NetApp

# 生成AIが漏洩の入り口に！？ストレージだから守れる機密情報

**井上 耕平**
ネットアップ合同会社
2025/03/18

# Agenda

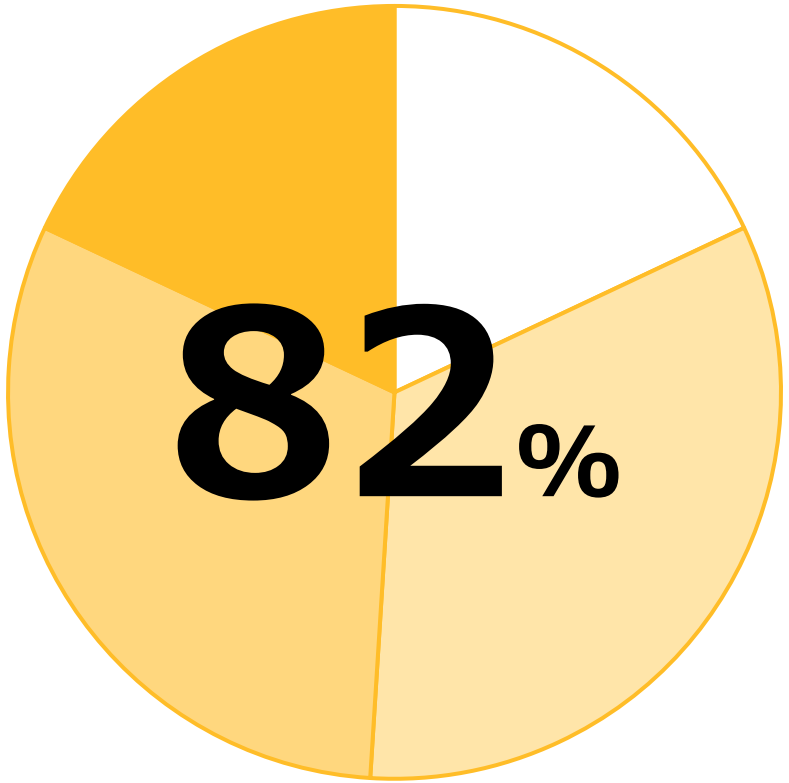
- 生成AIの登場によるセキュリティリスクの変化
- 生成AIにおいて意図せぬ情報漏洩を防ぐ考え方
- ストレージだからできる情報漏洩対策

# 生成AIの登場による
# セキュリティリスクの変化

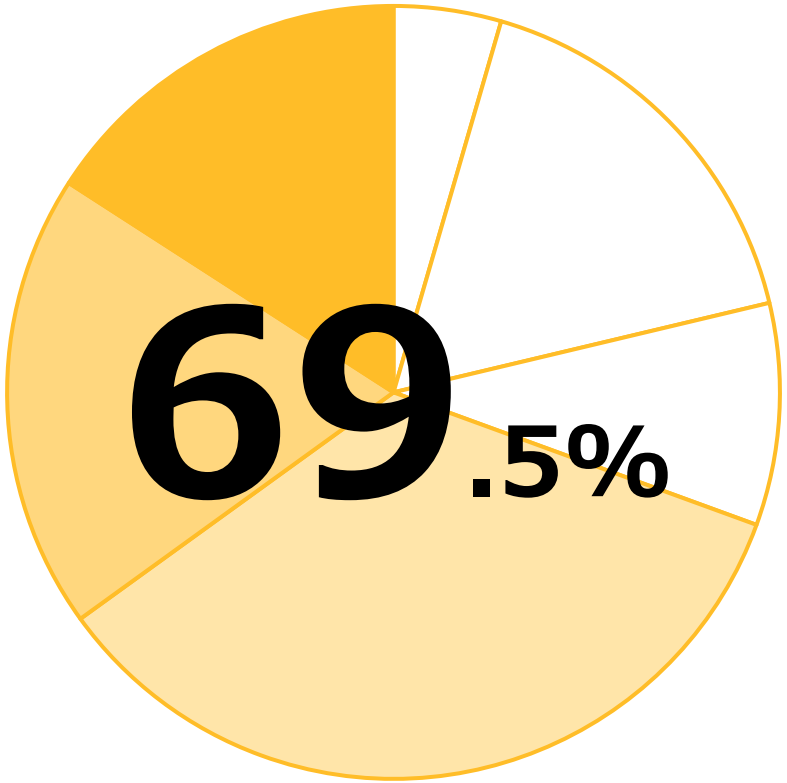
# 日本国内でも進む生成AIの取り組み

2022年11月30日にChatGPTが公開されて以降、グローバルのみならず、
日本国内においても取り組みが進んでいる（計画も含む）

**グローバル**におけるAIの取り組み状況（※1）




**国内**における生成AIの使用状況（※2）

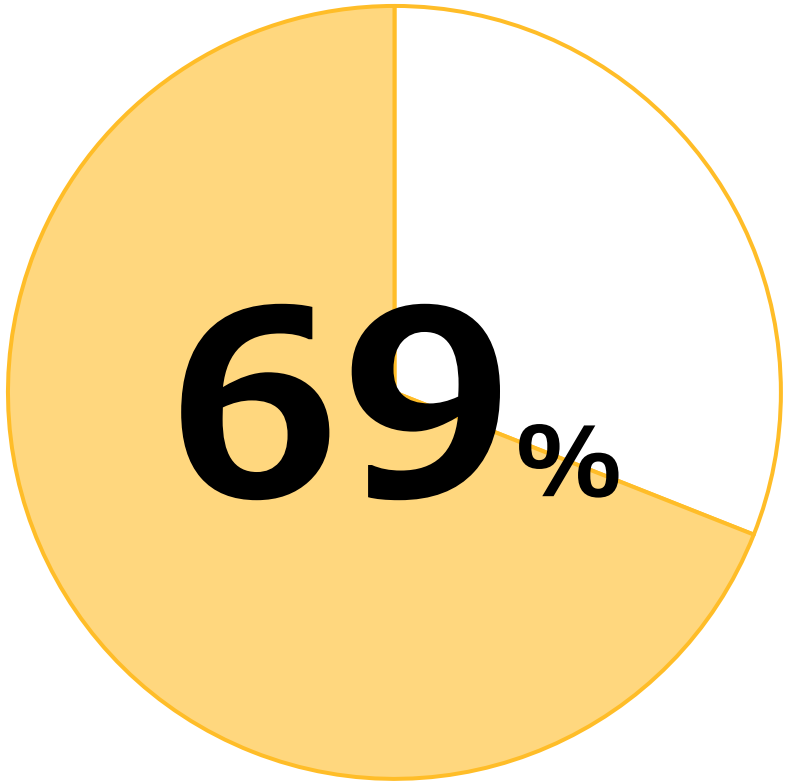



※1：出典 ”2024 CLOUD COMPLEXITY REPORT”, NetApp,（2024/04）、広義のAIを対象とした調査
※2：出典 ”企業IT利活用動向調査2024”, JIPDEC／ITR,（2024/05）

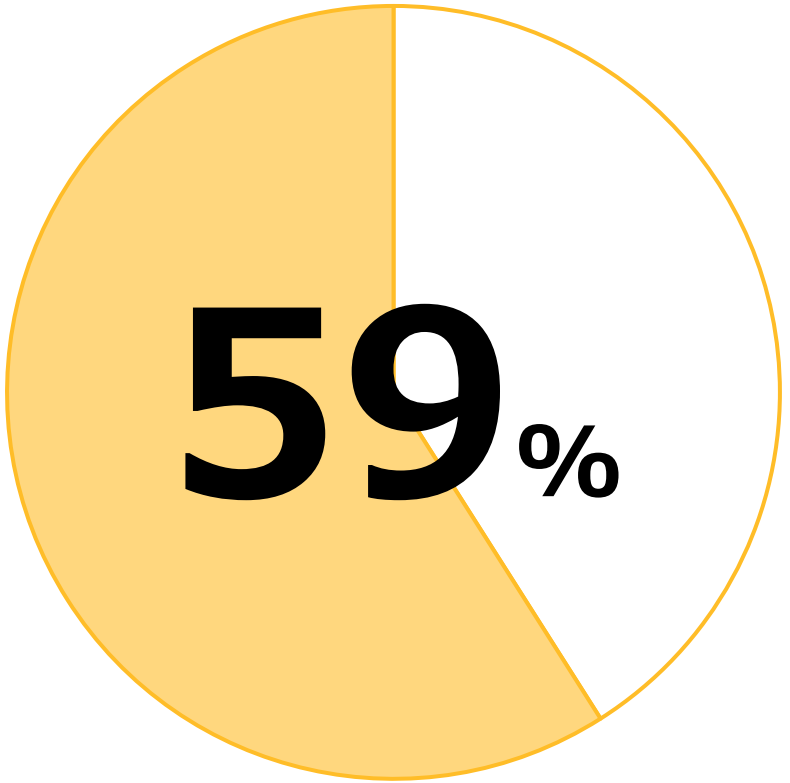
# 一方で、AI導入はセキュリティリスクの増加に繋がる

AIは企業の競争力を高める可能性を秘めているが、新たな仕組みの運用はセキュリティ上の懸念に。

AI導入はセキュリティリスクを増加させたか？（※1）




AI導入によるセキュリティ課題に悩まされているか？（※1）







※1：出典 ”2024 DATA COMPLEXITY REPORT”, NetApp,（2024/12）、広義のAIを対象とした調査

# 生成AIの導入にはどのようなリスクが存在するのか

特に、機密情報の漏洩は、OWASP Top10において前回（LLM06）から大幅に順位が上昇しており、
機密情報の漏洩を防ぐ方法に注目が集まっている

## OWASP Top10 for LLM Application（※1）

- LLM01: プロンプトインジェクション
- **LLM02: 機密情報の漏洩**
- LLM03: サプライチェーン
- LLM04: データとモデルのポイズニング
- LLM05: 不適切な出力処理
- LLM06: 過剰なエージェンシー
- LLM07: システムプロンプトの漏洩
- LLM08: ベクトル化と埋込の脆弱性
- LLM09: 不正確な情報
- LLM10: 際限のない消費

## 国内ガイドラインにおけるAIによるリスク（※2）

### 従来型AIから存在するリスク

- バイアスのある結果および差別的な結果の出力
- フィルターバブルおよびエコーチェンバー現象
- 多様性の喪失
- 生命、身体、財産の侵害
- データ汚染攻撃
- ブラックボックス化、判断に関する説明の要求
- エネルギー使用量および環境の負荷

### **生成AI**で特に顕在化したリスク

- 悪用
- **機密情報の流出**
- ハルシネーション
- 偽情報、誤情報を鵜呑みにすること
- 著作権との関係
- 資格などとの関係
- バイアスの再生産

※1：出典 ”OWASP Top10 for LLM Application”, OWASP,（2024/11）
※2：出典 ”AI事業者ガイドライン（第1.0版）別添1”, 経済産業省,（2024/06）

# 生成AIサービスにおいては既に情報流出事例が報告

現在報告されている事例は、生成AIサービス利用おけるものだが、生成AIシステム全体での情報漏洩パターンを考えると、扱うデータのライフサイクル全体を対象とした**幅広いデータセキュリティ**が求められる

## 報告されているインシデント

- サムスン電子社のChatGPTへの機密情報流出
  - 機密性の高い社内情報をChatGPTに入力してしまう事案が発生
  - 設備情報の流出が2件、 会議内容の流出が1件
  - 緊急措置としてChatGPTへの1質問あたりのアップロード容量を1,024B(バイト)に制限
- リートンテクノロジーズジャパン社のサービス脆弱性による個人情報流出
  - 同社が運営する対話型生成AIサービス「リートン」に脆弱性が見つかり、第三者が利用者の登録情報や入力したプロンプトを閲覧可能な状態だったと発表

## 考えられる情報漏洩のパターン

- 意図せぬ機密情報や個人情報の混入により、サービス利用時に出力可能になる
- 権限管理の不備によりアクセス権のないデータの情報が出力可能になる
- ストレージの設定不備によりデータが窃取される
- システムの設定不備によりモデルやナレッジベースが窃取される
- サービスの利用ログの管理不備により情報漏洩のインシデントが検知出来ない

# 生成AIの登場によるセキュリティリスクの変化まとめ

**国内においても多くのユーザーが**
**生成AIの取り組みを進める**

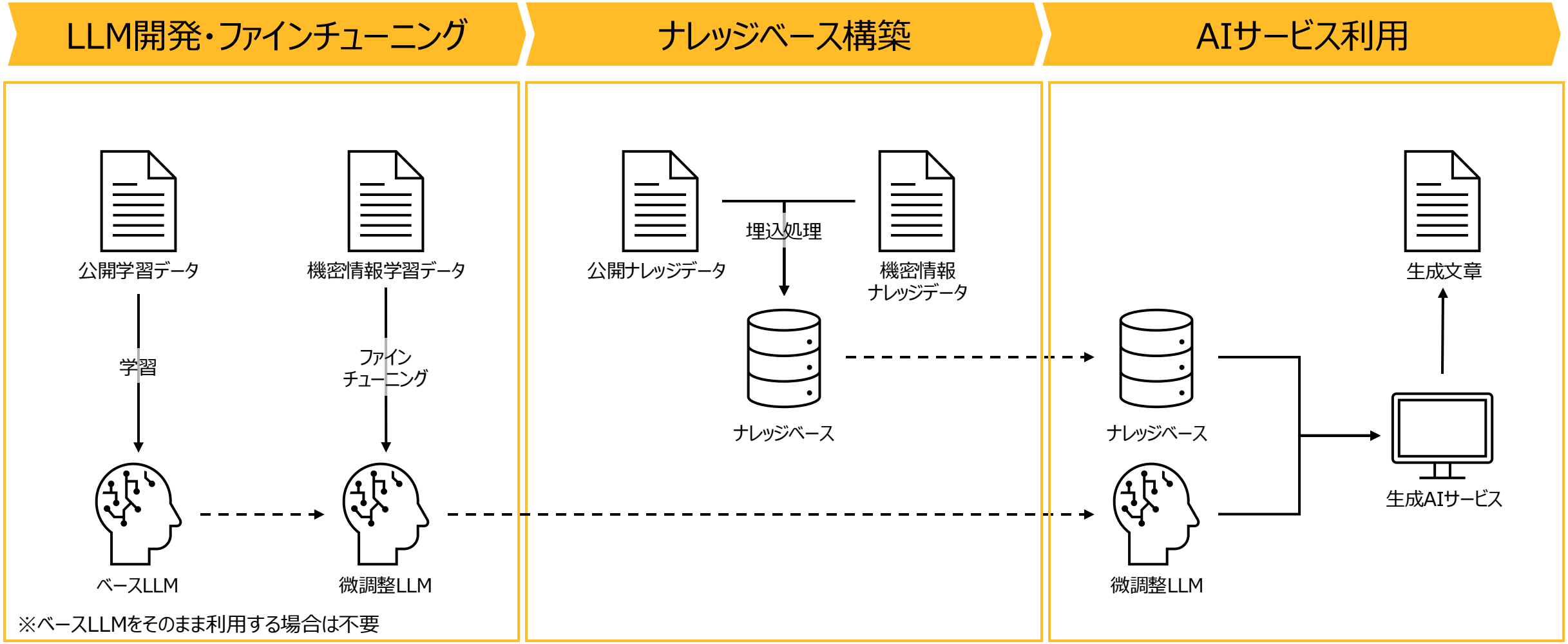
**一方で、セキュリティリスクも増加しており、**
**特に機密情報の漏洩に注目が集まる**

**機密情報は生成AIシステム内で形を変えて扱われるため、**
**ライフサイクル全体でのデータセキュリティが求められる**

# 生成AIにおいて意図せぬ情報漏洩を防ぐ考え方

# 生成AIシステム（LLM）で扱われるデータのライフサイクル

ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。



# 生成AIシステム（LLM）で扱われるデータのライフサイクル

ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。

| LLM開発・ファインチューニング | ナレッジベース構築 | AIサービス利用 |
| --- | --- | --- |

公開学習データ

学習

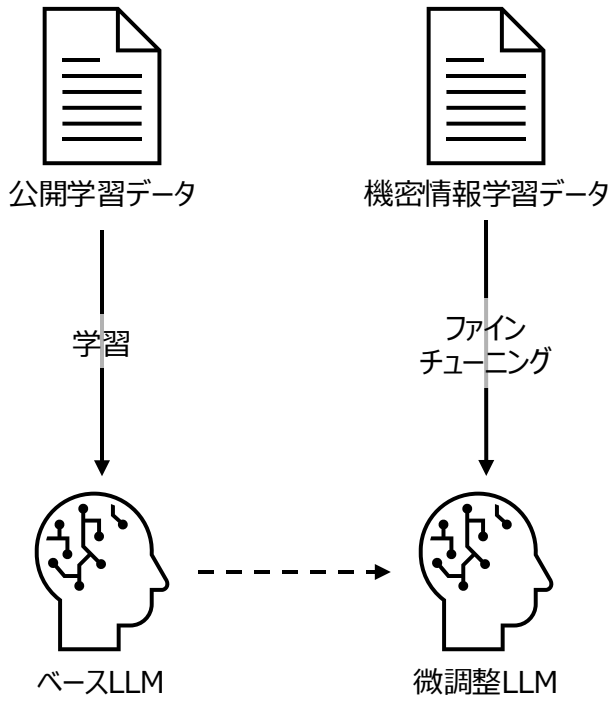
ベースLLM

※ベースLLMをそのまま利用する場合は不要

# 生成AIシステム（LLM）で扱われるデータのライフサイクル

ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。

| LLM開発・ファインチューニング | ナレッジベース構築 | AIサービス利用 |
| --- | --- | --- |

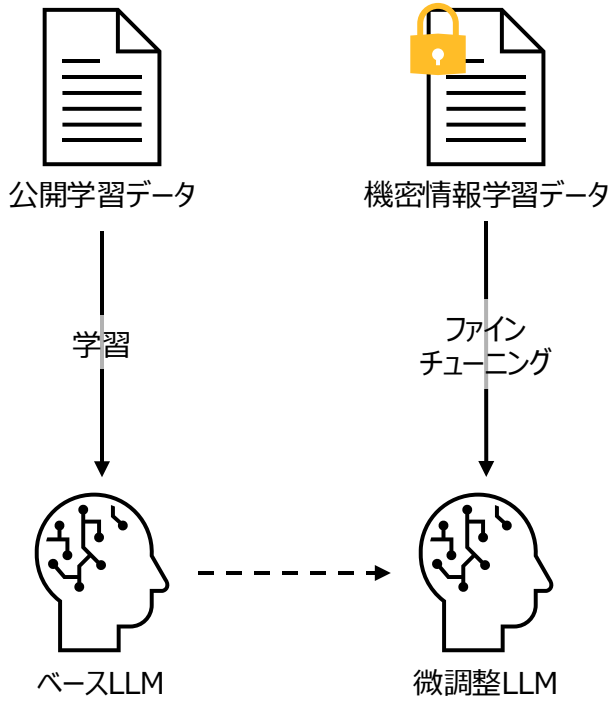



※ベースLLMをそのまま利用する場合は不要

# 生成AIシステム（LLM）で扱われるデータのライフサイクル

ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。

| LLM開発・ファインチューニング | ナレッジベース構築 | AIサービス利用 |
| --- | --- | --- |

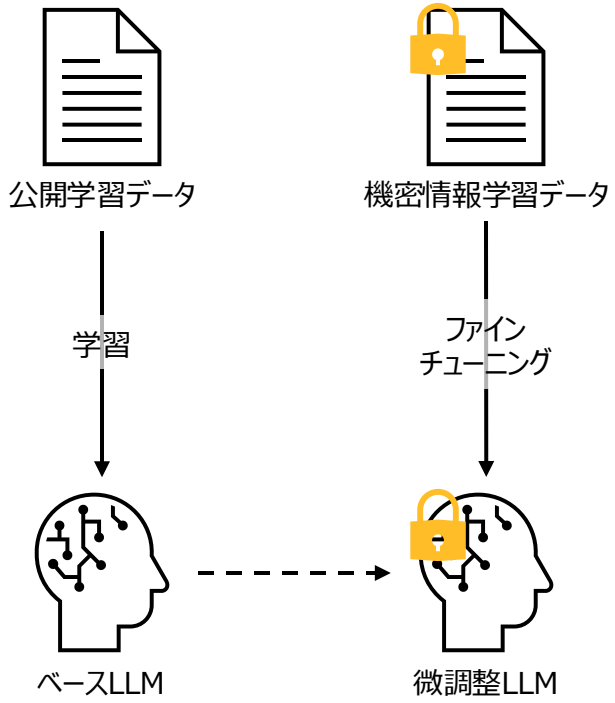



※ベースLLMをそのまま利用する場合は不要

# 生成AIシステム（LLM）で扱われるデータのライフサイクル

ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。

| LLM開発・ファインチューニング | ナレッジベース構築 | AIサービス利用 |
| --- | --- | --- |

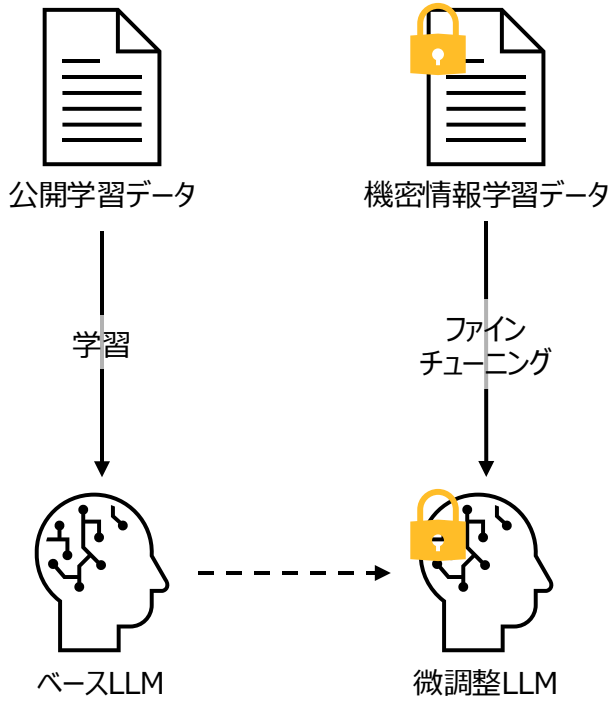



※ベースLLMをそのまま利用する場合は不要

# 生成AIシステム（LLM）で扱われるデータのライフサイクル

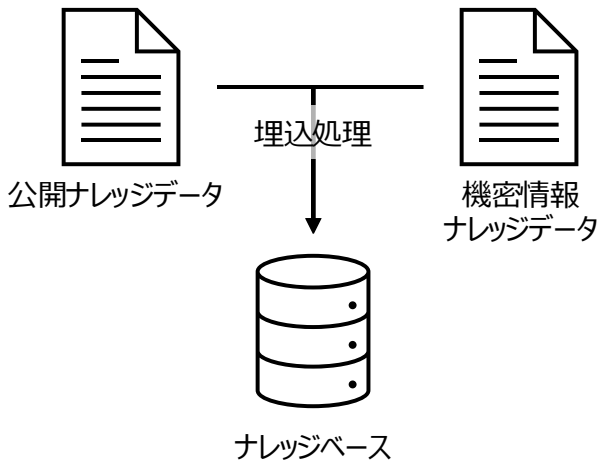
ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。

## LLM開発・ファインチューニング




※ベースLLMをそのまま利用する場合は不要

## ナレッジベース構築

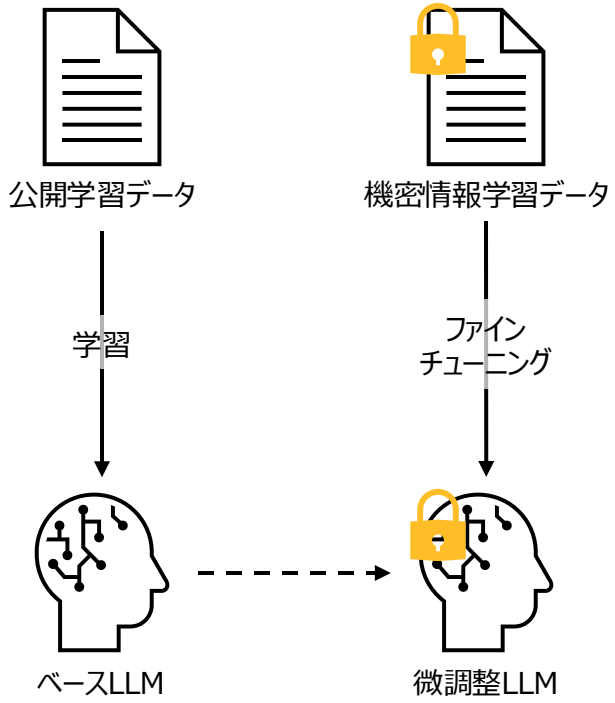



## AIサービス利用

# 生成AIシステム（LLM）で扱われるデータのライフサイクル

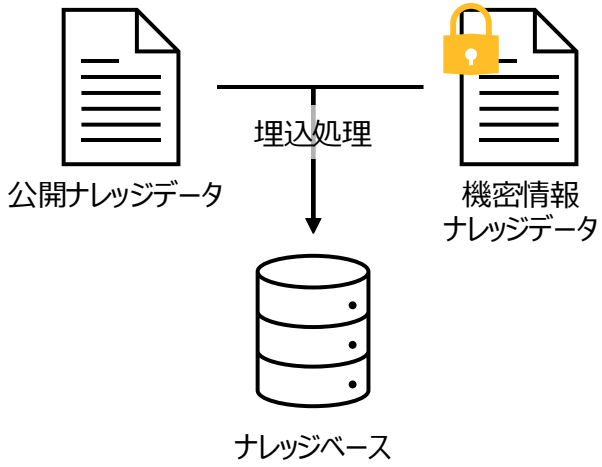
ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。

## LLM開発・ファインチューニング




※ベースLLMをそのまま利用する場合は不要

## ナレッジベース構築

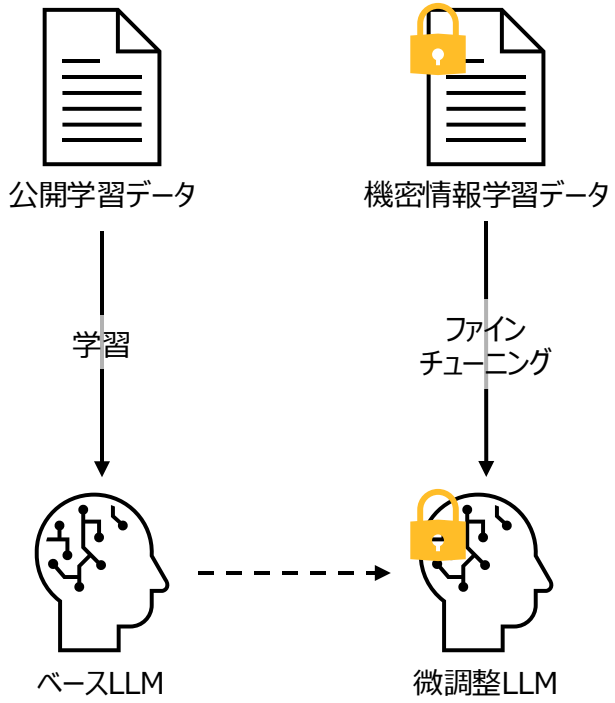



## AIサービス利用

# 生成AIシステム（LLM）で扱われるデータのライフサイクル

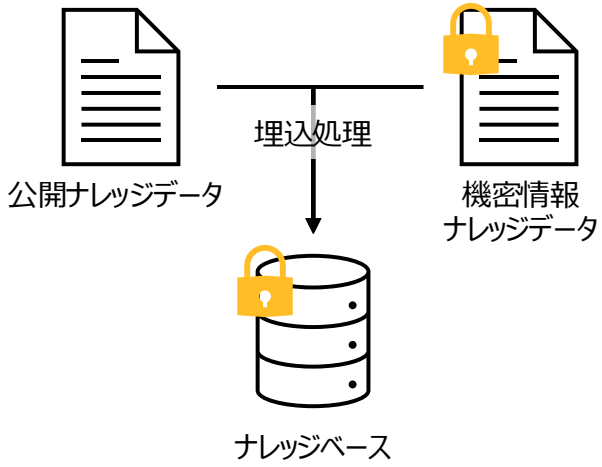
ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。

## LLM開発・ファインチューニング




※ベースLLMをそのまま利用する場合は不要

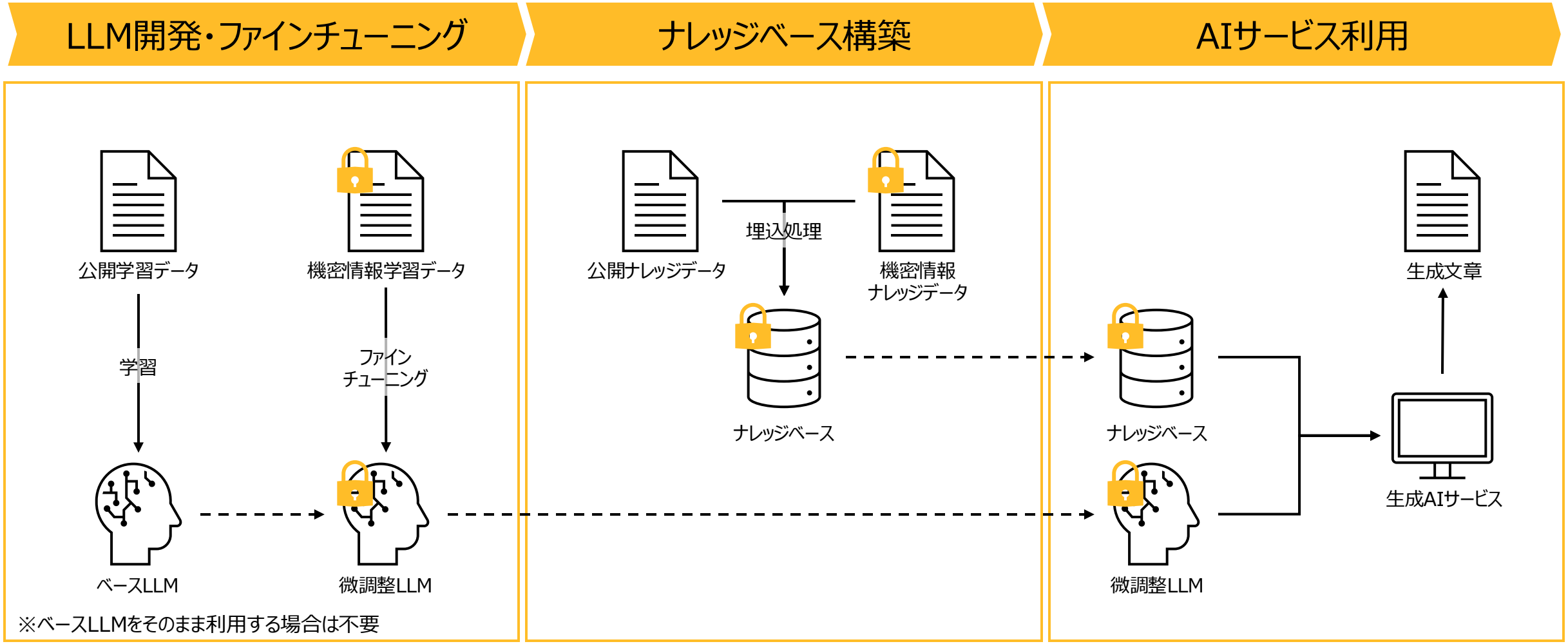
## ナレッジベース構築




## AIサービス利用

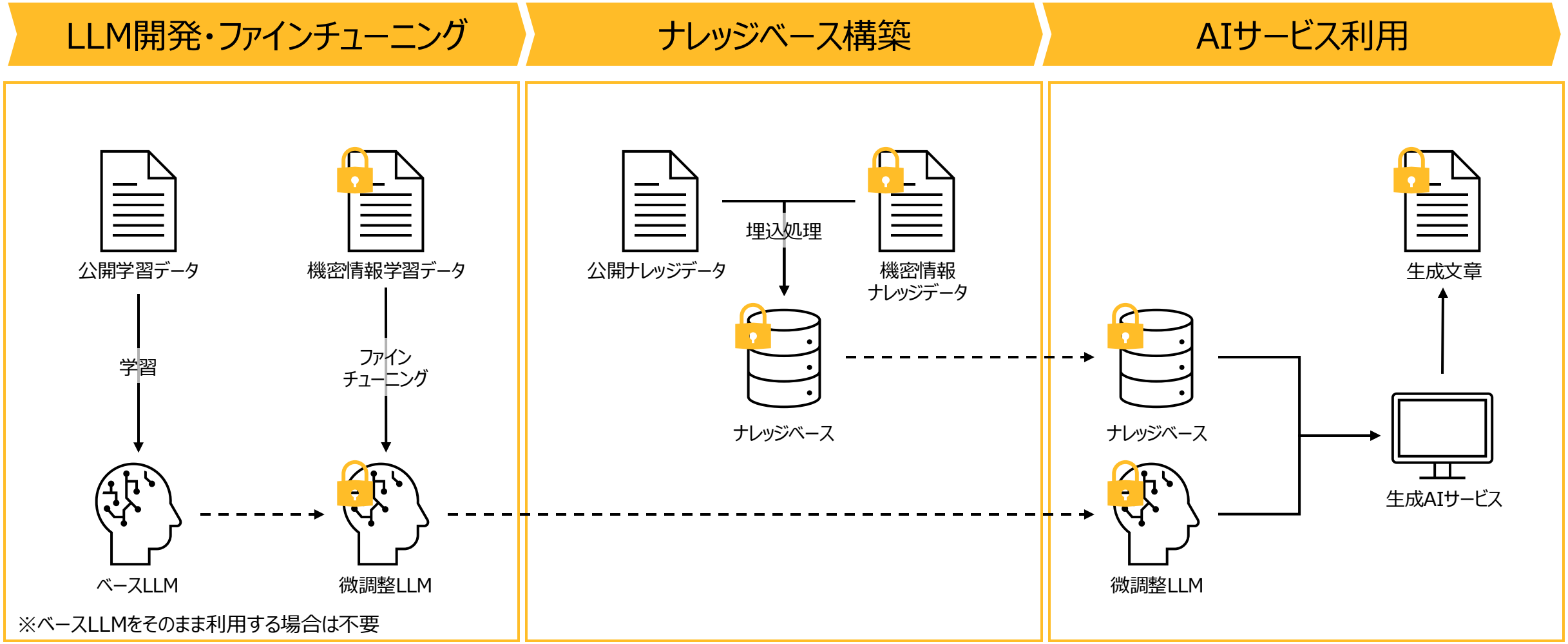
# 生成AIシステム（LLM）で扱われるデータのライフサイクル

ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。



# 生成AIシステム（LLM）で扱われるデータのライフサイクル

ファインチューニングや埋込処理をされた機密情報は、形を変えてLLMやナレッジベースの形で機密情報が保持される。
それらを活用した生成AIサービスから出力された情報にも機密情報が含まれる可能性がある。



# OWASP LLM and GenAI Data Security Best Practice（※1）

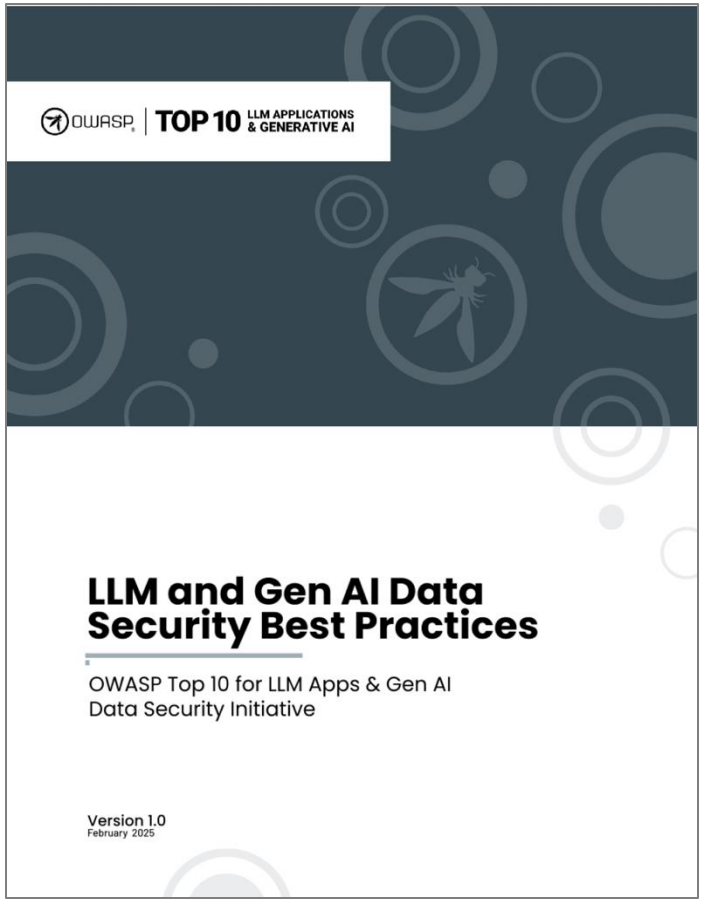
大規模言語モデル（LLM）は、その使用に伴う重大なリスクと脆弱性があるため、データのセキュリティは非常に重要。
データはすべてのLLMの「生命線」であり、その保護の確保には以下が含まれる：

**機密情報の保護**

**データの完全性と信頼性の確保**

**プライバシーに関する懸念への対応**

**新たな脅威への対応**



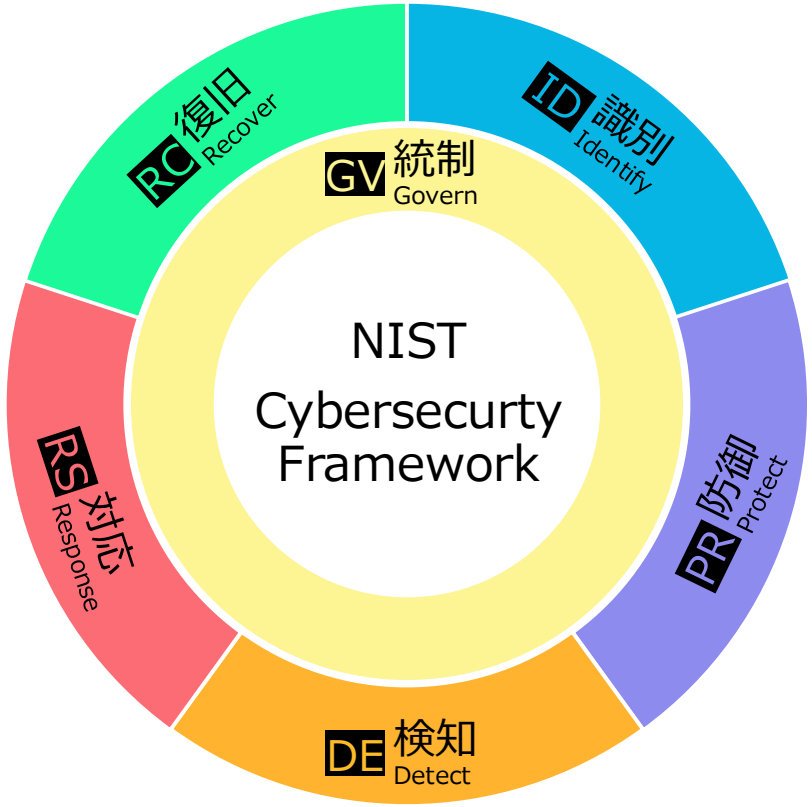

 

※1：出典 ”LLM and Gen AI Data Security Best Practices”, OWASP,（2025/02）

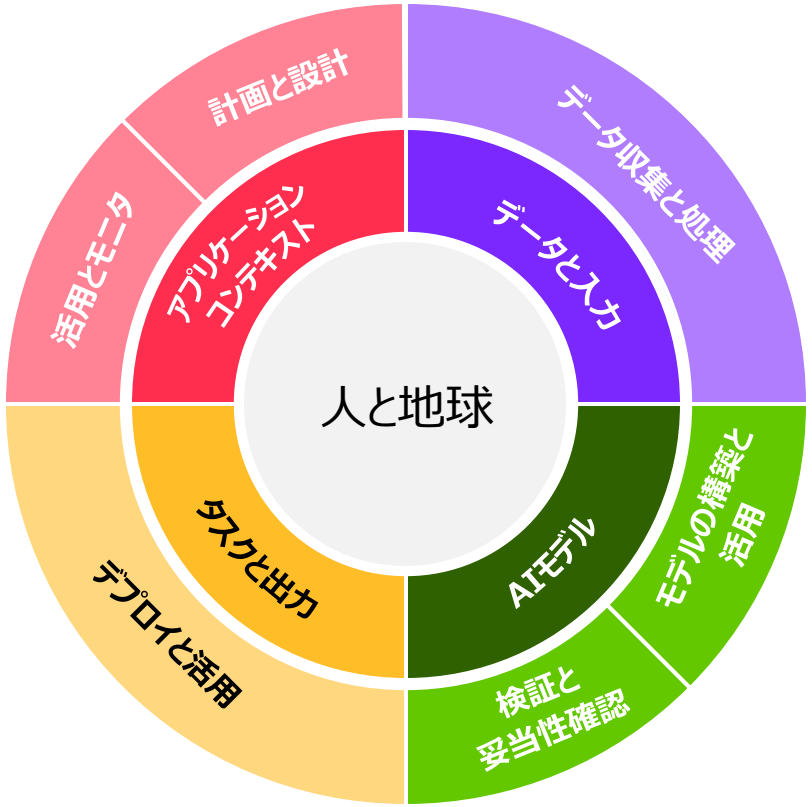
# 世界基準の「NIST」フレームワークをベースとした基本的な考え方

意図せぬ情報漏洩を防ぐ基本的な考え方はNIST サイバーセキュリティフレームワーク（CSF）をベースにしつつ、
AI特有のリスクに関しては、AIリスクマネジメントフレームワーク（AI RMF）も参照

サイバーセキュリティフレームワーク（※1）




AIリスクマネジメントフレームワーク（※2）

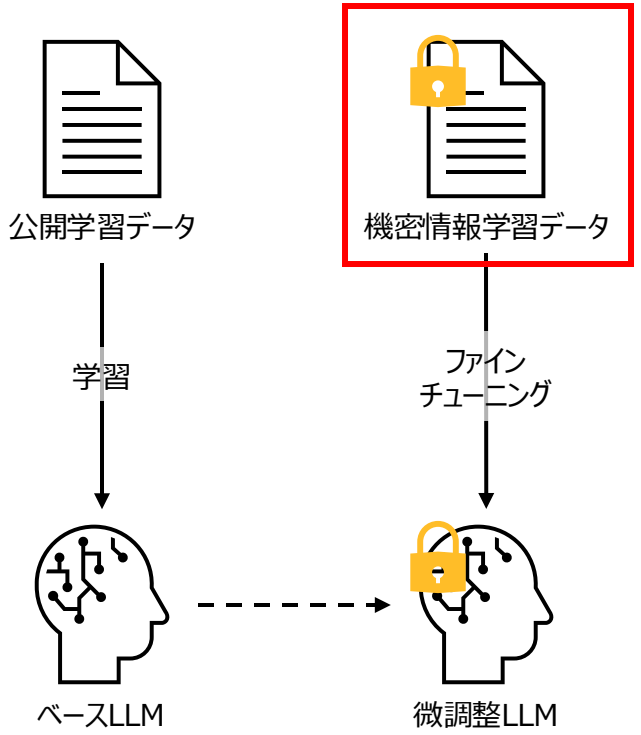



※1：出典 ”The NIST Cybersecurity Framework (CSF) 2.0”, NIST,（2024/04）
※2：出典 ”Artificial Intelligence Risk Management Framework (AI RMF 1.0)”, NIST,（2023/01）

# LLM開発・ファインチューニング時の考慮事項

一般的にLLM開発においては公開されている情報を元にモデルが開発されるが、
ファインチューニングを行う場合は、学習データに機密情報が含まれる可能性がある点に注意

## LLM開発・ファインチューニング




※ベースLLMをそのまま利用する場合は不要

**①学習データは保存時・転送時に暗号化されているか**

**②学習データは機密度などで分類されているか**

**③個人情報など機微な情報はマスクされているか**

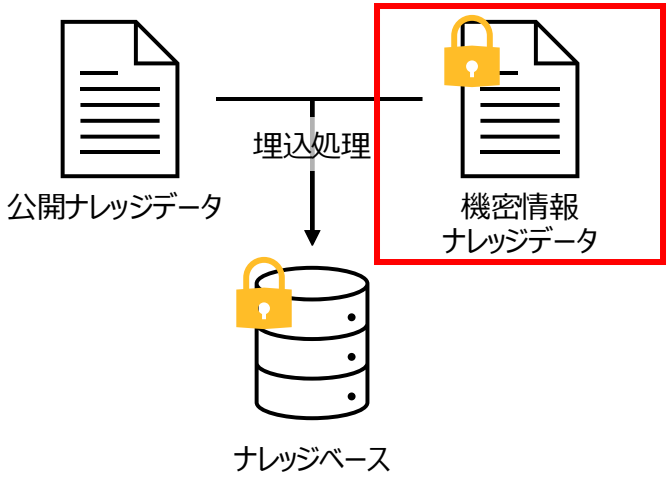
**④学習データやストレージの権限が適切に管理されているか**

**⑤学習データは版数管理、バックアップされているか**

# ナレッジベース構築時の考慮事項

生成AIシステムでは一般的に社内情報などを扱う場合、RAGと呼ばれるナレッジベースを用いる構成を取ることが多い。
その場合、機密情報が埋込処理によってナレッジベース内に存在していることに注意。

## ナレッジベース構築




**①ナレッジデータは保存時・転送時に暗号化されているか**

**②ナレッジデータは機密度などで分類されているか**

**③個人情報など機微な情報はマスクされているか**

**④ナレッジデータやストレージの権限が適切に管理されているか**

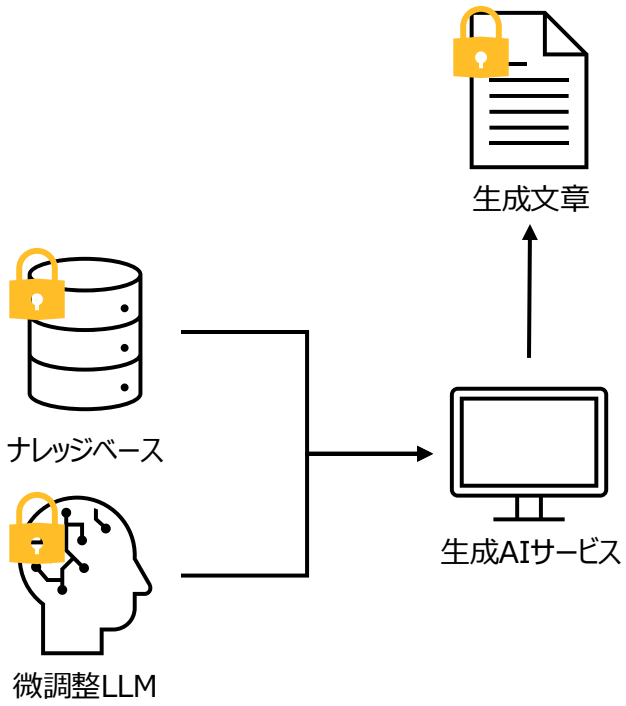
**⑤ナレッジデータの権限がナレッジベースにも同期されているか**

**⑥ナレッジデータは版数管理、バックアップされているか**

# 生成AIサービス利用時の考慮事項

AIサービス利用時はインシデントが報告されているとおり、情報漏洩が起こりやすい。
基本的なWebサービスの認証認可やインシデント発生時の証跡確保、レジリエンスまで考慮することが重要。

## AIサービス利用



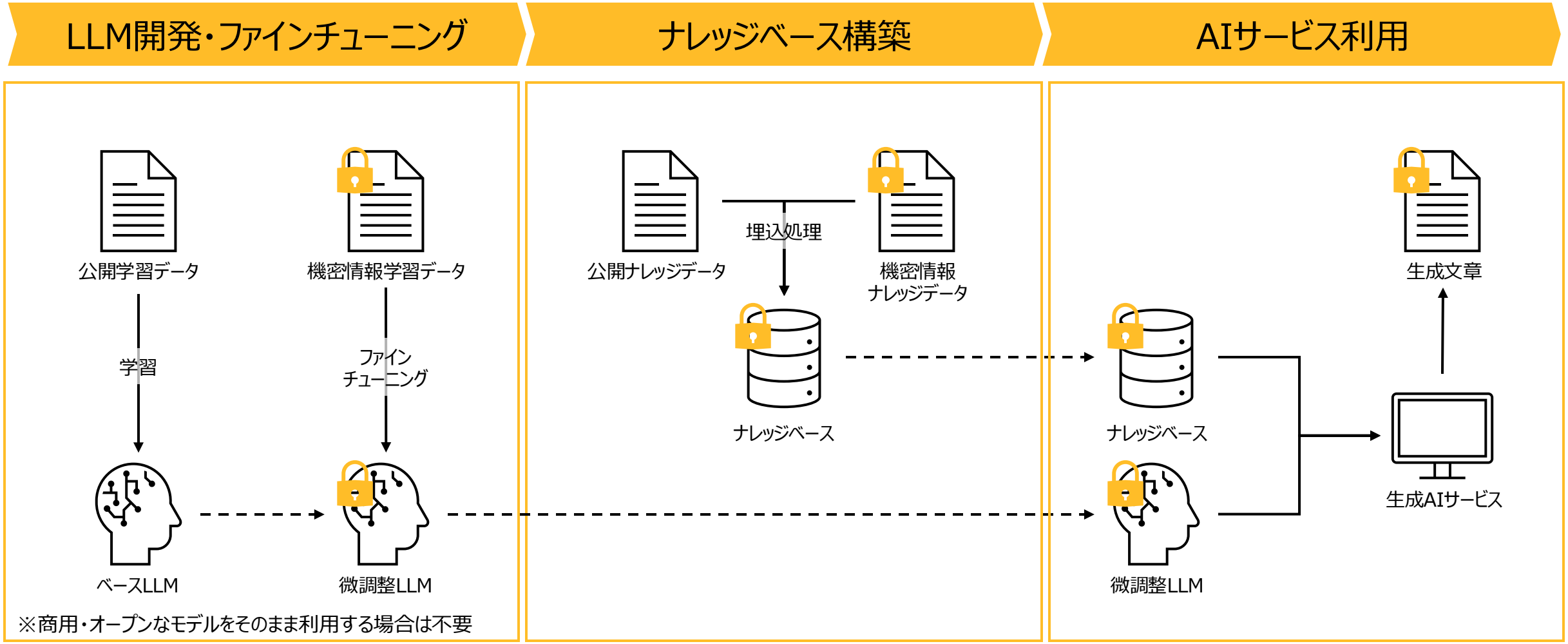

**①出力される情報がユーザー毎に権限管理されているか**

**②出力される情報や利用ログが管理されているか**

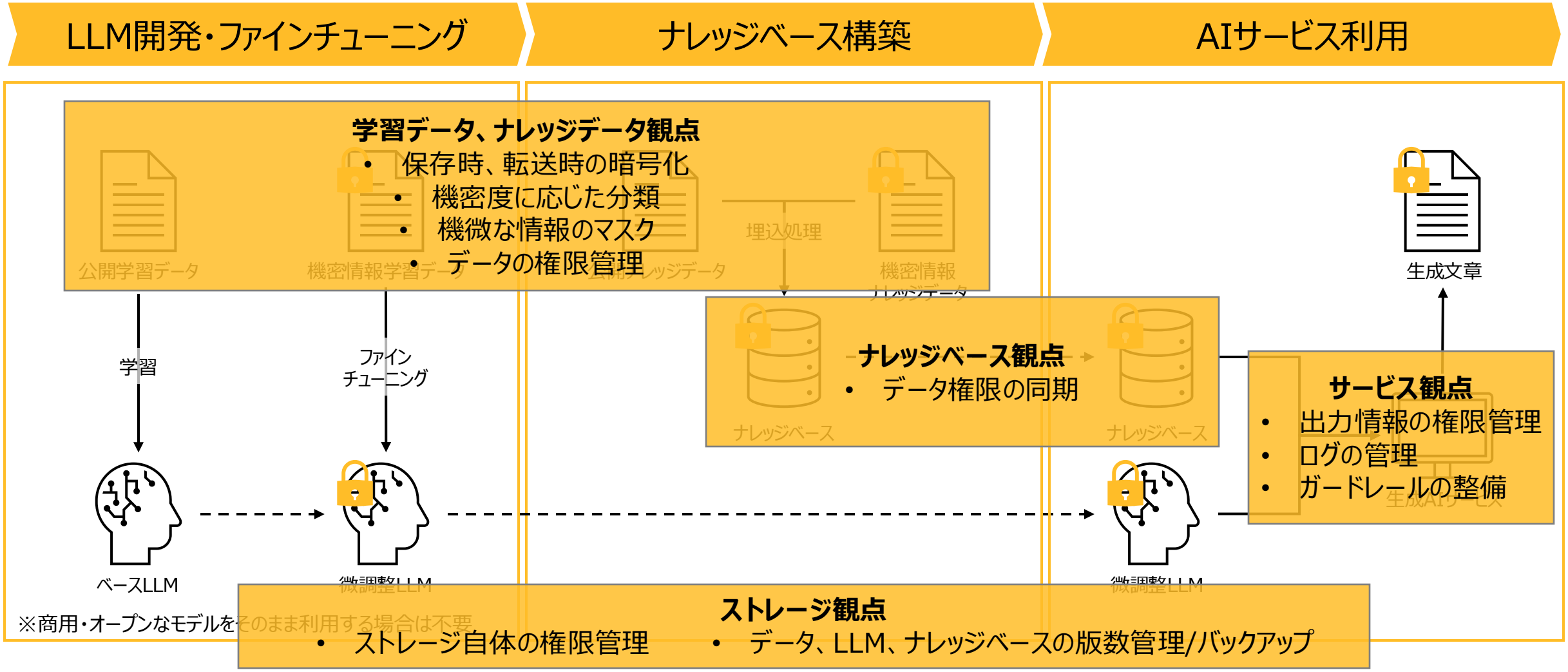
**③出力される情報のガードレールが整備されているか**

**④LLMやナレッジベースが版数管理、バックアップされているか**

# 生成AIシステムにおける考慮事項まとめ



# 生成AIシステムにおける考慮事項まとめ



# ストレージだからできる
# 情報漏洩対策

# 情報漏洩に備えるセキュリティ機能

## NetApp ONTAP テクノロジーを核に、生成AIシステムの情報漏洩に備える

ID 識別
PR 防御
DE 検知
RS 対応
RC 復旧

1. 生成AIシステム特有のセキュリティ
2. 保存時/転送時の暗号化
3. 最小権限の原則と権限保護
4. 各種ログの取得と保全
5. 保護された複数世代のバックアップ

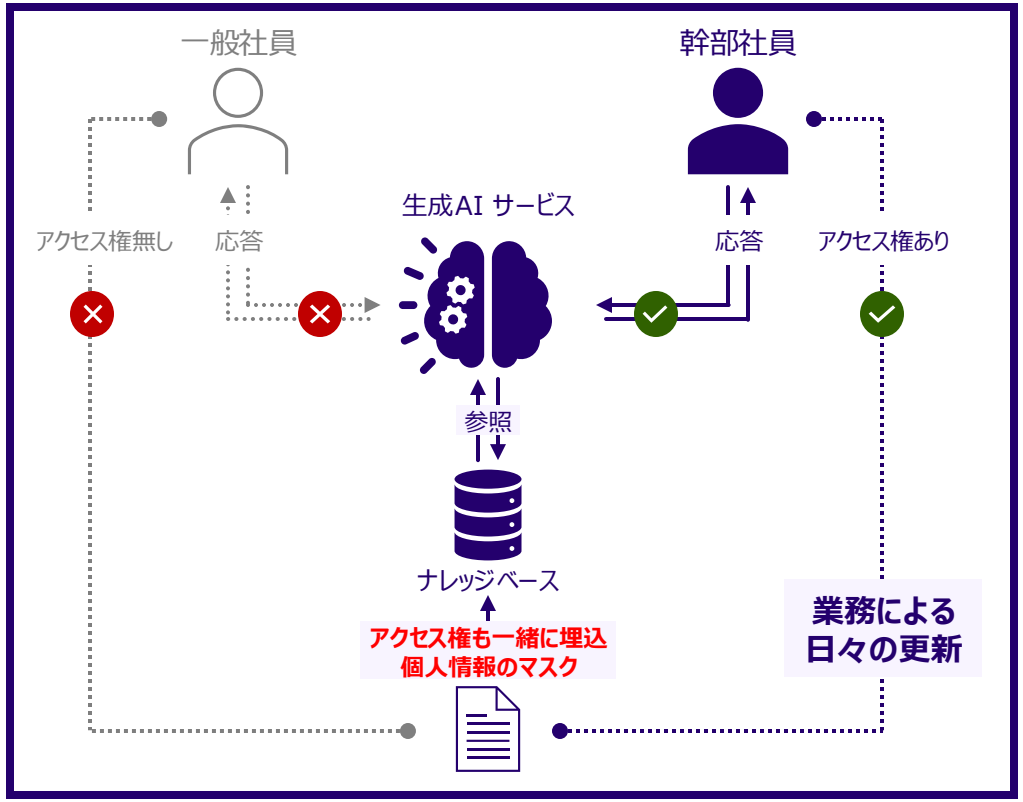
＋ AI for Security

# 1. 生成AIシステム特有のセキュリティ

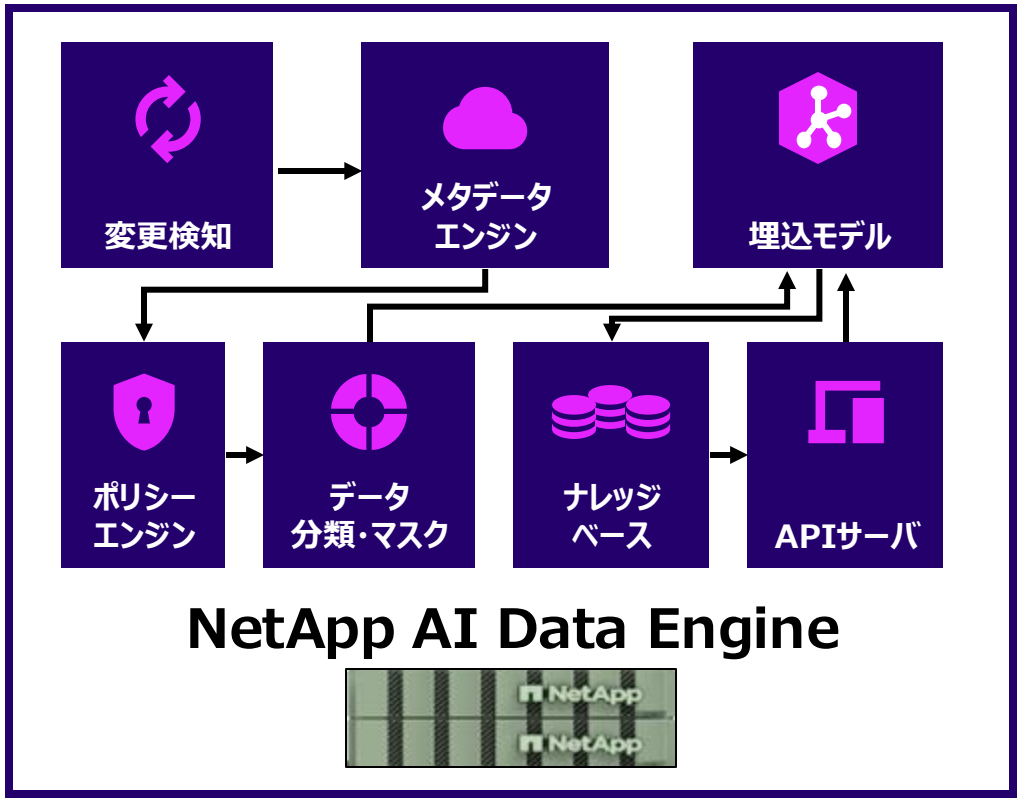
ネットアップは生成AIシステムにおけるデータセキュリティに早くから取り組みを進めており、参照実装を提供。
今後のアップデートで、ストレージ自体の機能でナレッジベースを提供予定。

VISION

## 権限管理やデータガードレイルを実現する参照実装




## 分類、マスク、ナレッジベースをストレージ内で完結



# 2. 保存時/転送時の暗号化

# 多層の暗号化機能を標準搭載

## ハードウェア、ソフトウェアの両面で
## 学習データやナレッジデータを暗号化

標準搭載

NetApp Storage Encryption（NSE）
- ✓ 専用ディスクによるハードウェアレベル暗号化

NetApp Volume Encryption（NVE）
- ✓ ボリューム毎のソフトウェアレベル暗号化

NetApp Aggregate Encryption（NAE）
- ✓ アグリゲート毎のソフトウェアレベル暗号化

転送時の暗号化
- ✓ Cluster Peering Encryption
- ✓ ユーザアクセスの暗号化（SMB暗号化、NFS krb5p）

# 3. 最小権限の原則と権限保護

## 様々な権限保護機能を標準搭載
### 管理者権限を利用した不正操作も防止

標準搭載

- SVMによるマルチテナントとセグメンテーション
  （SVM：Storage Virtual Machine）
- ロールベースアクセス制御
  （RBAC：Role-Based Access Control）
- 多要素認証
  （MFA：Multi-Factor Authentication）
- 複数管理者検証
  （MAV：Multi-Admin Verification）

# 4. ログの取得と保全

## 管理者とユーザーのログを保全（※1）

### ストレージに対する様々な操作やアクセスのログをしっかりと管理

標準搭載

- 管理者監査ログ（管理ログ）
- ユーザー監査ログ（アクセスログ）

※1：ストレージに対する管理ログやアクセスログが対象

# 5. 保護された複数世代のバックアップ

# 特許取得済技術「スナップショット」

## 学習データ、ナレッジデータ、LLM、ナレッジベースを複数世代バックアップ

標準搭載

ファイル書き換え保護機能「SnapLock」
- ✓ WORM機能によるデータやログの改ざん防止

瞬時に取得可能なバックアップ「Snapshot」
- ✓ データやモデルのバックアップ・版数管理

改ざん防止バックアップ「Tamperproof-Snapshot」
- ✓ サービス利用ログ等の証跡の改ざん・消去防止

別筐体・遠隔地へのバックアップ「SnapMirror」
- ✓ 必要に応じて3-2-1バックアップ戦略も実現

# 安全・安心なストレージを提供
## セキュリティの厳しい基準や規制をクリア

米国家安全保障局
NSA CSfC

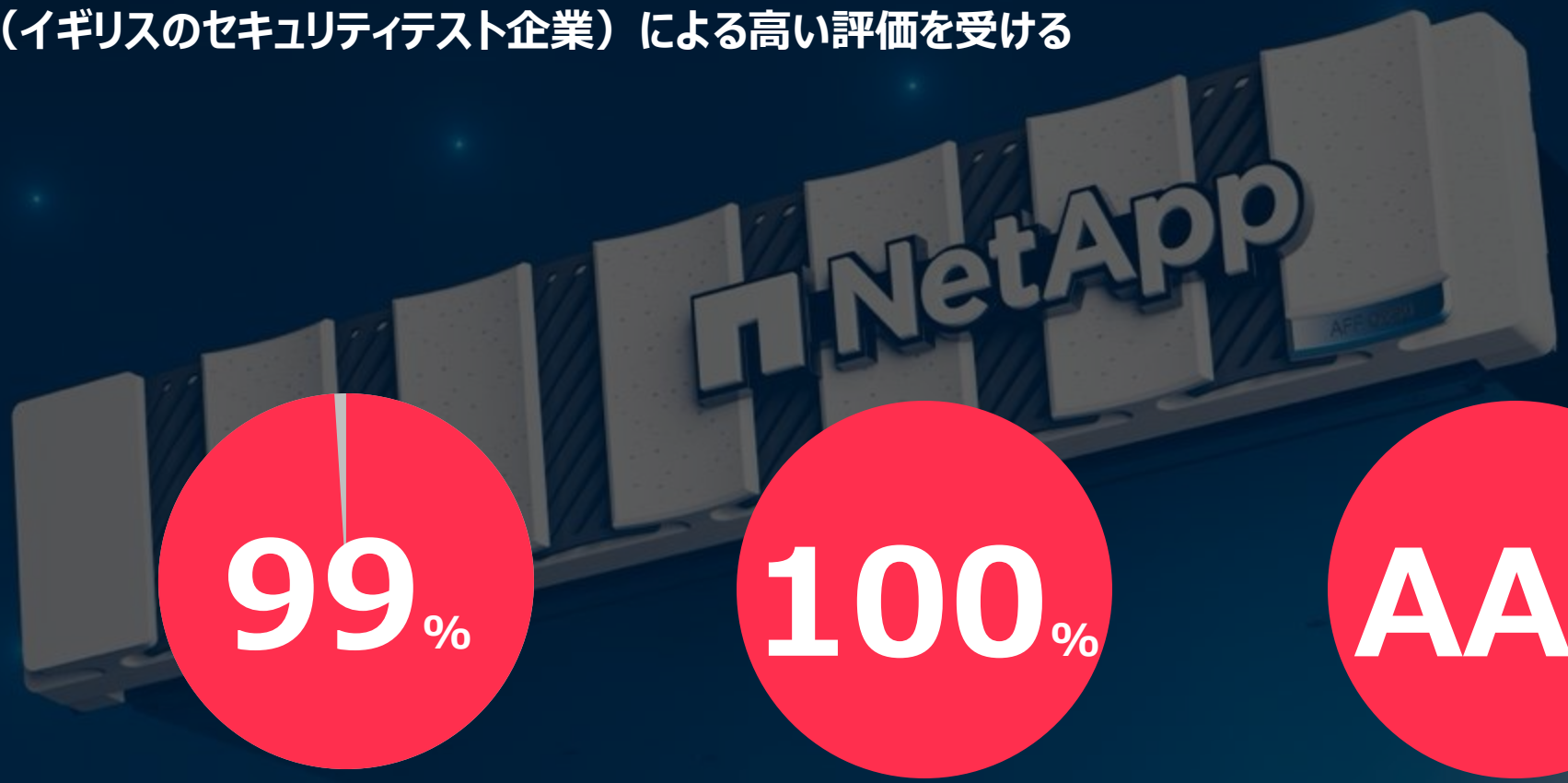
米国防総省
DoDIN APL

国際標準規格
ISO/IEC 15408-1

米連邦標準
FIPS 140-2/3

# AI for Security（AIを活用したセキュリティ向上）への取り組み

**NetAppのランサムウェア対策機能であるAnti Ransomware Protection（ARP）はAI活用により性能が向上し、SE Lab（イギリスのセキュリティテスト企業）による高い評価を受ける**




**99%**

**検知率**
**Detection Accuracy**

**100%**

**誤検知回避率**
**Legitimate Detection Accuracy**

**AAA**

**総合評価**
**Rating**

備考：本テスト結果は、お客様環境でのランサムウェア攻撃の検知率・誤検知回避率を保証するものではありません。

# KEY TAKEAWAYS

## 01

国内においてもAIの取り組みが増加している一方で、AI導入における
セキュリティリスクも顕在化。特に情報漏洩リスクの注目度は上がっている。

## 02

生成AIシステムにおける情報漏洩を防ぐためには、守るべき機密情報が
どこにあるのか、それがどのように変化して扱われているのかを理解し、
ベストプラクティスやフレームワークに沿った対策を行う事が重要。

## 03

生成AI特有のリスクへの対策機能や基本的なセキュリティ機能など、
ストレージだからできる漏洩対策で生成AIシステムの最後の砦としての
役割を果たす。

本資料は2025年3月の内容となっています。

最新の情報については、NetApp 担当営業・SE、

もしくは以下よりお問い合わせください

https://www.netapp.com/ja/forms/sales-contact/

# THANK YOU