

**東京大学・NTT・NEC、安心・安全を支える AI エージェント普及の実現に向け
6G/IOWN 基盤に 3 者技術を統合し、リアルタイム AR 支援の実証に成功
～AI エージェントが扱う大容量データの通信及び計算処理を最適化し遅延を低減～**

発表のポイント:

- ◆ 3 者の技術を統合し、AI エージェント実装時の通信・計算インフラの課題解決に貢献
- ◆ AI エージェントが扱う大容量データの通信及び計算処理を最適化し、実証により有効性を確認
- ◆ MWC 2026 Japan Pavilion への出展に採択され、研究成果及びコンセプトを国際的に発信予定

国立大学法人東京大学大学院工学系研究科(所在地:東京都文京区、研究科長:加藤 泰浩、同研究科中尾研究室教授:中尾 彰宏、以下「東京大学」)、NTT 株式会社(本社:東京都千代田区、代表取締役社長:島田 明、以下「NTT」)、日本電気株式会社(本社:東京都港区、取締役 代表執行役社長 兼 CEO:森田 隆之、以下「NEC」)は、安心・安全な社会を支える AI エージェント※1 普及の実現に向け、6G※2/IOWN※3 基盤に 3 者技術を統合しました。本取り組みでは、ストリーミングセマンティック通信技術※4、生成 AI 向けメディア制御技術※5、In-Network Computing (INC) アーキテクチャ技術※6 を統合し、AI エージェントに必要な大容量データ通信及び計算処理の最適化を実現しました。実証を通じて、提案技術の有効性を確認しました。本取り組みは、Mobile World Congress 2026(以下「MWC 2026」)Japan Pavilion での出展を予定しており、研究成果及びコンセプトの国際発信を行います。

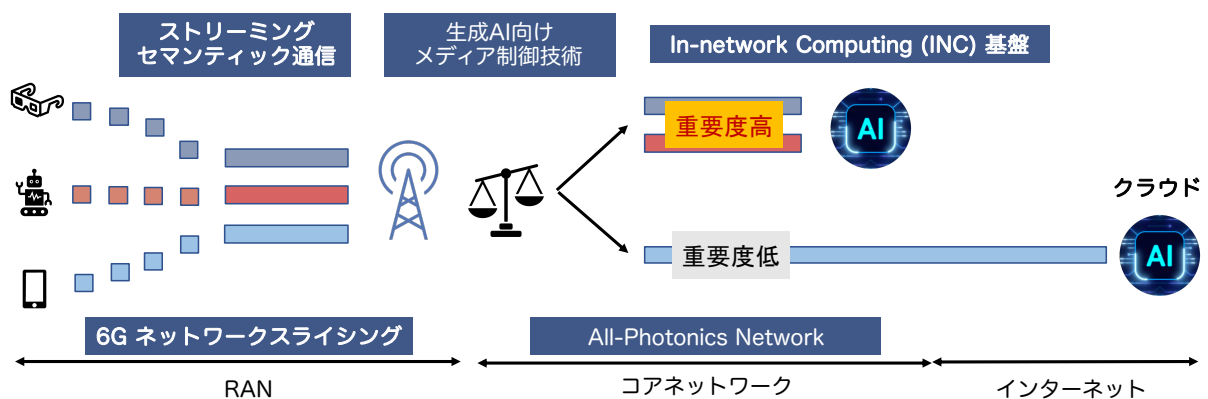


図 1. 6G/IOWN 基盤技術統合の全体像

1. 3 者連携の背景と意義

情報通信は、世界中のどこでも誰とでも利用できる共通基盤であり、その発展には国際連携及び国際標準化が不可欠です。そのため、グローバルな価値を起点に検討することが重要となります。近年では、グローバルな価値として、安心・安全の重要性が一層高まっています。

安心・安全な社会の実現には、災害や事故、サイバー攻撃など、刻々と変化する状況を迅速に把握し、適切な対応が求められます。しかし、こうした対応を人手や事前に定めたルールだけで行うことには限界があり、AI の活用が重要になります。人を支援する AI エージェントを用いる事で自律的かつリアルタイムに状況を認識でき、被害の未然防止や影響の最小化に貢献します。

一方で、AI エージェントを高度に機能させるためには、膨大なデータを低遅延かつ高い信頼性で処理・伝送できる次世代 ICT インフラが不可欠です。

このような背景のもと、6G 及び IOWN に強みを有する東京大学、NTT、NEC が、東京大学の「社会連携講座」制度のもとで結集し、安心・安全を支える AI エージェントの実現に向けた、6G/IOWN 基盤技術の研究開発に取り組んでいます。

2. 研究背景

近年の AI 活用は、人間からのプロンプト入力を起点に動作する形態が中心です。一方、今後の社会では、人間の明示的な指示に依存せず、センサなど非人間からの入力を起点として自律的に動作する AI エージェントの普及が進むと考えられます。このような進化により、周囲環境のデータを常時収集・監視し、異常の兆候を検知した際にリアルタイムで対応する、安心・安全を支える AI エージェントの実現が可能となります。

しかし、常時稼働型の AI エージェントの利用が拡大すると、マルチモーダルデータ※7 の量が爆発的に増加します。データの増加に対し、現行の ICT インフラでは次の 3 つの技術課題があります。

① 無線区間の帯域不足

多数のセンサや端末から膨大なデータが常時送信されると、無線区間において十分な通信帯域の確保が困難となります。

② センサデータの常時 AI 処理による計算負荷の増大

全てのセンサデータを計算量の大きい AI で常に処理すると、計算負荷が過大となり、リアルタイムな応答が困難となります。

③ AI の大規模化に伴う計算負荷・消費電力の増大

多様なセンサデータの解析には大規模 AI が必要となり、必要な計算能力・電力が増大します。

3. 技術のポイント

本研究では、これら 3 つの課題を解決するため、以下の技術アプローチを提案します。

① ストリーミングセマンティック通信技術(東京大学)

時間的意味連続性を通信制御に組み込むセマンティック通信の応用技術により、周囲の状況の変化を時間方向に追跡し、利用することで、単一フレームでは検出困難な周囲の状況も時間

的 文脈に基づいて検出可能となります。さらに、重要な情報を低データ量で送信することで無線区間における通信リソースを大幅に削減します。

② 生成 AI 向けメディア制御技術 (NEC)

AI エージェントの前段にデータ識別器を配置し、重要なセンサデータのみを選択的に AI エージェントへ入力することで、推論に必要な計算リソースを削減します。データ識別器は、AI エージェントの推論結果を学習し、重要なセンサデータを判別します。

③ In-Network Computing (INC) アーキテクチャ技術 (NTT)

1 つの巨大な AI にすべてを任せるのではなく、ネットワークに分散する小型の専門 AI や外部の情報源を組み合わせ、AI 処理の高効率・高信頼化を実現します。

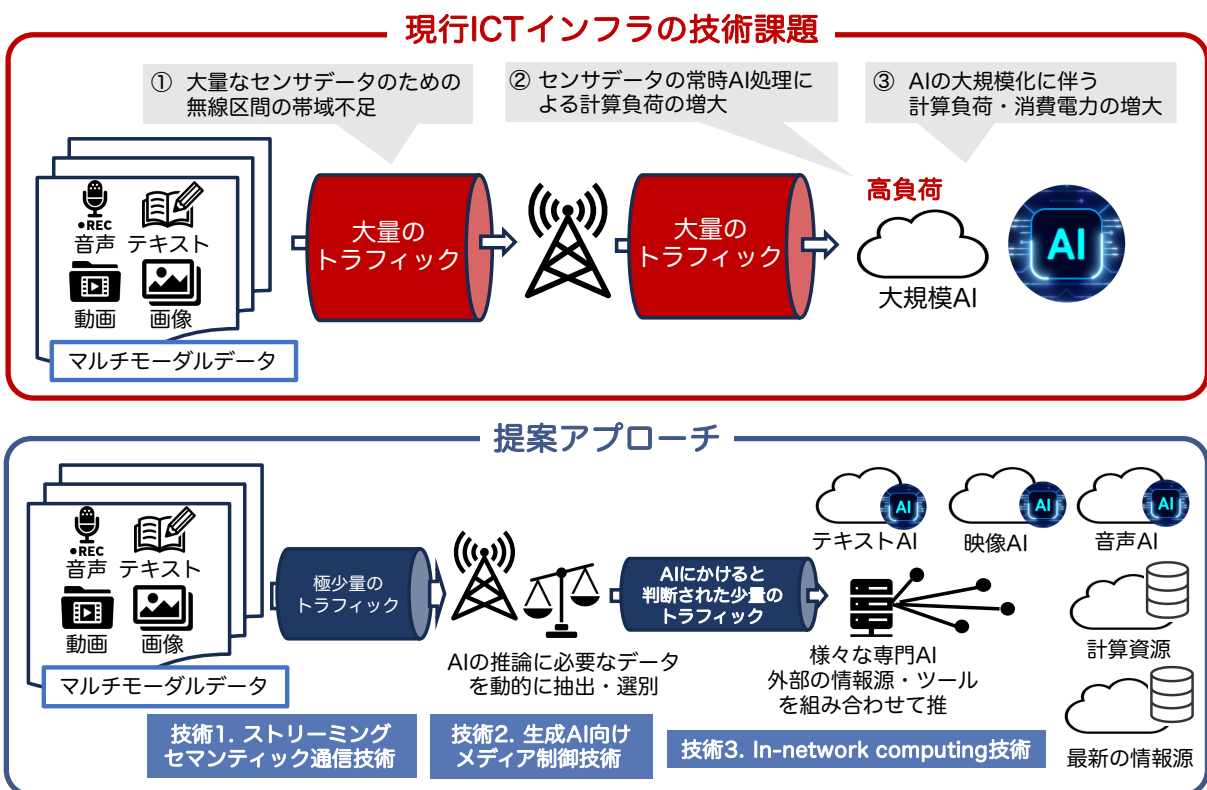


図 2. 技術課題と提案アプローチ

4. 実証の概要

本実証は、AI エージェント利用時におけるエンドツーエンド遅延※8 (以下、E2E 遅延) の特性を把握した上で、提案技術による通信及び計算処理の効率化が遅延低減に与える効果を検証することを目的として実施しました。実証では、危険シーンを含む動画データセット(60 秒、1,800 フレーム)を用い、センサから入力される映像を AI エージェントが処理する構成について、段階的な評価を行いました。対象とするユースケースは、AR グラスを装着したユーザの周辺環境を AI エージェントが継続的にモニタリングし、環境の変化やコンテキストを把握した上で潜在的なリスク兆候を予測・判断する、リアルタイム性が求められるシナリオです。

まず、事前評価として、センサから入力される全フレームを逐次的に AI 処理する構成における遅延特性を評価しました。その結果、フレームごとに処理待ち時間が累積し、E2E 遅延が時間の経過

とともに増大する傾向が確認されました。これは、ユーザの目の前で危険イベントが発生してから判断結果や指示が提示されるまでに要する時間が増大することを意味しており、リアルタイムな AR 支援のユースケースにおける課題であることが明らかになりました。

次に、事前評価の結果を踏まえ、本実証において提案技術を適用した構成の評価を行いました。その結果、通信量及び計算負荷が抑制され、動画全体を通じて E2E 遅延をほぼ一定に維持できることが確認されました。また、処理待ち時間が累積的に増大する傾向は認められませんでした。さらに、提案技術の適用による AI の推論精度の低下も確認されませんでした。以上より、提案技術は、リアルタイム性が要求される本ユースケースにおいて、AI の推論精度を維持したまま E2E 遅延を安定的に低減できることを示しました。



図 3.1. AR グラスをかけての実証の様子



図 3.2. AR グラスの画面

5. 各者の役割

① 東京大学:

ストリーミングセマンティック通信技術の研究開発を担当し、主に無線区間の帯域不足の技術課題解決に貢献

② NEC:

生成 AI 向けメディア制御技術の研究開発を担当し、主にセンサデータの常時 AI 処理による計算負荷の増大の技術課題解決に貢献

③ NTT:

In-network Computing (INC) 技術の研究開発を担当し、主に AI の大規模化に伴う計算負荷・消費電力の増大の技術課題解決に貢献

なお、各者技術の連携による相互作用により、各者が他の課題解決にも貢献しています。

6. 今後の展開

本取り組みは、MWC2026 Japan Pavilion への出展に採択されており、研究成果やコンセプトの国際発信を予定しています。今後は、研究成果の社会実装を視野に入れながら、「安心・安全」を支える AI エージェントと次世代 ICT インフラの実現に向けた研究開発を加速していきます。



※1. AI エージェント

AI エージェントとは、周囲の状況や入力データを基に目的を理解し、自律的に判断・行動しながらタスクを実行する AI 技術です。従来の人間からの指示（プロンプト）に基づく AI に加え、センサなどからの非人間起点の入力を用いて、継続的に環境を認識・判断する形態が注目されています。

※2. 6G

6G（第 6 世代移動通信システム）は、5G の次世代として研究が進められている移動通信技術であり、超高速・超低遅延・高信頼に加え、多数端末接続や AI 活用を前提とした通信基盤の実現が期待されています。

※3. IOWN (Innovative Optical and Wireless Network)

IOWN 構想とは、あらゆる情報を基に個と全体との最適化を図り、光を中心とした革新的技術を活用し、高速大容量通信ならびに膨大な計算リソースなどを提供可能な、端末を含むネットワーク・情報処理基盤の構想です。

※4. ストリーミングセマンティック通信技術

セマンティック通信は、データの「意味」や「重要度」に着目し、必要な情報のみを抽出・圧縮して送信する通信方式です。従来のビット単位の通信に比べ、通信量を大幅に削減しつつ、AI 処理に必要な情報を効率的に伝送できます。ストリーミングセマンティック通信技術は、セマンティック通信の応用技術で、周囲の状況（コンテキスト）変化を検知し意味の差分のみを伝送する事で、無線区間における通信リソースを大幅に削減します。

【関連文献】

小野翔多, 中尾彰宏, “時系列意味連続性を考慮したリアルタイム動画ストリーミング向けセマンティック通信方式”, 電子情報通信学会技術研究報告, March, 2026

※5. 生成 AI 向けメディア制御技術

生成 AI 向けメディア制御技術とは、センサや映像などの入力データを AI 処理の前段で分析し、重要度や必要性に応じて選別する技術です。これにより、AI 推論に必要な計算量を抑え、リアルタイム性の向上や計算資源の効率的利用が可能となります。

【関連文献】

K. Azuma, H. Itsumi and K. Nihei, “Similarity-based Frame Screening for Edge-Cloud In-Cabin Monitoring”, In Proc. of 2026 IEEE 23rd Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2026

※6. In-Network Computing (INC) アーキテクチャ技術

INC 技術とは、モバイルネットワークのコア内に AI 処理などの計算機能を統合し、重要な処理をネットワーク内部で完結させるアーキテクチャです。端末やネットワーク、クラウドなど、分散する計算リソースを有効に活用することで、低遅延かつ高信頼なサービス提供を実現します。

【関連文献】

林 健太郎, 平井志久, 松川達哉, 馬場宏基, “6G/IOWN 時代の高速なエンドエンド情報同期・連携技術「In-Network Service Acceleration Platform」”, NTT 技術ジャーナル, October, 2024

林 健太郎, 小野翔多, 武田知典, 中尾彰宏, “MCP 統合 LLM におけるネットワーク状態とコストを考慮した tool 選択システム”, 電子情報通信学会技術研究報告, December, 2025

Hiroki Baba, Kentaro Hayashi, Shiku Hirai and Tomonori Takeda, “Unified Control of Network and Compute Toward 6G In-Network Computing Service”, In Proc. of 2026 IEEE 23rd Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2026



※7. マルチモーダルデータ

マルチモーダルデータとは、映像、音声、センサ情報、位置情報など、異なる種類(モダリティ)のデータを組み合わせたデータのことです。

※8. エンドツーエンド遅延

エンドツーエンド遅延とは、データが送信元から受信先で処理・応答されるまでに要する全体の遅延時間を指します。通信遅延と計算処理遅延の双方を含み、リアルタイム性が求められる AI サービスにおいて重要な指標です。

■ 本件に関する報道機関からのお問い合わせ先

国立大学法人東京大学
大学院工学系研究科
広報室

NTT 株式会社
情報ネットワーク総合研究所
広報担当

日本電気株式会社
NEC コーポレートコミュニケーション統括部