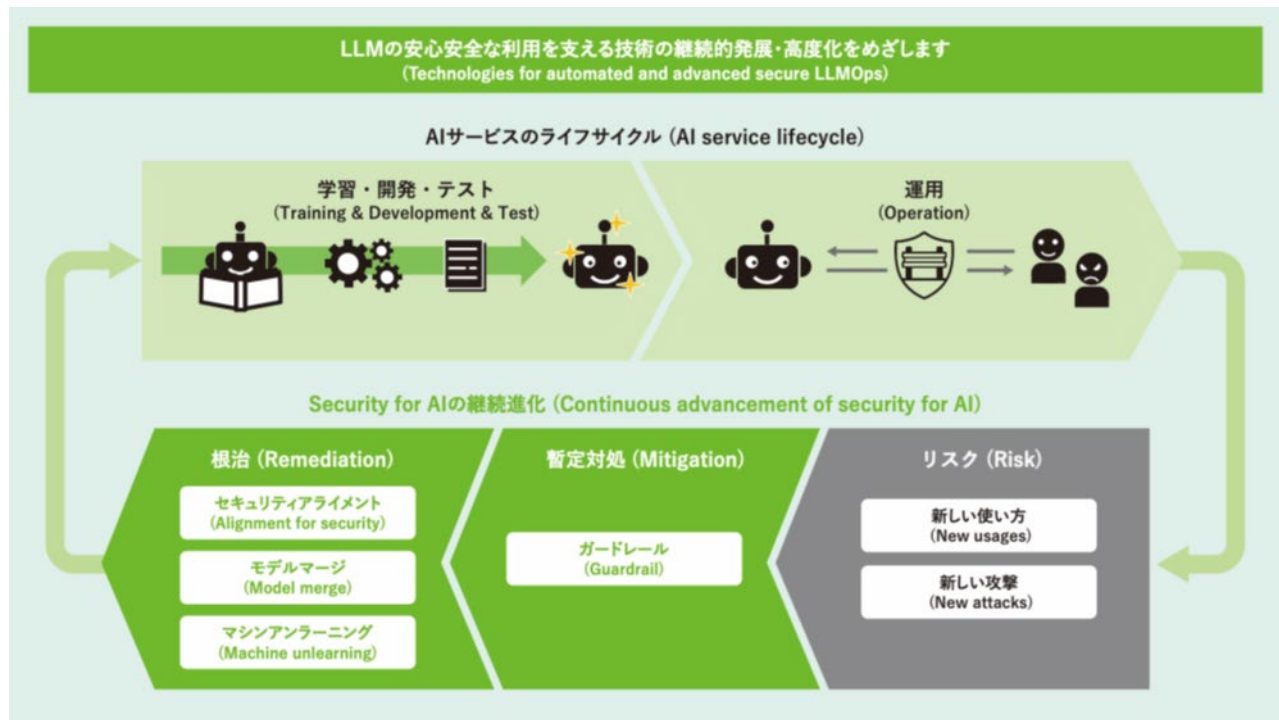


Promoting organizational AI usage by the integration and optimization of security for AI

Security for the entire lifecycle of AI

Background and Technical Challenges

The rapid advancement in AI technologies has been accompanied by growing concerns about the potential misuse of user-provided data in generative AI systems and unauthorized disclosure of sensitive information. As AI continues to evolve, such risks are expected to increase, highlighting our view that robust security measures are essential to ensure its safe and reliable use.



R&D Goals and Outcomes

As the use of AI expands, we aim to support the entire AI lifecycle—from training to operation—by integrating and optimizing two complementary approaches: mitigation enabling rapid response to emerging risks, and remediation requiring more time but resolving issues at their root.

Key Technologies

01 Core Technologies

- Technologies that support the entire AI lifecycle through methods to enhance AI safety and control AI inputs and outputs
- Machine unlearning that enable AI to forget unnecessary knowledge and prevent it from being generated in outputs

02 Key Differentiators

- Continuously advancing and optimizing security for AI technologies to respond swiftly to increasing risks arising from new applications and emerging attacks
- Machine unlearning that manipulates the internal representations of AI

Use Cases Information Technology (IT)

R&D phase Research

Technology Schedule FY25–26

Commercialization Schedule FY25–26

【Exhibitors】
NTT Social Informatics Laboratories

【Contact】
[Social Innovation Research Project](#)

【Co-exhibitors】

【Related Links】
<https://journal.ntt.co.jp/backnumber/2025/vol3709>