doesn't even have the "opt-out" toggle visible (since it's always opted-out), but they might allow disabling of certain logging.
- **Content Controls:** Neither OpenAI nor Anthropic allow custom content filtering rules by the customer yet, but they both have built-in **moderation filters**. For instance, OpenAI employs an automated content moderation system to block disallowed content (hate, violence, sexual abuse, etc.) and will refuse or flag those requests. Anthropic's Claude has its **Constitutional AI** approach that makes it resistant to producing toxic or sensitive outputs. For enterprise, reducing problematic outputs mitigates risk of harassment, discrimination, or data leakage via the model. However, these filters are vendor-defined. Microsoft and Google, by comparison, integrate their AI outputs with existing **DLP (Data Loss Prevention)** systems – e.g. Google Duet will abide by any DLP policies you set (if a user tries to use Duet to write an email containing sensitive info that violates a rule, it can get blocked). Microsoft's Copilot likewise respects M365 Compliance settings: it won't surface data the user isn't allowed to access and will presumably log its activities for compliance. In the OpenAI/Anthropic scenario, **enterprise must rely on policy and training** since you cannot yet enforce, say, "don't allow users to paste 16-digit numbers" on their platform.
- **Monitoring and API Controls:** OpenAI's enterprise API usage can be monitored through their dashboard or via API keys. Anthropic's Claude API similarly provides an API key per org. An enterprise can rotate keys, set quotas, or use a proxy to monitor prompts. OpenAI recently introduced **organisation-level API keys and usage insights**, which helps large teams see how the API is being used. These are not as full-featured as Azure's, but it's improving. There is also talk of **audit APIs** from these vendors – i.e. an API that lets a compliance officer fetch conversation records for audit. If not available yet, it's on the roadmap.
- **Integration with enterprise systems:** Microsoft and Google obviously integrate with their own suite (Teams, Gmail, etc.). Anthropic is adding

integrations like **GitHub sync** for Claude (to bring in code context) and likely will add other connectors. OpenAI doesn't yet have native integrations but can be manually integrated via API. While these features (like connecting internal knowledge bases) improve utility, they also raise risk – e.g. connecting Claude to your Confluence or SharePoint means the AI has access to lots of internal data. Thus, the security of that connector is crucial (Anthropic says the data you sync is used only as context and not stored beyond your org – it stays in your "Project" and is not used to train Claude globally).

Enterprise Contractual/SLA Features:

- **Service Level Agreements (SLA):** ChatGPT Enterprise likely offers an uptime SLA (perhaps 99.9%) and priority support. Anthropic Claude Enterprise presumably similarly offers support response time guarantees. Microsoft and Google have published SLAs (generally 99.9% uptime for core services, with credits if violated).
- **Customisation of Terms:** Large enterprises or government clients often negotiate custom contract addenda. For example, an NZ government department might require the contract to explicitly state compliance with NZ Privacy Act and that data will be stored in certain jurisdictions or that the provider will cooperate with Official Information Act requests. Microsoft and Google have standard government contract riders for that; OpenAI/Anthropic being newer may negotiate case by case. We expect that for sufficiently large deals (e.g. a Fortune 500 company or a government signing up), both OpenAI and Anthropic would accommodate **additional privacy, security, or liability clauses**. Already, Anthropic offers **a separate set of terms for commercial vs consumer** (they have "Terms of Service – Commercial" on their site), which likely include more robust indemnities and commitments than the consumer terms. OpenAI's Enterprise terms are not public, but likely similar (the consumer terms are quite limited, but the enterprise will have a tailored contract).
- **Indemnification and warranties:** While not publicly detailed, enterprise contracts for these AI likely contain warranties about data handling (e.g. warranting that "we will not use your data to train AI" as a binding clause, not just a marketing promise). If they breach that, you have contractual remedies. Indemnities may cover intellectual property of outputs or confidentiality breaches. Again, Microsoft leads here by offering an IP indemnity for Copilot output (a strong promise unique in the industry so far). Enterprises should ask for indemnification if AI output accidentally leaks third-party confidential data (though how that would happen is complex – presumably only if the model somehow regurgitated training data that was confidential, which OpenAI says their model should not do with private data).
- **Customer Success and Policy Guidance:** OpenAI and Anthropic now have teams to help enterprise deployments. They often assist with recommended

prompts, integrating the AI into workflows, etc. They may also advise on safety best practices. This soft feature is important – it means as an enterprise you're not just buying an API key, you get a partnership (which is something Microsoft and Google have offered for years via account teams).

Recommended Internal Controls (for any platform): Regardless of vendor features, enterprises should implement their own controls to mitigate risks:

- **Acceptable Use Policies & Training:** Clearly define what data can or cannot be input into the AI. For example, a bank might prohibit paste of full client personally identifying information (PII) or any account numbers into ChatGPT. Employees should be trained that while enterprise ChatGPT/Claude won't leak data to the public, it's still not a memory hole – treat it as you would any outsourced service. Also, train users to **verify AI outputs** before using them, to mitigate misinformation risk.
- **Data Classification and Redaction:** Before sending data to an AI, consider automating a redaction step. Tools can mask names or ID numbers with placeholders. This way you get the AI analysis on data without exposing real identities. If using OpenAI/Anthropic via API, one could build a middleware that strips out sensitive fields (e.g. replace "John Doe" with "[Person1]" in prompts). For Microsoft/Google's integrated AI, they leverage the fact that it runs on your tenant's data without exposing it externally – but still, if one were to prompt Copilot with extremely sensitive info, one should ensure that data was supposed to be in M365 to begin with (if yes, it's already protected by Microsoft's compliance boundary).
- **Output Review and Filters:** Treat AI outputs as draft. Particularly if the output will be communicated externally or used for decision-making, have a human in the loop to review accuracy and appropriateness. Internally, one could use a secondary AI or rules to scan outputs for any sensitive data before they are saved or shared. For example, if an employee asks ChatGPT Enterprise to write a summary of a client email, the output might inadvertently contain personal data from the prompt – the employee should ensure that's handled per privacy rules before forwarding it.
- **Logging and Monitoring:** Use available logs (from the vendor or via network logs) to monitor usage patterns. Unusual volume of data being sent or prompts containing certain keywords might indicate misuse. For instance, an enterprise proxy could flag if someone tries to paste "BEGIN PGP PRIVATE KEY" (an obvious sign of attempting to feed confidential material). Some companies have even blocked ChatGPT access at the firewall except for approved accounts, to ensure only the enterprise version is used (preventing users from going to the public site).
- **Key Management:** For API usage, manage the API keys securely. Don't hard-code them in public repos, rotate them periodically, and assign separate keys for different teams to isolate risk. Microsoft's Cloud App Security (Defender for

Cloud Apps) and Google's CASB can detect use of known SaaS – they can be configured to **detect and block uploads of sensitive data to unapproved apps**. An organisation could leverage those to ensure employees aren't secretly using personal ChatGPT with work data.

- **Network Isolation (advanced):** If extremely sensitive, run the AI in a contained environment. E.g., call the AI API only from a secure network segment, not from open internet. Use VPN or private links where possible (Azure OpenAI in a VNet, etc.). This reduces exposure if someone tries man-in-the-middle (though TLS already protects traffic). Also, ensure endpoint security on devices to prevent malware from exfiltrating data to an AI API.
- **VPC Peering / On-premise alternatives:** If absolute control is needed, some organisations explore hosting open-source LLMs on-premise (like Llama 2) to avoid external data transfer entirely. However, those models may not match GPT-4/Claude in capability and come with their own security and cost issues. Still, for the highest risk data, this is a viable internal control: use closed networks or self-hosted AI for that subset of tasks. For more moderate data, use the best-in-class external AI with the above controls.
- **Audit and Compliance Checks:** Periodically audit how the AI is used. Ensure it aligns with stated policy (for example, review a random sample of prompts from logs to confirm no one is pasting customer SIN numbers). Ensure the DPA is up to date and that any new features (plugins, etc.) are reviewed for compliance before enabling. Also, maintain an inventory of which departments are using AI and for what purposes – helpful for DPIAs and demonstrating compliance.
- **Incident Response Plan:** Extend your incident response to include generative AI. For example, if an employee reports "I think I pasted a client's address into ChatGPT by mistake," have a procedure: you might contact OpenAI (enterprise support) to ensure the conversation is deleted immediately rather than waiting 30 days, and analyse if that data could have been output to any other user (likely not, but due diligence). Also, if the AI produces something inappropriate (e.g. biased output that, if used, could cause a legal issue), treat it as an incident to learn from – maybe adjust the prompt patterns or provide feedback to the vendor.

Enterprise Risk Feature Comparison to M365/Google: Microsoft 365 and Google Workspace provide many built-in compliance features that help mitigate AI risks by default. For example, **Data Loss Prevention (DLP)** rules you set up (like "flag if an email contains a tax file number") will also apply if a user tries to have Copilot email that info – Copilot's output will be scanned and blocked if violating policy. Also, **Customer Lockbox** (in M365) ensures support engineers cannot access your data without explicit approval – a level of control you won't get from OpenAI at this time. Microsoft Purview and Google Vault allow eDiscovery on all content, including (eventually) AI-produced content. In contrast, with OpenAI/Anthropic, if you need to

do eDiscovery on AI conversations, you'd have to rely on stored chat histories and export them. This is manageable if retention is turned on; you could ask OpenAI to export all conversations in a date range. It's just not as turnkey as with Microsoft/Google where it's one admin console search.

Emerging Features: We anticipate both OpenAI and Anthropic will introduce more enterprise controls: e.g. **Bring Your Own Key** encryption, on-premise deployment options via cloud marketplaces, more admin analytics, and fine-grained policy settings (maybe "disable code generation" for certain groups, etc.). Microsoft and Google will keep enhancing their AI integration with compliance (for example, Google's **Content Safety API** might be used to filter Duet AI outputs for toxicity). For now, enterprises adopting OpenAI or Anthropic should **compensate with strong internal governance** as described. It's telling that Morgan Stanley built an entire "evaluation framework" and retrieval limits for their GPT-4 usage – they didn't just plug it in blindly; they put guardrails and constant monitoring of output quality and compliance, which is a model approach to risk mitigation.

---

5. Perception vs. Reality: Enterprise Hesitation and Case Studies

Generative AI's meteoric rise has brought along a mix of excitement and fear in enterprises. It's important to separate **perceptions (myths or initial impressions)** from the current **reality** now that enterprise-grade options exist.
Perception: *"If employees use ChatGPT, our data will be sucked into a public dataset and could leak to others."*
Reality: This was a valid concern for the **free** version of ChatGPT – user inputs were indeed used train the model (until users could opt-out) – and there was at least one incident where a ChatGPT bug exposed snippets of other users' chat history. Those events in early 2023 spooked many companies into outright bans. For example, major firms like Samsung, JPMorgan, Apple, Amazon, and Deutsche Bank banned employees from using ChatGPT in 2023. In New Zealand, the Ministry of Business, Innovation and Employment (MBIE) took a cautious approach by **banning staff from using ChatGPT or other AI** in June 2023 due to privacy and security concerns. These bans were often temporary measures until governance could catch up. **Today**, with ChatGPT Enterprise and Claude Enterprise, the data is not used to train the AI and is kept confidential. The **enterprise versions effectively address the data leak concern** by providing isolation. If an employee uses ChatGPT Enterprise, another company's model will not suddenly know about it or produce it. The **Morgan Stanley case** is illustrative: Initially, financial firms were very wary, but Morgan Stanley partnered with OpenAI under strict conditions, ensuring *"OpenAI doesn't learn anything about our private information… they're not training their models on our*

data", as one executive noted. Morgan Stanley was able to deploy GPT-4 to 20,000 staff by curating the use (only internal data, and OpenAI's model hosted in a segregated manner) and it has been successful without breaches. This shows that once the *reality* of data control was established (via contract and tech safeguards), adoption became possible.

Perception: *"These AI tools aren't secure or compliant enough for regulated industries."*

Reality: Initially, regulators did raise red flags. Italy's data authority said ChatGPT wasn't transparent and possibly lacked legal basis under GDPR – leading to a temporary ban in March–April 2023. OpenAI quickly implemented age checks, privacy disclosures, and an opt-out feature to comply. The Italian ban was lifted, and no other EU country banned it after those changes. In Australia, as mentioned, the Office of the Victorian Information Commissioner investigation led to a state dept ban, citing failure to assess privacy properly. However, regulators are not universally against it – many are actively exploring guidelines. The NZ Privacy Commissioner's guidance doesn't prohibit GenAI; it outlines expectations like doing a PIA and ensuring necessary/proportionate use. The *reality* is that compliance can be achieved. For example, **Germany's Deutsche Bank** reportedly allowed controlled use of an LLM after careful vetting (they had banned it initially). **Deloitte**, a Big Four firm, is highlighted as an Anthropic Claude Enterprise customer, indicating even firms handling sensitive client data are embracing these tools with the right agreements. **Government adoption** is slower but happening – the Singapore government, for instance, launched a GOV AI platform using OpenAI tech internally with added safeguards. In NZ, some agencies are running limited trials under the oversight of the Government Chief Digital Officer.

Perception: *"The AI might say something that gets us sued (e.g. libelous or biased content)."*

Reality: This is a risk – generative AI can produce incorrect or even problematic outputs. Enterprises worry about "AI hallucinations" giving wrong answers to customers or outputs that reflect biases. To mitigate this, enterprise AI usage is mostly internal or human-reviewed for now. Microsoft and Google emphasise a human-in-the-loop approach for their copilots. OpenAI and Anthropic have improved their models' reliability (Claude 2 and GPT-4 are much more factual than earlier models, but not infallible). They also allow some customisation: OpenAI's system message or Anthropic's "constitution" can be tailored to reinforce company policies (to a degree). Real-world example: **KPMG** is using Azure OpenAI to assist auditors; they manage risk by limiting the AI to certain tasks and verifying outputs. The NZ Privacy Commissioner actually warns about accuracy – agencies must ensure outputs are verified for accuracy and bias before use. The reality is, with proper use

(not relying on AI for final decisions without check), this risk can be managed. Many companies treat AI suggestions like a junior staffer's work – helpful but requiring review.

Perception: *"If we put data in the cloud (OpenAI/Anthropic), it's automatically a breach of confidentiality or privacy."*

Reality: Cloud services can be used without breaching confidentiality as long as contractual and technical measures are in place. For instance, attorney-client privilege can be maintained using these tools if done carefully – some law firms are experimenting with AI for legal research but ensuring no client identifiers are input. Banks worried about customer confidentiality have seen cases like **Samsung** where an employee reportedly pasted source code into ChatGPT (free) and that data might have become part of OpenAI's training set, theoretically risking exposure. This led to Samsung banning it. The lesson learned pushed OpenAI to create enterprise solutions to avoid that scenario altogether. Now, using ChatGPT Enterprise, that Samsung engineer's paste would *not* leave Samsung's control in terms of training, and could be deleted. So the **reality** has evolved such that confidentiality can be preserved. Still, caution is needed: if an organisation has rules like "no client data on any third-party system without clearance," those rules apply here too. Generative AI should be evaluated like any vendor handling sensitive data. With NDAs, DPAs, and the fact that OpenAI/Anthropic promise not to look at the data except for abuse, one can argue it is not fundamentally different from other cloud outsourcing.

Perception: *"Our employees will misuse the AI or feed it garbage in, garbage out."*

Reality: This is more of a change management issue than a tech issue. Early on, some employees tried using ChatGPT for tasks it wasn't suited for, or even asked it to do things against policy. But with guidelines, employees can become adept at using AI productively and safely. In practice, many enterprises that banned ChatGPT informally found employees using it on the side via personal accounts – a shadow IT issue. Offering an approved, governed AI tool can actually reduce that and let employees innovate in a safer sandbox. For instance, **PwC** announced a deal to provide ChatGPT (via Azure) to tens of thousands of employees with monitoring in place, precisely to channel use responsibly rather than have untracked use. The NZ government's tactical guidance encourages "safe experimentation" with generative AI rather than prohibition – implying that with the right guardrails, benefits can be realised. The reality is employees will use these tools (because they enhance productivity dramatically), so enterprises are shifting from blanket bans to **controlled enablement** on enterprise-ready platforms.

Perception: *"OpenAI and Anthropic are startups – can we trust them like we trust Microsoft or Google with our data?"*

Reality: OpenAI and Anthropic, while younger companies, have taken significant steps to build trust – such as hiring security teams, obtaining certifications, and being transparent about policies. They both have substantial funding (OpenAI backed by Microsoft, Anthropic by Google and others) and are unlikely to vanish overnight. That said, they do not have the decades-long enterprise track record. Some risk-averse organisations may still prefer to access OpenAI's models *through* Microsoft (Azure OpenAI) because then Microsoft is the contracted entity with all its reliability and compliance. This is a valid strategy – it effectively puts a Microsoft wrapper around OpenAI tech. Anthropic's partnership with Google (via the Vertex AI platform) plays a similar role. In NZ, using Azure OpenAI via the local Microsoft datacentre could address sovereignty and support concerns (Microsoft would handle support and the model runs in Azure). So while direct trust in the startups is growing (with SOC 2 reports to back it up), there are options to leverage them under an umbrella of a trusted cloud provider. The reality is that regulators and large customers *are* engaging directly with OpenAI/Anthropic now – for example, OpenAI reportedly is working on a UK government pilot with an NHS agency. Over time, their trust level is approaching that of established vendors, especially as they continue to shore up compliance (Anthropic getting ISO 27001, etc., is a good sign).

Case Studies (NZ or similar)

- *New Zealand context:* As noted, MBIE banned GenAI for staff in 2023, but this is likely a temporary stance. Other departments may be experimenting in sandboxes. NZ businesses vary – a Datacom survey found about 50% of NZ companies had started adopting AI, but many lacked formal policies. This suggests a gap between use and governance that needs closing. One NZ law firm (Buddle Findlay) commented that the Privacy Commissioner's guidance means businesses must be very cautious, but not necessarily avoid AI – rather get explicit about purposes and safeguards.
- *Adoption:* **Morgan Stanley** (US but globally relevant) – as described, deployed "AskResearch" GPT-4 assistant with a custom knowledge base, after rigorous evaluation. This indicates a high degree of trust achieved when using GPT-4 in a controlled, fine-tuned manner.
- **Italy's Government** – after initial ban, now working with OpenAI on a task force to ensure compliance, showing that even regulators are coming around once their concerns are addressed with facts and changes.
- **Deloitte, AWS, Google** – Deloitte is using Claude (Anthropic) internally; AWS is offering Anthropic models to enterprise customers via Bedrock; Google Cloud signed an agreement with NZ's DIA (Dept of Internal Affairs) to make AI

available under some conditions – so the ecosystem is moving to enable safe use.

- **Rejection:** Some organisations remain sceptical. For example, some EU banks are still keeping a ban until clearer guidance from data protection authorities emerges. These cases often cite lack of clarity on data handling. But as OpenAI and Anthropic publish more about their enterprise controls, the factual basis for extreme caution diminishes.

Misconceptions: There is a lingering misconception that "ChatGPT is trained on everything you type." As we've detailed, this is false for business accounts. Another misconception is that using these AI might violate laws like GDPR outright. Not if done with a proper DPA and if you're not processing sensitive personal data without a lawful basis. It's similar to using any cloud SaaS – you must check the compliance boxes. The **Office of the Privacy Commissioner (NZ)** expects that if personal information is involved, the agency will ensure rights like access and correction can be met. Generative AI may not easily allow an individual to retrieve or delete "their data" from the AI's brain (if it was in training data). That remains a tricky area – but note that enterprise use typically doesn't involve feeding new personal data to train the model, so it's more about the prompts/outputs, which can be deleted.

Overall: Enterprises were right to be cautious early on; some high-profile mishaps underscored that (e.g. A Samsung employee pasting confidential code into a free AI, or someone using an LLM to draft a public document and it inserted a fictional citation, embarrassing the firm). But the landscape in late 2024 and 2025 is one of rapid **professionalisation of AI offerings**. With **ChatGPT Enterprise and Claude Enterprise, many initial fears are directly addressed**. What remains is for enterprises to adapt their risk management – using the available controls and demanding any additional ones needed. Adoption is picking up as a result: a survey might show the majority of big companies are now at least piloting these tools. In New Zealand, we will likely see early adopters in sectors like tech, education, and finance who craft careful policies to satisfy the Privacy Act and industry rules, rather than outright bans.

The **perception that generative AI is a "wild west" is becoming outdated**. Reality is, it's becoming a governed service much like any other cloud service, though with some unique challenges (like unpredictable outputs). The key to bridging hesitation is education and transparency – knowing exactly what the commitments (no training, deletion, security) mean and having evidence (certifications, case studies) that they are upheld. Confidence grows as success stories emerge (e.g. "Company X deployed GPT-4 and saw a productivity gain with no security incidents in 6 months"). Each of those reduces the fear of the unknown.

As one more data point: **93% of companies in an Australian survey said they planned to ban ChatGPT** in early 2023, yet by late 2024 many of those are instead seeking "secure and private" AI alternatives – which often means enterprise versions of OpenAI/Anthropic or similar. This swing shows that once proper enterprise options exist, the calculus changes from fear to cautious optimism. The OPC's guidance basically says: do your homework, but we *encourage innovation* (they list benefits like efficiency and better service design). So the enterprise trend now is moving from "block AI" to "embrace AI responsibly."

Visual Matrix Scorecard: AI Platforms vs Enterprise Criteria

Below is a comparative **scorecard** summarising how OpenAI (ChatGPT) and Anthropic (Claude) stack up against Microsoft 365 and Google Workspace on key enterprise-readiness dimensions. ✓ indicates full support/compliance, 🔶 indicates partial or emerging support, and ❌ indicates not available or a gap:

Criterion	OpenAI ChatGPT (Free/Plus)	OpenAI ChatGPT (Team/Enterprise)	Anthropic Claude (Free/Pro)	Anthropic Claude (Work/Enterprise)	Microsoft 365 E5 (Office 365)	Google Workspace (Enterprise)
Data used for model training	Yes (by default, to improve models)*Can opt-out?* 🔶 (history off stops training use)	**No** (no training on customer data by default) ✓	No (default is no training on chats) ✓	**No** (no training on customer data) ✓	**No** (customer data not used to train AI) ✓	**No** (customer data not used to train AI) ✓
Data retention (default)	Indefinite (chats saved until user deletes) ❌ *Deleted chats* – 30 days retention	Admin-configurable (e.g. 30 days or less) ✓ *Deleted chats* – purged in 30 days	~90 days (prompts auto-deleted on backend) 🔶 *Flagged* – up to 2 yrs (for abuse)	Configurable (30 days minimum, can agree shorter) ✓ *Flagged* – up to 2 yrs (safety)	Customer-controlled (via retention policies) ✓ *Deleted* – purged per policy (e.g. 30 days default)	Customer-controlled (via Vault retention rules) ✓ *Deleted* – purged per admin settings
Data ownership	User owns inputs & outputs (per policy) ✓, but OpenAI has broad rights on content (free user license)	Customer retains ownership; OpenAI only has rights to operate service ✓	User owns content; Anthropic has rights mainly to provide service ✓	Customer retains ownership; Anthropic acts as data processor ✓	Customer owns data; Microsoft has no rights except to run service ✓	Customer owns data; Google only processes per customer instructions ✓
Data segregation	Shared environment (consumer service); no customer-specific isolation ❌	Dedicated enterprise workspace; logical tenant isolation ✓	Shared consumer service; no tenant concept ❌	Dedicated org workspace; strong logical isolation ✓	Dedicated tenant per org; strong isolation ✓	Dedicated tenant per org; strong isolation ✓

Criterion	OpenAI ChatGPT (Free/Plus)	OpenAI ChatGPT (Team/Enterprise)	Anthropic Claude (Free/Pro)	Anthropic Claude (Work/Enterprise)	Microsoft 365 E5 (Office 365)	Google Workspace (Enterprise)
Encryption in transit & rest	Yes (HTTPS TLS 1.2+; encrypted storage) ✓	Yes (TLS 1.2+; AES-256 at rest) ✓	Yes (standard cloud encryption) ✓	Yes (TLS; AES-256 or equivalent at rest) ✓	Yes (TLS 1.2/1.3; AES-256 at rest) ✓	Yes (TLS 1.3; AES-256 at rest) ✓
Single Sign-On (SSO)	No (login via OpenAI account only) ❌	Yes (SAML SSO, Azure AD/Okta integration) ✓	No (Claude AI uses its own login or Google auth, but not enterprise SSO) ❌	Yes (SAML/OIDC SSO and domain-based access) ✓	Yes (Azure AD/Entra ID for M365) ✓	Yes (Google Workspace Identity / SAML) ✓
Multi-factor Authentication	Yes (via OpenAI login – supports 2FA app) ✓	Yes (through SSO provider controls) ✓	Yes (via standard login with Google account 2FA) ✓	Yes (through SSO enforcement) ✓	Yes (Azure AD MFA, Conditional Access) ✓	Yes (Google 2-Step, or SSO) ✓
Role-Based Access Control	N/A (single user only) ❌	Yes (Admin vs Member roles; fine-grained feature access control) 🔶 (basic roles exist)	N/A (consumer user only) ❌	Yes (Primary Owner, Owners, Members with different privileges) ✓	Yes (Admin roles, e.g. global admin, compliance admin, etc.) ✓	Yes (Admin roles, delegated roles for various services) ✓
Audit logging (user activity)	Minimal (perhaps IP logging, no user-accessible audit) ❌	Planned/Basic (enterprise usage stats available; full audit log features evolving) 🔶	Minimal (no customer audit logs) ❌	Yes (Audit logs of Claude Enterprise usage, available to admin – in development) 🔶	Extensive (Unified Audit Log, real-time user activity logs; admin export) ✓	Extensive (Admin logs, Access Transparency logs for Google ops) ✓
Incident response / support	Community support, no SLA ❌	Priority support, 24/7 critical incident response ✓ (SOC 2 in place)	Email support for Claude.ai, no enterprise SLA ❌	Dedicated support with SLAs (Anthropic support team, likely 24/7 for critical) ✓	24/7 Premier Support available; formal incident notification processes ✓	24/7 Enterprise support; formal incident management (per DPA) ✓

Criterion	OpenAI ChatGPT (Free/Plus)	OpenAI ChatGPT (Team/Enterprise)	Anthropic Claude (Free/Pro)	Anthropic Claude (Work/Enterprise)	Microsoft 365 E5 (Office 365)	Google Workspace (Enterprise)
SOC 2 Certification	No (not for free consumer service) ❌	**Yes** – SOC 2 Type II audited ✓	No ❌	**Yes** – SOC 2 Type II audited ✓	Yes – SOC 2 Type II (for cloud services) ✓	Yes – SOC 2/3 audited ✓
ISO 27001 Certification	No ❌	No (in progress/ not announced) 🔶	No ❌	**Yes** – ISO/IEC 27001:2022 ✓	Yes (Office 365 is ISO 27001 certified) ✓	Yes (Google has ISO 27001 for Workspace) ✓
Other certifications	N/A (none for free tier) ❌	SOC 2, CSA STAR Level 1; GDPR/CCPA alignment; can sign BAA (HIPAA) 🔶	N/A ❌	ISO 42001 (AI MSM) ✓; HIPAA-ready (BAA available) ✓; CSA STAR Attestation; pursuing more	FedRAMP, IRAP, CJIS, HIPAA, GDPR, etc. ✓ (broad compliance portfolio)	FedRAMP Moderate, BCR, HIPAA, GDPR, etc. ✓ (broad compliance)
Data residency options	No (data processed in US/global) ❌	Not yet (no customer-controlled region; primarily US) 🔶	No ❌	Not yet (hosted in US by Anthropic, though available via EU clouds indirectly) 🔶	Yes (choose tenant region; NZ/AU/EU/US available) ✓	Yes (data region policies – EU/US; upcoming APAC region) ✓
DPA & SCCs (legal)	No DPA for free users; Std Terms only ❌	Yes – DPA available with SCCs; OpenAI as Processor ✓	No (consumer terms only) ❌	Yes – DPA available; Anthropic as Processor ✓	Yes – comprehensive DPA & SCC part of contract ✓	Yes – comprehensive DPA & SCC incorporated ✓
HIPAA BAA	No (not for consumer use) ❌	Yes (will sign BAA for Enterprise/API) 🔶	No ❌	Yes (will sign BAA for Claude commercial) 🔶	Yes (Microsoft will sign BAA; O365 is HIPAA eligible) ✓	Yes (Google will sign BAA; Workspace can be HIPAA compliant) ✓
Customer-managed encryption keys (CMEK)	No ❌	No (encryption is managed by OpenAI) ❌	No ❌	No (encryption managed by Anthropic) ❌	Yes (Customer Key for Exchange/SharePoint, etc.) ✓	Yes (Client-side Encryption with customer keys for Drive, Gmail, etc.) ✓

Criterion	OpenAI ChatGPT (Free/Plus)	OpenAI ChatGPT (Team/Enterprise)	Anthropic Claude (Free/Pro)	Anthropic Claude (Work/Enterprise)	Microsoft 365 E5 (Office 365)	Google Workspace (Enterprise)
Private network connectivity	No (internet only) ❌	Not directly (internet TLS; use Azure OpenAI for VNET option) 🔶	No ❌	Not directly (internet; available via AWS/GCP integration for private networking) 🔶	Yes (ExpressRoute, private links available) ✓	Partial (VPN or proxy; Google does not offer direct MPLS, but BeyondCorp secures access) 🔶
Integration with DLP/Compliance	No native DLP integration ❌	Not yet (must rely on external proxy or user policy) ❌	No ❌	Not yet (no native DLP; rely on policy) ❌	**Yes** – existing DLP, retention, eDiscovery apply to all content (including AI-assisted) ✓	**Yes** – DLP and Vault apply; Duet AI honours data loss rules ✓
Misuse monitoring & filtering	Basic (OpenAI content policy filters) ✓	Basic (OpenAI automated moderation of prompts/outputs) ✓	Basic (Claude refuses disallowed content) ✓	Basic (Claude's constitutional AI prevents many policy violations) ✓	Advanced (tenant-level policies, e.g. block keywords via DLP; plus MS's own AI content filters) ✓	Advanced (enterprise settings + Google Safe AI system for toxic content) ✓
Enterprise support & roadmap	N/A (no enterprise features) ❌	Yes (dedicated account support, fast model updates, new enterprise features frequently) ✓	N/A ❌	Yes (Claude Enterprise roadmap – e.g. larger context, audit logs, etc. – active development) ✓	Yes (well-established roadmap, uservoice feedback, etc.) ✓	Yes (well-established, with customer councils for new features) ✓

Key observations: OpenAI and Anthropic enterprise offerings score highly in **data control, encryption, and basic compliance**, though they lack some of the **granular admin tools and regional options** that Microsoft and Google provide. The free or personal versions of ChatGPT/Claude are **not enterprise-ready** (no DPAs, weaker data protections), and thus not suitable for business use of sensitive data. Microsoft 365 and Google Workspace excel across almost every category due to their maturity – especially in audit, integration with compliance tools, and global compliance coverage. OpenAI/Anthropic are quickly closing the gap in **core areas like no-training commitments and SOC 2 compliance**, meaning for many enterprises they are now acceptable vendors when used via their enterprise plans. Still, organisations with requirements like **data residency in specific countries, customer-managed keys, or deep admin oversight** will find those features only in Microsoft/Google environments at present.

The **scorecard shows** that if an enterprise's primary concern is **data misuse for AI training**, OpenAI and Anthropic Enterprise have solved that (✓). If the concern is **broad regulatory compliance and internal control**, Microsoft and Google currently provide a more complete toolbox, but OpenAI/Anthropic are steadily improving and can be supplemented with third-party solutions.

Practical Toolkit for Safe Enterprise AI Adoption

Finally, based on the analysis above, we present a toolkit of practical steps and resources for enterprises to confidently deploy OpenAI or Anthropic AI solutions (or similar) while meeting data control, security, and compliance obligations. This includes a **checklist** for readiness, suggested **contract clauses**, and a high-level **risk assessment template**:

Enterprise AI Adoption Checklist

1. **Data Inventory & Classification:** Identify what data your users might input into the AI (e.g. source code, customer emails, personal data). **Classify** this data by sensitivity. *Only allow data of an appropriate sensitivity level* into the AI. For example, disallow secret/confidential data or regulated PII unless specifically approved.
2. **Vendor Selection:** Use the **enterprise-tier** offerings. *Do not* permit use of personal/free AI accounts for work purposes. Ensure the chosen plan (ChatGPT Team/Enterprise or Claude Enterprise) explicitly meets your data usage requirements (no training, deletion, etc.). Verify any necessary **certifications** (e.g. SOC 2 report) via the vendor's Trust Portal.
3. **Contracts & Policies:** Sign a **Data Processing Addendum** (DPA) with the AI vendor. Include **confidentiality clauses** that cover any sensitive data

processed. If required, get a **Business Associate Agreement** (for health data) or any industry-specific addendum. Ensure **Standard Contractual Clauses (SCCs)** are in place if data will leave your jurisdiction (EU/NZ). Negotiate an **acceptable liability cap** and if possible an **IP indemnity** for AI outputs. Incorporate the vendor's commitments (no data sharing, retention limits) as obligations in the contract for enforceability.

4. **Security Configuration:** Enable **Single Sign-On** integration so that only authenticated staff can use the AI platform. Enforce **MFA** via your identity provider. Assign admin roles to a few trusted individuals and enable any available **audit logging** or monitoring features on the AI platform. If the platform allows, set the **data retention setting** to the minimum necessary (e.g. 30 days or less). Disable any features that allow broad sharing of data (for example, OpenAI's shareable chat links, if enabled, or Anthropic's public sharing of GPTs, which by default are off for enterprise).
5. **Access Control:** Limit which users or departments have access initially (e.g. a pilot group). Use the principle of **least privilege** – not everyone may need access to the AI if their role doesn't require it, especially during a trial phase. For API usage, restrict API keys to necessary systems and secure them in a vault.
6. **Internal Policy & Training:** Develop an **AI Acceptable Use Policy**. Include guidelines such as: Do not input personal data about individuals unless it's been anonymised; do not rely on the AI for final decisions without review; do not attempt to bypass safety filters. Train employees on this policy and on how the AI may be used beneficially. Emphasise the point that **no customer-identifiable or highly sensitive info should be entered** unless specifically authorised and logged. Leverage the Privacy Commissioner's guidance – e.g. make sure staff know to **check AI outputs for accuracy and bias** before use.
7. **Privacy Impact Assessment (PIA):** Conduct a PIA (or DPIA under GDPR) before rolling out the AI tool. Document the types of personal data that might be processed, the vendor's safeguards (refer to OpenAI's or Anthropic's privacy documentation), and the expected impacts on individuals. For instance, note that prompts may contain personal data and that individuals have rights to have that data deleted – coordinate how you'd fulfil such requests (likely by deleting conversation records). Address NZ Privacy Principles in this PIA, showing how you'll ensure openness, security, limited retention, ability to correct information if an AI summary is wrong, etc..
8. **Pilot and Evaluate:** Start with a controlled **pilot project**. Monitor the outputs closely and gather feedback. Evaluate whether any **hallucinations or inappropriate outputs** occur and adjust guidance accordingly. Also monitor if users are adhering to input guidelines. This pilot phase can also produce case studies to convince stakeholders of the value or highlight adjustments needed.

9. **Monitoring and Logging:** Set up monitoring: if the AI platform provides **audit logs**, regularly review them for any unusual activity. Additionally, implement **network monitoring** – for example, log all API calls to OpenAI/Anthropic with metadata (but not contents) to detect anomalies (like a single user sending huge volumes of data). If feasible, use a Cloud Access Security Broker (CASB) or proxy that can enforce DLP rules on AI usage (some security vendors are releasing "AI usage control" tools that intercept prompts).
10. **Ongoing Compliance Checks:** Periodically re-assess the arrangement. Stay updated on any **policy or feature changes** by the AI vendor. For instance, if OpenAI changes how data is handled or adds a new feature (plugins, web browsing) that might send data to third parties, evaluate and configure it off if risky. Ensure **data deletion commitments are fulfilled** – you may do periodic requests for confirmation of deletion for audit purposes. Keep an eye on evolving regulations (like EU AI Act) and be prepared to adjust usage or get additional assurances from the vendor when those come into force.

Key Contract Clauses to Include (or verify) When Engaging an AI Vendor

- **No Training on Customer Data:** The contract should state that the vendor will **not use or incorporate your inputs, outputs, or any customer data for the training or improvement of any AI models** outside your own instance. This clause solidifies the verbal promise in legal terms.
- **Data Retention & Deletion:** Specify the agreed retention period (e.g. "Vendor shall retain customer conversation data for no longer than 30 days, after which it will be permanently deleted from all systems, barring legally required exceptions"). Include a commitment that upon termination of contract, all customer data will be deleted. Require the vendor to certify data deletion if requested.
- **Data Segregation and Access:** Include language that **your data will be segregated from other customers' data**, and that the vendor will implement strict logical access controls to prevent any unauthorised access or commingling. Also: "Vendor personnel will only access customer data for troubleshooting or abuse monitoring purposes, and only on an as-needed basis, subject to confidentiality obligations". If possible, require that **access logs** of vendor personnel are available on request (or at least retained for audit).
- **Confidentiality:** Treat any data shared with the AI as you would other sensitive data – a robust confidentiality clause binding the vendor (and its employees, subcontractors) to keep your information confidential, with no disclosure to third parties. This should survive termination of the contract.
- **Security Measures:** The contract or an appendix should detail security controls the vendor commits to: e.g. "Vendor shall maintain ISO 27001

certified information security program, including encryption of data at rest and in transit, regular penetration testing, and compliance with SOC 2 requirements." If specific certifications or audits are required (SOC 2 Type II report, PEN test summary), list those and perhaps set a schedule for receiving updates (e.g. annually).

- **Compliance with Laws:** Include a clause that the vendor represents and warrants that in providing the service, they will comply with applicable data protection laws (GDPR, Privacy Act 2020, etc.) and assist you in fulfilling your obligations. For example, they should assist with data subject requests (e.g. help delete or export an individual's data if somehow it ended up in prompts) and with breach notifications in a timely manner.
- **Data Residency or Transfer:** If data residency is a requirement, specify where data will be stored/processed. If not possible to restrict fully, ensure the contract at least requires adherence to SCCs for cross-border transfer and possibly request notification if new regions are brought into play.
- **Audit Rights:** For highly sensitive situations, include a right to audit or have a third-party audit the vendor's controls, or at minimum to review their audit reports (SOC 2 report, ISO certificate). Microsoft and Google contracts often include the ability for customers to request an audit report or even visit data centres under controlled conditions – smaller vendors may not allow physical audits, but should allow paper audits.
- **Indemnification:** If possible, get an indemnity for third-party claims arising from the service. Two areas: (1) **Data breach** – if the vendor's negligence causes a breach of your data, they should cover direct losses (this is often capped, but even a capped indemnity is good to have). (2) **Intellectual property** – if a third party claims the AI output or the model itself infringes IP and your company is sued, the vendor should defend you. OpenAI's standard terms don't necessarily offer this, but enterprise deals might. Microsoft notably offers this for Copilot – try to negotiate similar from OpenAI/Anthropic if outputs will be widely used.
- **Limitation of Liability:** This will be there to protect the vendor, but ensure it's not too low to be meaningful. Negotiate higher caps for confidentiality breaches or regulatory fines. Also consider a **carve-out from liability cap for data misuse** – e.g. if the vendor *were* to use data for training after all (a breach of contract), that should be outside any cap (since that was the main trust item). This may be hard to get, but it signals the importance.
- **Service Performance & SLA:** Define uptime expectations (e.g. 99.5% uptime) and support response times (critical issues in 1 hour, etc.). While not directly compliance, a strong SLA ensures reliability (an outage can also become a compliance issue if it disrupts service to customers). Also include perhaps a clause that the vendor will not introduce any code or data that creates security vulnerabilities (in other words, no malware, and they will maintain the service to be secure).

- **Termination Assistance:** If you end the contract, ensure you can retrieve your data (any stored conversations, custom models) and that the vendor will certify deletion. Also ensure you're not locked out of your data during the term – you should be able to export conversation logs as needed for legal hold or compliance investigations (OpenAI provides an export option for user data in settings; enterprise should have similar).
- **Regulatory Cooperation:** If you are regulated (finance, govt, etc.), include that the vendor will cooperate with regulatory requests or audits relating to the service. E.g. if the Privacy Commissioner or a financial regulator asks how your data is handled by the AI vendor, the vendor should promptly provide info to help you satisfy the inquiry.

These clauses align the vendor's obligations with your compliance needs and ensure you have recourse if any assurances fail.

Risk Assessment Template (Simplified)

When evaluating the introduction of ChatGPT or Claude enterprise into your environment, consider the following risk categories and control questions. This template can be used as part of a Privacy Impact Assessment or security review:

- **1. Data Privacy Risk** – *What personal or sensitive data might be exposed to the AI service?*
    - Identify data types (PII, financial info, proprietary business data).
    - **Risk:** Personal data could be processed overseas or stored by vendor.
    - **Mitigation:** DPA with SCCs in place? Data minimisation strategies (no names, use IDs)? Consent or notification to individuals if needed? Ability to delete data (Yes, via admin delete in 30 days). Residual risk after vendor's no-training commitment (probably low if only transient processing).
    - **Risk Level:** Low/Med/High (depending on if personal data is involved and its sensitivity).
    - **Actions:** E.g., "Only anonymised data will be used. DPA signed. Acceptable risk." Or if high, "Do not proceed until no personal data or further controls."
- **2. Security Risk** – *Could use of the AI lead to data breach or unauthorised access?*
    - Consider the platform's security: encryption, access control, vendor SOC2, etc.
    - **Risk:** Data intercepted or account compromised.
    - **Mitigation:** SSO/MFA enabled? Strong passwords? Logs monitored for unusual access? Vendor certifications (SOC2) reviewed (Yes).
    - Also, **internal misuse**: An employee might input something sensitive or get a sensitive output and mishandle it.

    - **Mitigation:** Employee training and policy (in place), content scanning of outputs.
    - **Risk Level:** (Probably medium, as with any cloud service – mitigated by encryption and SSO).
    - **Actions:** E.g., "Implement CASB monitoring of AI traffic. Risk acceptable with controls."
- **3. Compliance/Governance Risk** – *Does use of the AI comply with contractual, legal, and regulatory obligations?*
    - Check any industry regulations: e.g. client confidentiality (for lawyers or bankers), bank secrecy, export control (if the AI might output or be trained on controlled technical info).
    - **Risk:** Violation of law or contract (e.g. sharing client data with a third party vendor might breach client contract).
    - **Mitigation:** Obtain client consent or amend terms, or choose not to input client data. Ensure vendor contract has needed clauses (checked). PIA done (this document).
    - **Regulatory engagement:** If in highly regulated industry, have we notified/consulted regulator? (Optional step – e.g. some banks talk to their prudential regulators before using new tech).
    - **Risk Level:** Varies – e.g. for a government dept, if no personal data, compliance risk may be low; if in healthcare with PHI, risk high without BAA.
    - **Actions:** E.g., "Sign BAA for HIPAA compliance – done. Ensure random audits for compliance – scheduled."
- **4. Ethical/Reputational Risk** – *Could the AI produce harmful content or decisions that cause reputational damage or unfair outcomes?*
    - **Risk:** AI gives biased advice or wrong info used in decision-making affecting people. Could lead to complaints or bad press.
    - **Mitigation:** Human review of outputs (established policy). Use of vendor's safeguards (Claude's constitution, OpenAI's filters). Test the AI on sample queries to see if it produces biased results; document outcomes.
    - **Risk Level:** Medium – always a possibility of a bad output, but mitigated by review.
    - **Actions:** e.g., "Implement mandatory review of any external communications drafted by AI. Track any incidents of inappropriate output."
- **5. Operational Risk** – *What is the impact if the AI service is unavailable or has an error?*
    - **Risk:** Outage of service could interrupt business process. Or AI provides an erroneous answer that leads to internal error.

    - **Mitigation:** Not mission-critical system (if true) – have fallback to manual process. SLA in place with vendor for uptime (noted). Validate critical outputs with a second method.
    - **Risk Level:** Low for outage (if not critical), Medium for errors.
    - **Actions:** "Establish fallback workflow for when AI is down (e.g. revert to manual drafting). Provide user training to double-check important answers."
- **6. Vendor Risk** – *Evaluate the vendor's stability and compliance posture*.
    - Look at OpenAI/Anthropic corporate maturity: will they support the product long-term? Do they have sub-processors (like cloud hosts) you need to be aware of?
    - **Risk:** Vendor could change terms or suffer a breach.
    - **Mitigation:** Lock in terms via contract for period. Monitor vendor news (set Google Alert for "OpenAI breach" etc.). Have exit strategy (ability to turn off if needed quickly).
    - **Risk Level:** Medium (new vendor but backed by big tech).
    - **Actions:** "Quarterly review vendor's compliance updates. Prepared to disable if any major incident until resolved."

Fill out such a template with identified risks, existing controls, needed controls, and resulting risk rating. The goal is to demonstrate to compliance officers (and regulators, if needed) that you systematically addressed potential issues and either mitigated them or accepted them knowingly.

By following the checklist, nailing down contract protections, and conducting a thorough risk assessment, an enterprise can responsibly integrate powerful AI tools like ChatGPT or Claude into their workflow. This proactive approach turns what was a high-risk proposition (the "wild west" days of early 2023 GenAI) into a well-managed technology adoption by 2025 – unleashing productivity gains while keeping data secure and regulators satisfied.

APPENDIX

Short answer (TLDR)

No. You cannot treat ChatGPT Plus or Claude Pro as a sealed vault for sensitive material. They are safer than the free tiers, but they are still **consumer products**. Your text is held on the provider's servers, may be reviewed by staff for abuse, and—unless you use the right settings—may still feed future model training (ChatGPT Plus) or limited safety-model training (Claude Pro). Real, if low-probability, leakage vectors remain.

What actually happens to your prompts and files?

Feature	ChatGPT Plus (OpenAI)	Claude Pro (Anthropic)
Default training use	**Yes** – prompts, files and outputs help retrain core models. You must turn off **"Improve the model for everyone"** in *Settings → Data Controls* to stop it.	**No** – Anthropic pledges not to train the main Claude model on consumer data. Only three carve-outs: (1) you explicitly submit feedback, (2) you join an opt-in tester programme, or (3) the prompt is flagged for policy violation (then it can be used to improve *safety* models, not the core model).
Retention if you **do nothing**	Chats stay in your history indefinitely. Deleted chats are removed from all systems within **30 days**.	Chats remain in your history until you delete them. Once deleted they disappear from back-end storage within **30 days**.
Retention for **flagged** content	30 days minimum (OpenAI may keep longer "where legally required").	Inputs/outputs kept **up to 2 years**, safety metadata **up to 7 years**.
Human access	Limited, logged, "need-to-know" access for support, safety checks, legal requests.	Same: authorised staff may review flagged or feedback conversations.
Encryption	TLS 1.2+ in transit, AES-256 at rest.	Same (required for ISO 27001).
Certifications	None specific to consumer tier (SOC 2 covers Enterprise only).	SOC 2 Type II and ISO 27001 cover the whole hosting environment, including Pro.

Key risks you still carry

Risk	How it could bite you	Mitigation options
Unintended training (ChatGPT Plus)	If you forget to disable "Improve the model", your text may become part of OpenAI's training corpus and could re-emerge in a future model output.	Toggle the setting off **before** pasting anything sensitive; use *Temporary Chat* for transient work.
Safety-flag retention (both tools)	A prompt containing, say, a patient diagnosis plus a swear-word could be auto-flagged. That entire record might then sit for two years and feed *safety* classifiers (Claude) or be reviewed by staff (OpenAI).	Keep sensitive or regulated data out of casual prompts. If you must include it, redact personal identifiers first.
Staff or contractor access	Both firms allow limited human review. Malicious insiders are unlikely but remain a theoretical breach point.	Rely on vendor controls (access logs, background checks). For higher assurance, move to the Enterprise/API tier where audit rights and bespoke DPAs apply.
Service bugs & breaches	Example: March 2023 cache bug exposed other users' chat titles and snippets.	Bugs are rare but not eliminated. Avoid putting trade secrets or regulated data in any consumer chatbot.

Risk	How it could bite you	Mitigation options
Jurisdiction & oversight	Data resides in the US. NZ Privacy Act IPP 12 treats that as an offshore disclosure; you must ensure "comparable safeguards". Consumer terms give you no DPA or audit rights.	If compliance matters, insist on Enterprise + DPA, or use Azure OpenAI / Google Vertex AI where you can pick the region.
No customer-managed keys	You cannot bring your own encryption keys. If the provider is breached, encrypted data might still be at risk.	Accept the risk or upgrade to a service that offers CMEK (Microsoft 365, Google Workspace, some cloud LLM deployments).

Practical guidance

1. **Assume nothing is private by default.** In ChatGPT Plus you *must* switch off "Improve the model"; in Claude Pro no switch is needed, but flagged or feedback data still travels.
2. **Strip identifiers.** Replace names, account numbers, or patient IDs with placeholders before prompting.
3. **Use Temporary Chat (ChatGPT) or delete conversations immediately.** That limits retention to 30 days.
4. **Keep truly sensitive or regulated data on enterprise-grade channels.** If you need contractual guarantees (no training, short retention, audit logs, data-processing agreement), use ChatGPT Team/Enterprise, Claude Work/Enterprise, or a region-locked cloud deployment.
5. **Educate staff.** Most leaks come from user error—pasting keys, client data, or unreleased code into a consumer bot. Regular reminders cost little and close the biggest gap.
6. **Monitor public disclosures.** Both providers publish incident reports; subscribe to their security feeds.

Bottom line

- **ChatGPT Plus:** Safe **enough for casual work** if you disable training and avoid personal or highly confidential data. Not suitable for material governed by strict secrecy, privacy, or sector rules.
- **Claude Pro:** Slightly stronger by default—no core-model training—but the same caveats on retention, staff access, and jurisdiction apply.
- **Real-world risk:** Low-frequency, high-impact. A mis-click, a future policy change, or an unforeseen bug could expose data. If that exposure would be career-ending, use an enterprise plan or an on-prem solution instead.

Real-world confidentiality risks for a New Zealand business

Issue	ChatGPT Plus / Claude Pro (consumer, paid)	ChatGPT Team / Enterprise ↔ Claude Team-Enterprise (business)	Microsoft 365 Copilot	Google Workspace Duet AI
Training on your data	**ChatGPT Plus:** *ON by default* – you must switch off "Improve the model for everyone". If you forget, prompts and outputs may be mixed into future model training **Claude Pro:** *OFF by default* – Anthropic only trains on (i) feedback you submit and (ii) safety-flagged prompts for trust-and-safety classifiers, not for the main model.	Both vendors contractually exclude model-training unless you opt-in. OpenAI spells this out for Team/Enterprise; Anthropic does the same for Claude Work plans	Microsoft commits that prompts and responses *never* train foundation models.	Google promises Workspace content is not used to train models outside your domain.
Data location & control	U.S.-hosted; no data-residency choice; no DPA. Deleted chats leave back-end storage within 30 days, but only after manual deletion er.	Admins can set retention (min 30 days) and sign DPAs; audit API available; still US/EU data centres only.	Covered by existing Microsoft 365 regional commitments and NZ-aligned security certifications; customer chooses tenancy region; Copilot traffic stays inside the Microsoft 365 boundary.	Follows Workspace data-region rules and inherits all Workspace ISO 27001/SOC 2, FedRAMP High, etc.; content stays in-tenant.
Human access	Limited staff review for policy abuse (both vendors). Safety-flagged items may be retained up to two years (Anthropic) or indefinitely if required by law (OpenAI).	Same, but Enterprise tiers let admins monitor access via audit logs.	Microsoft disables Azure-OpenAI abuse monitoring for Copilot; prompts stored as M365 artefacts and governed by Purview controls.	Google stores prompts as Workspace objects, protected by existing IAM/DLP; staff access governed by Workspace support model.
Independent certifications	None specific to consumer tier.	SOC 2 Type II for ChatGPT Team/Enterprise and Claude Team/Enterprise.	Microsoft 365 suite (incl. Copilot) carries ISO 27001, SOC 1/2, PCI-DSS, HIPAA, NZ-ISM alignment, etc.	Google Workspace has ISO 27001, SOC 2/3, FedRAMP High (Gemini).

Issue	ChatGPT Plus / Claude Pro (consumer, paid)	ChatGPT Team / Enterprise ↔ Claude Team-Enterprise (business)	Microsoft 365 Copilot	Google Workspace Duet AI
Recent incident history	ChatGPT cache bug (March 2023) leaked other users' chat titles and some billing data. No comparable public breach for Claude to date.	None publicly disclosed.	No Copilot-specific breach reported.	No Duet-specific breach reported.

Key confidentiality concerns for NZ organisations

1. **Automatic opt-in risk (ChatGPT Plus).** If a user forgets to disable training, proprietary text can enter OpenAI's corpus and later surface in model output.
2. **Safety-flag retention (both vendors).** A prompt containing personal health details *and* disallowed language could be flagged and kept for up to two years (Anthropic) for classifier tuning.
3. **Jurisdictional exposure.** Consumer plans store data in the US. Under Information Privacy Principle 12 this is an offshore disclosure; you must be satisfied "comparable safeguards" exist or obtain explicit authorisation Privacy Commissioner.
4. **No customer-managed keys, limited audit rights.** Only enterprise tiers (or Azure OpenAI/Vertex AI) let you set retention, review logs, or bring your own encryption keys.
5. **Platform bugs and insider threats.** Low-probability events, but the March 2023 ChatGPT bug shows they happen. With no DPA or indemnity, liability rests with you.
6. **Hallucinated or re-generated secrets.** Even when not trained, LLMs may reproduce sensitive snippets pasted earlier in the same session; careless sharing can leak data to recipients.

Microsoft 365 Copilot and Google Duet avoid points 1–2: they run inside existing enterprise boundaries, honour tenant permissions, and exclude customer data from model training. They also let you apply NZ-centric controls (Purview, Workspace DLP) and sign DPAs already accepted by most local regulators.

How to use ChatGPT or Claude safely when confidentiality really matters

Level	Recommended practice
Quick, low-risk tasks	• Strip personal identifiers, client names, account numbers. • In ChatGPT Plus: *immediately* disable "Improve the model for everyone" and use **Temporary Chat** for one-. • In Claude Pro: keep feedback buttons off for sensitive prompts.
Team adoption	• Move to **ChatGPT Team** or **Claude Team**, where training is off by default and SOC 2 controls apply. • Set retention to the minimum (30 days) if business benefit outweighs UX loss. • Enforce SSO and role-based access, and review audit logs fortnightly.
Enterprise / regulated data	• Choose **ChatGPT Enterprise** or **Claude Enterprise**, or host GPT-4 in **Azure OpenAI** or Claude via **AWS Bedrock** for AU-Sydney region to keep data in Australasia.

Level	Recommended practice
	• Sign the vendor DPA and privacy schedule; link it to your own confidentiality agreement. • Activate customer-managed keys (OpenAI CMEK preview) or VPC peering where offered. • Apply NZ Privacy Act privacy-impact assessment and update your Information Security Policy to include generative-AI controls.
Process guard-rails	• Mandate **redaction tools** or macro-assisted templating before staff paste text. • Classify outputs as *internal-only* unless reviewed by a human subject-matter expert. • Monitor vendor trust-portals and RSS feeds for breach notices.

Bottom line

- For day-to-day creative drafting **without personal or high-value data**, ChatGPT Plus (with training disabled) and Claude Pro are "good-enough" but still weaker than Microsoft Copilot or Google Duet on contractual safeguards.
- Where **client confidentiality, trade secrets, or regulated information** are in play, rely on enterprise-grade deployments from OpenAI or Anthropic, or use Copilot/Duet that already sit inside your secure SaaS stack.
- Under NZ law you remain the "agency" responsible for any offshore disclosure. Follow the Privacy Commissioner's checklist—obtain leadership sign-off, run a Privacy Impact Assessment, and be transparent with individuals.

Adopt the enterprise tiers, embed redaction and retention policies, and you can reach a risk posture comparable to (though still not stronger than) Microsoft 365 or Google Workspace.