特集論文

AIサービスの顧客共創を支える安全なデータ利活用基盤

中井綱人*
Tsunato Nakai

*情報技術総合研究所(博士(工学))

Secure Data Utilization Platform Technologies Supporting Customer Co－Creation in AI Services

要　旨

AIサービスは，顧客接点で蓄積されるログ，画像，センサー値，文書，運用ノウハウなどのデータを活用することで，更に高精度で顧客に適したサービスを提供できる。特に，複数顧客や複数拠点が持つデータや知見を生かして，一つのAIサービスを継続的に高度化していくことへの期待が高まっている。一方，こうしたデータの多くは個人情報，企業機密，知的財産を含むため，単純に集約・学習することは難しい。すなわち，複数顧客の機微なデータを安全に利活用できる基盤が不可欠である。三菱電機は，安全なデータ利活用基盤の構築に有用となる技術として，データとその処理を保護するセキュア実行環境(TEE：Trusted Execution Environment)，及び生データを外部に出さずに共同で学習する連合学習の検討を進めている。

1. ま え が き

AIの活用は研究開発段階にとどまらず，公共，空調・ビル設備，FA，金融，医療など幅広い現場へ広がっている。特に生成AIやAIエージェントの進展によって，AIは単なる分析技術から，業務を支援して新たなサービス価値を生み出す源泉になりつつある。サービスの利便性を高めるには，顧客接点や現場運用を通じて蓄積されるログ，画像，音声，センサー値，文書，利用履歴，運用ノウハウなどを継続的に活用することが重要である。それに加えて，複数顧客や複数拠点の知見を反映し，一つのAIサービスを継続的に改善していくことへの期待も高まっている。

一方，このようなデータの多くは個人情報，企業機密，知的財産を含んでおり，単純に集約して活用することは難しい[(1)]。したがって，AIサービスを高度化したい一方で，その前提になるデータ共有や運用には大きな制約がある。この課題に対する重要な取組みが，データと処理を保護するTEEと，生データを外部に出さずに共同で学習する連合学習である。さらに両者を組み合わせることで，分散したデータを活用しながら，学習過程やモデルも含めて保護する安全なデータ利活用基盤を実現できる。

本稿では，まず**2章**で，安全なデータ利活用基盤の適用シナリオとして，複数顧客の機微なデータを生かしてAIサービスを高度化したいというニーズと，その実現に向けた課題を述べる。次に**3章**で，それらの課題に対応する基盤技術として，TEE，連合学習，及びその組合せの考え方を述べる。さらに**4章**で今後の展望を述べる。

2. 安全なデータ利活用基盤の適用シナリオ

安全なデータ利活用基盤の価値は，単にデータを守ることだけでなく，これまで機密性の高さゆえに十分に活用できなかったデータを，安全性を確保しながらAIサービスへ結び付けられる点にある。特に，組織ごと，現場ごと，機器ごとに分散して蓄積されるデータを適切に活用できれば，AIモデルの精度向上だけでなく，導入範囲の拡大や継続的なサービス改善にもつながる。

この章では，公共，空調・ビル設備，FA，金融，医療などを例に，複数顧客の機微なデータを用いて一つのAIサービスを継続的に賢くしたいというニーズと，その実現に向けた課題を述べる。

2.1　公共空間の見守りと災害時捜索

公共分野では，監視カメラやセンサーから得られる画像・映像データを活用することで，不審行動の検知，混雑状況の把握，転倒や異常行動の早期発見など，多様なAIサービスが期待される。また，災害時には，現場映像や捜索データを解析することで，要救助者の発見や状況把握の高度化が可能になる。これらは，安心・安全の向上に直結する代表的な

ユースケースである。さらに，複数の拠点や自治体，機器群がそれぞれ持つデータを生かせれば，単独の現場では得にくい多様な事例をモデルへ取り込むことができて，見守りや災害対応でのAIの実用性を高めることができる。

一方，公共空間や災害現場で取得されるデータには，個人の行動や位置に関する情報が含まれることが多く，そのまま集中管理・共有することには慎重さが求められる。それに加えて，公共分野では安定かつ長期的な運用継続性も重要である。そのため，この分野では，データを守るだけでなく，必要な処理をどの環境で安全かつ継続的に実行するか，さらに複数地点の知見をどのように生かすかを一体で考える必要がある。

2.2　空調・ビル設備分野での最適制御

空調やビル設備分野では，温度，湿度，人流，設備稼働状況，利用時間帯など，多数のセンサー情報が日々蓄積されている。これらのデータを活用することで，快適性と省エネルギー性を両立する運転最適化や，設備異常の予兆検知，人や空間に応じた制御の高度化が期待される。AIサービスの観点では，機器の販売後も運用データを活用してサービスを改善し続けられる点が大きな価値になる。さらに，複数の建物や設備から得られる知見を生かせれば，個別設備に閉じた最適化ではなく，多様な現場を踏まえたAIモデルの高度化が期待できる。

しかし，建物の利用状況や運転履歴には，生活パターンや業務実態を推測できる情報が含まれる場合がある。また，設備運用の知見そのものが事業上の競争力につながることも多い。それに加えて，エッジ機器上での継続運用や実装性も重要な制約になる。そのため，この分野では，データ，処理，モデルの保護に加えて，長期運用しやすい形で分散した知見を活用できる基盤が必要になる。

2.3　FA分野での品質向上と異常検知

FA分野では，製造ラインの画像，設備状態，検査結果，運転条件など，製品品質や生産性に直結するデータが大量に存在する。AIを活用すれば，不良品検知，製造異常の早期把握，設備の予知保全などを高度化できて，現場の省人化や品質向上に貢献できる。特に，多様な製品や設備条件を学習したモデルほど，現場での適用範囲が広がりやすい。さらに，複数工場や複数ラインに分散した知見を生かせれば，単独拠点では得られない異常パターンや不良兆候もモデルへ取り込める可能性がある。

一方，製造データには，製品仕様，工程条件，不良傾向など，企業にとって重要な機密情報が含まれる。そのため，工場間や企業間でデータをそのまま共有することは難しいことが多い。さらに，停止コストの高い現場で安定運用できることも重要である。したがって，この分野では，データ秘匿だけでなく，現場制約に耐える運用性も含めて考える必要がある。

2.4　金融分野での不正検知

金融分野では，取引履歴，口座利用状況，アクセスログなどのデータをAIで解析することで，不正取引や異常行動の早期検知が期待される。不正検知は，被害抑止に直結するだけでなく，利用者の安心感やサービス品質の向上にもつながる重要な領域である。それに加えて，複数の組織や拠点で蓄積される不正傾向を生かせれば，更に多様なパターンに対応した検知性能が期待できる。

しかし，金融データは極めて機密性が高く，顧客情報や各社の業務ルールを含むため，慎重な取扱いが求められる。また，不正検知モデル自体にも，各組織が蓄積した検知ノウハウが反映されるため，モデル保護の観点も重要になる。それに加えて，金融分野では監査性や説明責任も常時強く求められる。そのため，この分野では，データとモデルを守るだけでなく，運用の継続的透明性も考慮した基盤が必要になる。

2.5　医療分野での診断・予防支援

医療分野では，画像，検査値，診療記録，センサーデータなどを活用することで，疾病の早期発見，重症化予防，診断支援，治療計画支援などへの応用が期待される。医療AIの価値は，単に精度が高いことだけでなく，多様な患者データや運用環境に対応できることにもある。そのため，複数医療機関のデータや知見を生かすことができれば，更に多様な症例に対応できるAIサービスが可能になる。

一方，医療データは個人情報として特に慎重な扱いが必要であり，単純なデータ集約には大きな制約がある。また，医療機関ごとにデータ形式や運用が異なるため，AIの出力が現場によってばらつかず，医療従事者が判断根拠を理解できることが重要になる。そのため，実際のAI導入には技術面だけでなく，信頼性や説明可能性も求められる。したがって，

この分野では，機微なデータを保護しながら，異なる現場の知見をどのように統合していくかが重要な課題になる。

2.6　適用シナリオに共通する課題

それらの各シナリオに共通しているのは，複数顧客や複数拠点が持つ機微なデータや知見を活用し，一つのAIサービスを継続的に賢くしたいというニーズである。一方，その実現には，単にデータを預かって学習させればよいわけではない。データを守るだけでなく，処理を守る，モデルを守る，さらに分散した知見を安全に活用することが求められる。したがって，安全なデータ利活用基盤では，データ，処理，モデル，さらに分散した知見の活用を一体で考える必要がある。

3. 安全なデータ利活用基盤構築技術

この章では，**2章**で示した課題に対する基盤技術として，TEE，連合学習，及びその組合せの考え方を述べる。

3.1　TEEによる保護

TEEは，CPU内部に設けられた隔離領域で，データやAIモデルを保護しながら処理を実行する技術である。通常の実行環境では，処理時間の都合上，データやモデルのAI処理は上記隔離領域の外の通常領域メモリー上で行われる。そのため，メモリー上に展開された機微なデータやモデルが不正な参照，解析，改ざんを受ける懸念が残る。当社では，CPUが備えるTEEを活用しつつ，さらに暗号処理の小型・高速実装技術を組み合わせることで，保護と実用的な処理性能の両立に取り組んでいる(**図1**左)[(2)]。

AIサービスでTEEが有効なのは，学習時だけでなく，推論時にも機微な情報を扱う場面が多いためである。例えば，顧客の利用履歴や現場のセンサー情報を用いて推論する場合，入力データそのものが保護対象になる。また，AIモデル自体にも，長年蓄積した業務知識や運用ノウハウが組み込まれており，知的財産として守る必要がある。TEEは，こうしたデータとモデルの双方を保護しながら，クラウドやエッジ機器上でAI処理を実行できる点に特長がある。さらに，TEEを活用することで，OSやユーザーの信用性を担保できない外部委託先や共有インフラ上でも，更に安心してAI処理を実行しやすくなる。

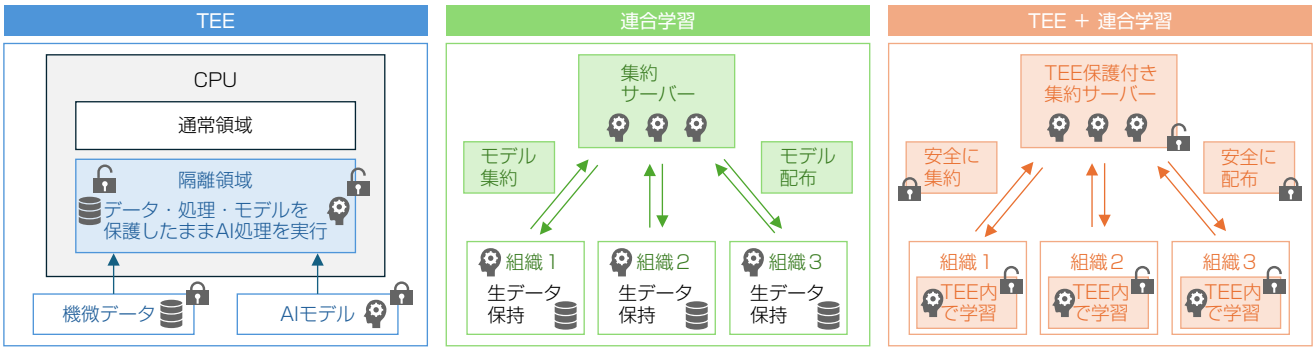

図1－AIサービスの顧客共創を支える安全なデータ利活用基盤の構成

3.2　連合学習による分散データ活用

連合学習は，複数の拠点や機器が持つ生データを1か所へ集めることなく，それぞれの場所で学習を行い，学習結果だけを持ち寄って全体のモデルを改善する技術である。AIサービスでは，公共空間や災害現場の画像・映像データ，空調・ビル設備の運転データ，FA現場の稼働データ，金融機関ごとの取引データ，医療機関ごとの診療・検査データなど，価値の高いデータが分散して存在することが多い。しかし，それらは個人情報，企業機密，知的財産に関わるため，そのまま共有することが難しい。連合学習は，このような条件の下でも，分散したデータを活用してAIモデルの性能向上を図れる点に意義がある。この方式の利点は，各参加者がデータの管理主体であり続けながら，共同で学習できることである。特に，単独の拠点では十分な件数や多様性を確保しにくい場合でも，複数の現場が協調することで，更に実用的なAIモデルに近づけることが期待できる(**図1**中央)。

一方，生データを外部に出さなくても，やり取りされる学習結果やモデル情報から，元のデータに関する情報が推測されるおそれがある。また，悪意ある参加者による改ざんや，不正な学習妨害への対策も必要になる。したがって，連合学習は有力な手法である一方で，安全性を更に高めるためには，追加の保護手段と組み合わせて活用することが必要である。

3.3　TEEと連合学習を組み合わせた高セキュア化

安全なデータ利活用基盤を実現する上では，TEEと連合学習を単独ではなく組み合わせて活用することが有効な場合がある。構築上のポイントは，どのデータを拠点に残して，どの処理をTEE内で保護し，どのモデル情報を共有するかを設計する点にある。連合学習は，生データを外部へ出さずに学習できる点で優れているが，モデル情報や集約処理の保護には課題が残る。これに対してTEEを組み合わせることで，学習結果の集約，モデルの配布，機微な処理の実行を隔離環境内で行い，更に高い安全性を確保できる[3]。この構成を当社が持つ知見と技術によって実現し，各拠点や各機器のデータはその場に保持しつつ，学習に必要な処理や共有されるモデル情報を保護しながら全体最適を図れる基盤を構築する。つまり，データを守る，処理を守る，モデルを守る，という三つの観点を同時に満たすことを可能にする(**図１**右)。

また，TEEと連合学習を組み合わせた基盤は，単なる防御技術にとどまらず，AIサービスを継続的に成長させる。顧客は，安全性が確保されることで，これまで活用が難しかった機微なデータや，組織横断でのデータ活用に踏み出しやすくなる。その結果，サービス導入の障壁を下げるだけでなく，運用開始後も継続的に学習と改善を進められる。

4．今後の展望

今後，安全なデータ利活用基盤で保護すべき対象は，従来の画像やセンサー値に加えて，大規模言語モデル(Large Language Model：LLM)が扱う文書，知識，対話履歴，AIエージェントが用いる業務手順や判断ロジックへ広がっていく。特に検索拡張生成(Retrieval-Augmented Generation：RAG)やAIエージェントシステムでは，参照知識，システムプロンプト，推論過程，外部連携処理まで含めて守る視点が重要になる。また，将来は単独組織内の活用にとどまらず，複数組織や複数拠点が協調してAIを高度化する場面が増えると考えられる。TEEのような保護技術と連合学習のような分散協調技術を適切に組み合わせることで，安全性を確保しながらAI活用の幅を広げられる。

さらに，これらの技術を実際のAIサービスへつなげるため，適用シナリオを想定した試作の開発も進めている(**図２**)。**図２**の試作は，モデル集約を担うサーバーと，学習・推論を行う３台のエッジ機器(クライアント)から構成される。**図２**中の“TEEを用いたエンドツーエンド高セキュア連合学習システム”は，サーバーとクライアントを含む全体構成を指す。試作を通じて，各クライアントのデータ・処理・モデルの保護，及び複数拠点の知見を安全に活用する流れを確認し，性能・運用面の課題を明らかにしていく。

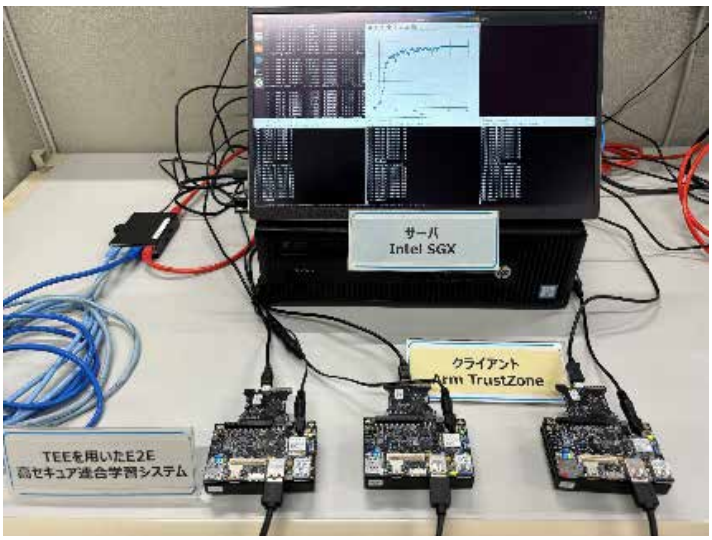

図２-サーバーと３台のエッジ機器で構成する試作

5. む　す　び

AIサービスの顧客共創を支える安全なデータ利活用基盤技術として，TEEと連合学習の取組みを基に，データ，処理，モデルを守りながらAI活用を進める考え方を述べた。公共，空調，FA，金融，医療などの適用シナリオに共通するのは，安全に使えること自体がAIサービスの成立条件である点である。

今後も，LLM，RAG，AIエージェントへ展開しながら，機微なデータやノウハウを守りつつ価値の高いAIサービスの実現に貢献していく。

参 考 文 献

(1) 小関義博，ほか：AIの安心・安全を守るセキュリティー技術とプライバシー技術，三菱電機技報，98，No.8，5-01～5-04 (2024)
(2) Nakai, T.，et al.：Co-designing Trusted Execution Environment and Model Encryption for Secure High-Performance DNN Inference on FPGAs，2024 IEEE International Symposium on Circuits and Systems(ISCAS) (2024)
(3) 中井綱人，ほか：ビザンチンロバストな連合学習における学習モデル保護の基礎検討，2022 Symposium on Cryptography and Information Security (2022)