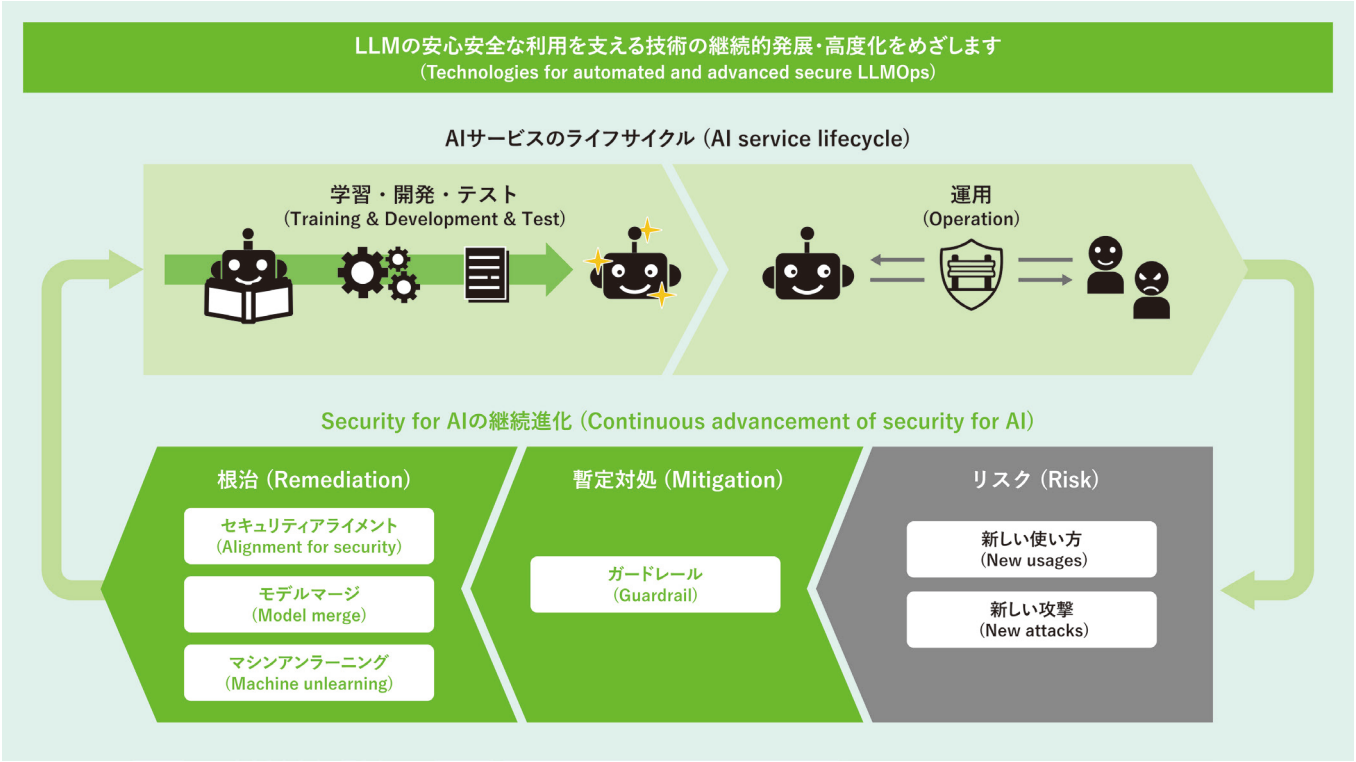
#リスクマネジメント #安心・安全 #ガバナンスと信頼性 #サイバーセキュリティ

Security for AI技術を統合・最適化することで、組織のAI導入を推進

AIライフサイクルを支えるセキュリティ

背 景 – 技術課題

AIは急速な勢いで進化していますが、一方で、生成AIサービスに入力した情報の不適切な利用や、機密情報が漏洩するリスクも明らかになっています。今後も、AIの進化と同時に脅威も増えていくことが想定され、さらなるAI活用には、セキュリティを高めることが必要です。



研 究 目 標 – 成果

AI活用を広げ、AIの利用を継続していく中で、顕在するリスクに対し、迅速な対応が可能な「暫定対処」技術と、時間は要するが根本的な対処が可能な「根治」技術を統合・最適化することで、学習から運用までのAIライフサイクル全体を支えることをめざしています。

技 術 ポ イ ン ト

01 要素技術

- AIの安全性を高める手法とAIの入出力を制御する手法でAIライフサイクル全体を支える技術
- 不要な知識をAIに忘却させ、出力しないようにするマシンアンラーニング技術

02 市中技術差異点

- 継続的にSecurity for AI技術の進化を図り、統合・最適化することで、新しい使い方や新たな攻撃に伴って増大するリスクにも迅速に対応
- LLMの内部表現にアプローチした忘却手法を提案

利用シーン	情報技術	**R&Dフェーズ**	研究
技術確立予定時期	FY25 - FY26	**ビジネス化予定時期**	FY25 - FY26

【出展企業】
NTT株式会社 社会情報研究所

【問い合わせ先】
社会イノベーション研究プロジェクト

【共同出展社/社外連携先】
–

【関連Link】
https://journal.ntt.co.jp/backnumber/2025/vol3709