



IOWN: Quantum Leap

NTT 執行役員
研究企画部門長

木下 真吾 Shingo Kinoshita

本記事は、2025年11月19～26日に開催された「NTT R&D FORUM 2025—IOWN: Quantum Leap」における、木下真吾NTT 研究企画部門長の基調講演を基に構成したもので、Quantum Leapについて紹介します。



NTT R&D FORUM 2025の概要

今回のNTT R&D FORUM 2025のテーマは「IOWN: Quantum Leap」です。「Quantum Leap」には2つの意味があり、1つは「量子力学的な飛躍」、もう1つは「ビジネスの劇的な進化」を表しています。技術面とビジネス面の両面で新たな飛躍を遂げたいとの思いを込めています。

では、なぜ今“Quantum”なのか。現在のAI（人工知能）時代において、必要となる計算量は急速に増大しています。図1は、横軸が年数、縦軸が計算リソースを示していますが、特にAIが本格化した領域から急激に傾きが変わり、膨大なリソースが必要になっていることが分かります。

一例としてGPT-4は、1回の学習に約4万MWhの消費電力が必要で、これは原発約40基分に相当するといわれています。こうした背景から、AIが今後さらに進化していく中で、今のように計算リソースを増やし続けるだけで良いのかという疑問が、世界共通の技術課題になっています。

この課題に対し、NTTは4つのアプローチで進めています（図2）。まず図2左にある2つは、古典的な情報処理の範囲での取り組みです。1つはAIの実行環境そのものを効率化する「IOWN」です。もう1つはAI自体を軽量化し効率化することで消費電力を抑える生成AI「tsuzumi」です。しかし、この2つだけでは不十分であり、これからは古典領域から量子領域へ処理の

範囲を広げていきます（図2右）。それが、計算性能を飛躍的に向上させる「光量子コンピュータ」と人間の脳に近い極めて高効率なAIをめざす「量子AI」です。NTT 研究所としては、この4つの取り組みを、古典と量子を交えながら進めていく方針です。

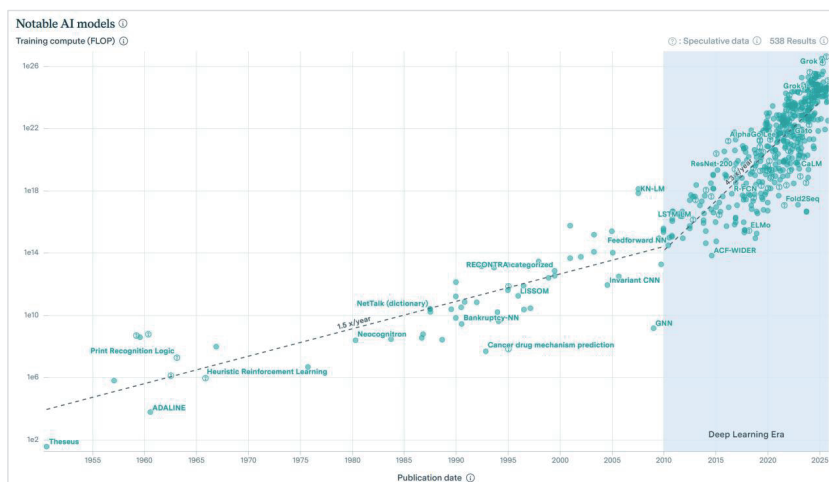
今回このアプローチを踏まえて、「Quantum」「IOWN」「Gen AI（生成AI）」について、NTTの研究成果とその意義を紹介します。

Quantum

■量子コンピュータ

最新のコンピュータやAI技術でも解決が難しい問題に対し、量子コンピュータの

AIの登場によって加速度的に計算機リソースが増加している



* Epoch AI, 'Data on AI Models'. Published online at epoch.ai. Retrieved from 'https://epoch.ai/data/ai-models' [online resource]. Accessed 23 Oct 2025.

図1 AI時代：膨大な計算機リソース

期待が高まっています。量子コンピュータの研究は近年急速に進展し、産業化の波が押し寄せています。市場予測では、2030年代には数兆円規模に成長し、期待はかつてないほど高まっています。量子コンピュータには、超電導、中性原子、光などさまざまな方式があります。

NTTは光通信技術から生まれた光量子方式に挑戦し、光の特性を活かすことで、高速・省電力で圧倒的なスケールビリティを実現していきます。通信との親和性も高く、未来のネットワークとつながる力を持っています。私たちは光の力で量子計算を次のステージへ導きます。それは、エネルギー効率を極限まで高めた持続可能な社会の新しい計算基盤です。産学連携による技術開発の加速と光量子の開発コミュニティの拡大を図り、2030年に汎用的に大規模計算が可能な光量子コンピュータの実現をめざします。将来はIOWNのコンピューティング基盤上で地球規模の量子コンピューティングを可能にします。

■光量子コンピュータの基本動作

光量子コンピュータの基本動作について説明します(図3)。最初に、量子光源によって量子光を生成します。続いて、量子どうしの関係性を保つために「もつれ」をつくります。その後、このもつれを網目状に広げていくことでクラスタを生成します。このクラスタ状態に対して、各量子へ操作を施すことがプログラミングに相当し、最後にそれらを測定することで計算を進めていきます。光量子コンピュータは、この一連の流れで計算を実行します。

光量子コンピュータの性能を向上させるためには、主に3つのポイントがあります。

まず1番目は量子光源の品質を高めることで、生成される量子光がより良質になります。2番目は「もつれ」の数を増やすことで、大規模な計算が可能になります。3番目は測定器のスピードを上げることで、高速な計算が実現できます。これら3点が現在の重要な技術課題となっています。

実際の光量子コンピュータのソフトウェ

アとハードウェアの構成について説明します。まず量子光源の部分で量子光を生成し、次に光学系装置の中でもつれを生成し、さらにそのもつれを拡大してクラスタを形成します。最後に、観測・制御装置でプログラミングと測定を行う、これが基本的な動作の流れです。NTTが特に取り組んでいるのは、量子光源と観測・制御装置の領域で、ここでは「光パラメトリック増幅器」を用いています(図4)。

光パラメトリック増幅器は、光通信分野で中継増幅や波長変換に使われていたデバイスですが、これを量子コンピュータに適用できないものかと挑戦してきました。その結果、まず光源としては非常に高品質な量子光の生成に成功しました。さらに測定についても、この中継に使われていた増幅の技術を応用することで、非常に高速な測定が可能になっています。光通信と光量子コンピュータは非常に親和性が高く、送信機の部分では光源の高品質化や波長変換の技術がそのまま活用できます。さらに、多重や光ファイバ自身の低損失化・マルチコアファイバ技術は、光量子コンピュータの光学系装置に適用できます。また、受信機で用いられる高速検出技術や光増幅技術も測定器に利用できます。

このように、光通信で長年培ってきた技術が、光量子コンピュータのさまざまな部分に活かされています。

■光量子コンピュータの特徴

量子コンピュータというと、超電導方式がよく知られていると思います。この方式で大部分を占めているのは冷凍・真空装置

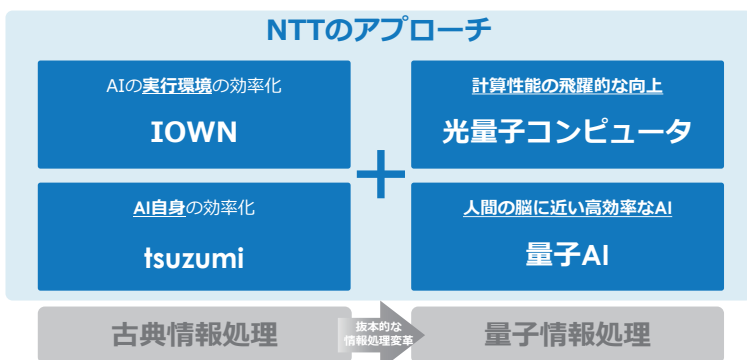


図2 NTTのアプローチ

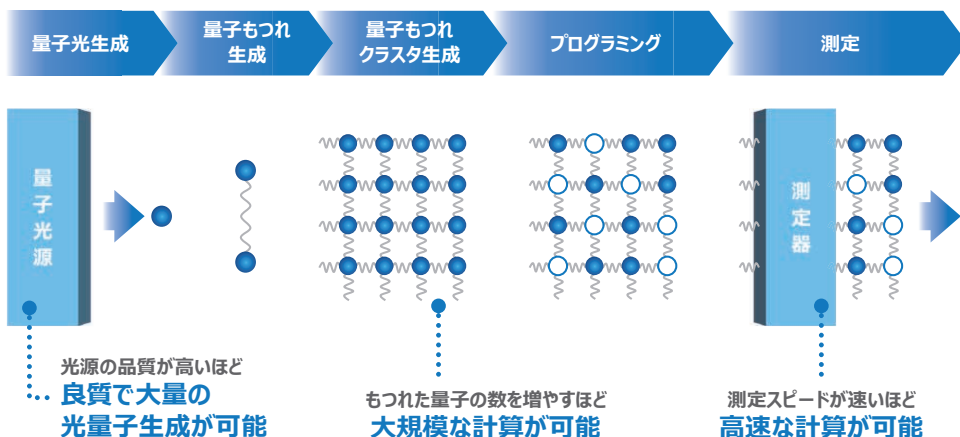


図3 光量子コンピュータの基本動作

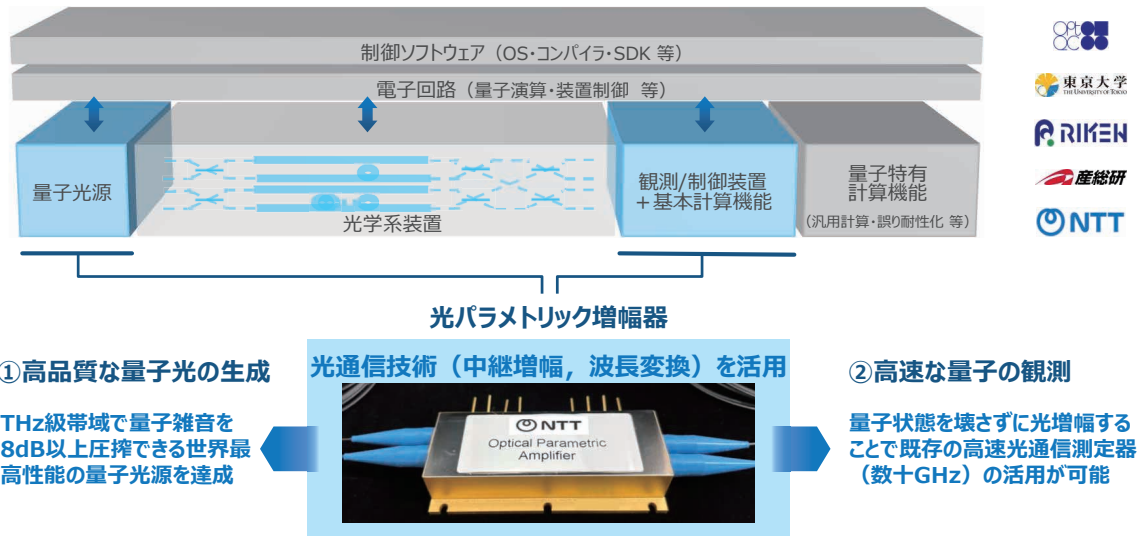


図4 光パラメトリック増幅器

で、その先端に小さな量子チップが載っています。この冷凍・真空装置は構造的に非常に大型で、技術的にも小型化が難しいとされています。

一方、光量子方式では主に光源やミラーなどの光学系装置を用いるため、巨大な冷却装置が不要で、常温常圧で動作します。装置全体を小型化しやすいという大きな利点があるのです。また、量子コンピュータの性能を高めるうえでは、空間多重・時間多重の観点から、いかに計算量や並列性を上げていくかが重要です。

他方式では主に空間的な並列化に頼り、チップ上で量子もつれ状態を並列化することで性能向上を図ります。しかし、前述のとおり冷凍・真空装置は小型化が難しく、並列化を進めれば進めるほど必要なスペースが膨大になっていきます。チップ自体は集積化できても、冷凍・真空装置は形状・構造の観点から縮小が困難であり、結果として装置全体を大規模化せざるを得ないと考えています。

また、光量子方式は特に時間多重で大きなメリットがあります。時間軸に沿って量子のもつれを連続的に重ね合わせていくことで、量子ビットを無限に増加させることができます。さらに、量子ビットを時間方向に生成する間隔を短くし、動作周波数を上げることで、多重度を増やすことが可能になります。現在は10 GHz程度の動作ですが、光源の高品質化や測定器の高速化によって、数10 GHzから数100 GHzまで高

めることができると考えられています。さらに光通信の特徴である波長多重を活用することで、単一の光ファイバの中に複数波長を同時に流すことができ、100 GHzから10 THzまで多重化を拡張できる可能性があります。

これらの特徴は非常に省スペースであること、さらに光の周波数で動作するため超高速であること、最小限の電気回路で済むため省電力であること、そして今まで開発してきた技術を活用することができるため低投資で、圧倒的なスケーラビリティが実現できるのではないかと世界的に大変注目されています。

これまでは1台の量子コンピュータの話でしたが、これを世界規模で複数台並べ、さらにその間をIOWN APNと全光量子中継技術を使うことにより、世界規模で分散した光量子コンピュータネットワークが将来的に実現できるのではないかと考えています。そして、NTTは東京大学発のベンチャー企業であるOptQCと世界最高の光量子コンピュータの実現に向け、連携協定を締結しました。2027年には1万量子ビット、2030年には100万量子ビットという規模で、スケーラブルで信頼性の高い光量子コンピュータの実現をめざしていきます。

IOWN

■IOWNのロードマップ

IOWNのロードマップについて、改め

て紹介します。まず、IOWN 1.0では、ネットワークの領域において、データセンタとデータセンタの間を完全に光化することをめざします。続いてIOWN 2.0以降ではコンピューティング領域に入ります。2.0では、サーバとサーバのボード間を完全光化します。3.0では、ボード上のチップとチップ間を光化します。そして4.0では、チップ内部のダイとダイの間を光化することをめざします(図5)。

■IOWN光コンピューティング

NTTは進化する未来のコミュニケーションを光による低消費電力で支えるため、ハードウェアとソフトウェアの両面からIOWN光コンピューティングにより解決をめざします。ハードウェア面においては、電気ではなく信号を光でつなぐ信号に置き換え、消費電力の削減と低遅延化による高速通信を実現します。ソフトウェア面においては、従来、高度な処理を必要とする大規模なコンピュータには、部品の組合せに制限があるため、筐体単位での増設が必要となり、使わない部品にも電力を消費していましたが、この制限を取り払い、必要な機能を必要な分だけ使えるリソースの配分技術により、処理の高度化と消費電力の削減を実現します。

大阪・関西万博の会場では、AIを用いた映像処理に、これらの技術を活用し、必要ときに必要なだけリソースを稼働させるなど、消費電力を最大8分の1まで削減しています。将来的には超高速ネットワーク

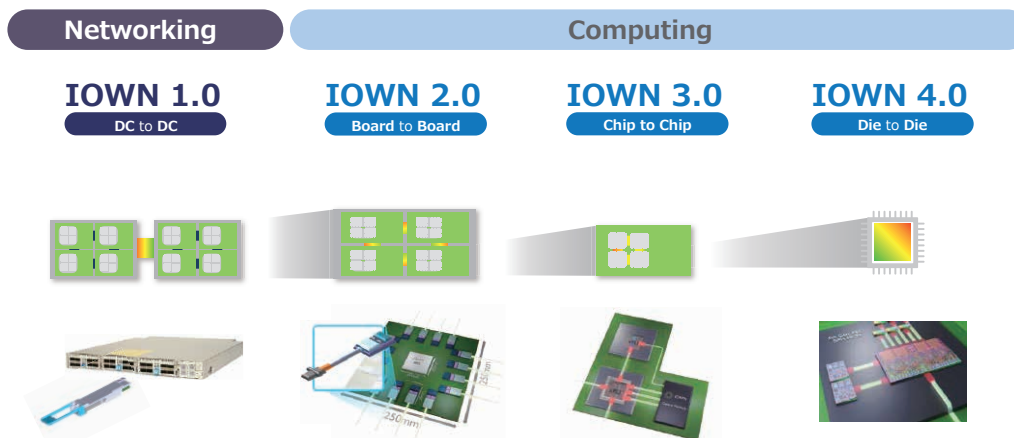


図5 IOWN ロードマップ

であるAPNを組み合わせることで、遠隔地のコンピュータをあたかも1つの高性能なコンピュータのように、利活用再生可能エネルギーや発電施設近傍での効率的な電力利用が可能となります。

伝送距離と伝送スピードについて、従来は100 kmを超えるような長距離通信において光ファイバが使用され、10 cmや1 mなどの非常に近い距離に関しては電気通信が使用されていました。しかし、近年ではただか10 cm程度の距離にもかかわらず14.4 T, 15 Tbit/sのスピードが必要になってきており、すでに電気では伝送できない状況にまできています。そして、この10 cm, 1 m以内の距離をいかに効率良く光で伝送するかが技術的な課題となっています。

■PEC-2スイッチにおける消費電力削減

私たちは効率の良い伝送技術の開発に挑戦し、「光電融合デバイス (PEC-2)」を試作しました (図6右)。従来の光通信スイッチは、トランシーバ装置の前面に差し込み、通信を行います。光は途中までしか届かず、装置内部に入った後は電気配線で信号が伝送されます (図6左)。この電気配線部分は約30 cmですが、例えば15 Tbit/sの通信を行う場合、わずか30 cmの距離でも大きな発熱が生じ、消費電力が増えるという課題があります。

一方、私たちが開発しているPEC-2スイッチでは、光がスイッチ用のASICの間際まで届くようになっており、これにより電気配線の距離を3 cmほどに短縮でき、消費電力を削減できます。このPEC-2スイッチは2026年度末に商用化することを先

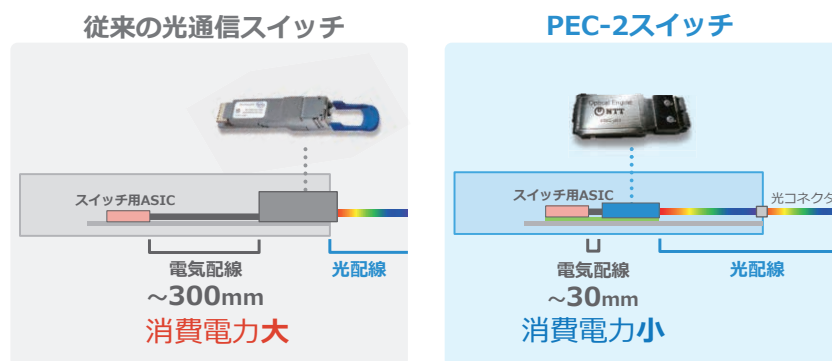


図6 PEC-2 スwitchにおける消費電力削減

日発表しました。製品化に向け、ASICはBroadcom、光エンジン、スイッチモジュールはNTTイノベティブデバイスが製造します。これを、Acctonがスイッチという装置に仕上げ、さらにGPUサーバと組み合わせることによって、光コンピューティングを実現していく流れとなります。

PEC-2スイッチが実現した暁には、GPU・クラスターの運用がどのように変わるかという点について説明します。NTTはソフトウェアの開発も行っています。前述のGPUクラスターはAccton, Drut, Fsasのコンポーザブルサーバ、あるいはDELL, Supermicroの汎用的なGPUサーバなどで構成することができます。これらのGPUを、光電融合スイッチで結合し、さらにDCIコントローラを用いることで、多数のGPUリソースを最適効率化することをめざしています。

例えば、現在東京・大阪・名古屋・福岡にそれぞれGPUが4つずつあり、各拠点のGPUの空き・故障の監視やトラフィックの監視、消費電力の監視を一元的に管理

できるようになっているとします。このとき、管理監視するだけではなく、ダイナミックな制御も可能です (図7)。水色・紫色・橙色の部分はすでにタスクが動作しているGPUで、灰色の部分が空いているGPUです。ここへ新たにAIの推論タスクが4つ来たとします。このタスクは4つともバラバラのGPUに配置しても構わないのですが、東京からのレスポンスタイムを優先してほしい、という要件があるとします。この場合、DCIコントローラは東京に2つ、大阪に1つ、名古屋に1つというかたちで配置をします (図7の赤色部分)。これはマニュアルではなく、GPUの状況を監視しながら自動で行う点が特徴です。

さらに、AIの学習タスクが来たとします。学習タスクは分散させると性能が低下するため、可能な限り集中させて配置する必要があります。これは、並んだGPU間の通信が大きく影響するためです。そこで例えば、福岡にある推論タスクを東京に移動させ、福岡の4つのGPUを空け、そこに新しい学習タスクを割り当てることで、全体

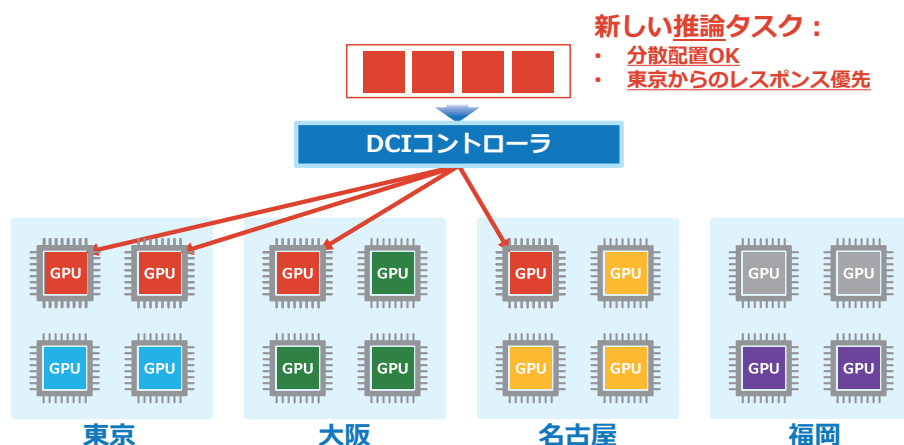
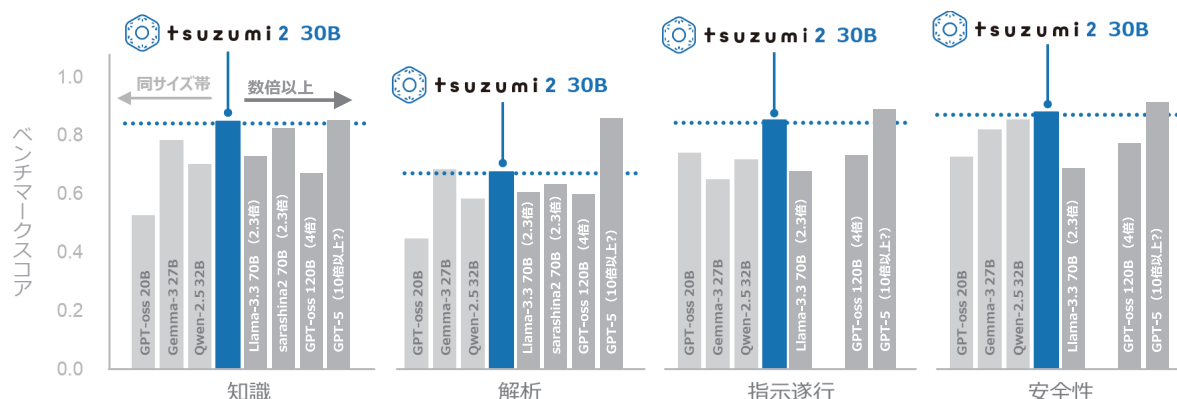


図7 IOWN光コンピューティング

- 同サイズ帯のモデル（Gemma-3等）では世界トップクラスの日本語性能
- 数倍以上大きなモデル（GPT-oss 120B, GPT-5等）にも引けを取らない性能でコストパフォーマンスに優れる



※評価条件

- 知識: llm-jp-eval (JAQKET (日本語QAタスク), NIIILC (質問応答), JCommonsenseQA (常識推論能力)), pfgen-bench (PFN日本知識ベンチ) の平均値
- 解析: llm-jp-eval (Wikipedia読み指定, Wikipedia寄り受け解析) の平均値
- 指示遂行: M-IFEval_ja
- 安全性: AnswerCarefully (x0.2)
- 知識と解析においては、事前学習モデルを使用。ただし、GPT-oss 20B/120B, GPT-5は事前学習モデルが公開されていないため事後学習モデル (Reasoningモデル) を使用
- 指示遂行と安全性において、全て事後学習モデルを使用。sarashina2 70Bは事後学習モデルが公開されていないため対象外

図8 日本語性能のさらなる向上

の最適化を自動的に行うことが、DCIコントローラの担う仕組みです。

GEN AI (生成AI)

■tsuzumi 2

2024年3月にtsuzumiを発表しました。このときのパラメータのサイズは7B (70億パラメータ) になります。そして今回2025年10月にtsuzumi 2を発表しました。こちらのパラメータサイズは30B (300億パラメータ) となっています。なぜ30Bなのかというと、現在最新のGPU 1台で動作する最大のサイズだからです。1 GPUが実現すると何が良いのかというと、オンプレミスやプライベートクラウドとして、非常に低

コストかつ高セキュアに運用できるからです。そのため、私たちはこのGPUのパラメータサイズにこだわって開発を進めています。

tsuzumi 2の特長は4つあります。そのうちの1番目が「日本語性能のさらなる向上」です。図8の横軸には「知識」「解析」「指示遂行」「安全性」と書かれていますが、これは代表的な日本語ベンチマークテストの種類を示しています。縦軸はスコアで、高いほど性能が良いことを意味します。中央のブルーの棒がtsuzumi 2の成績で、左側の3本の棒は、tsuzumiと同じくらいのサイズ帯の他社製品の成績です。左からOpenAIのGPT-oss 20B, GoogleのGemma-3 27B, AlibabaのQwen-2.5

32Bと並んでいますが、世界でもっとも優れているといわれているこの3つに対してもtsuzumi 2は日本語性能において優れているという結果が出ました。さらにパラメータサイズがもっと大きいものとも比較しています。右側には、Llama-3.3 70B, ソフトバンクのsarashina 2 70B, GPT-oss 120B, GPT-5と並んでいますが、GPT-oss 120Bはtsuzumi 2の4倍ほど大きなサイズとなっています。そしてGPT-5はおそらく10倍以上大きいだろうといわれています。tsuzumi 2は、これらと比較しても引けを取らない性能となっており、非常に効率的な日本語LLMなのではないかと考えています。

2番目の特長は「特化モデルの開発効率

の向上 (F.T) です。今回の tsuzumi は、汎用をねらうだけでなく、特化型も対象としています。金融、医療、自治体などの大規模データを幅広く学習させています。その結果として、金融分野における性能評価も行っています。具体的には、金融のベンチマークとして「ファイナンシャル・プランニング技能検定 2 級テスト (FP2)」を用い、200問、500問、1900問と段階的に学習させた場合に、成績がどのように向上するかを評価しました。Google の Gemma においては初期状態で39点を獲得し、1900問の予習を実施することによって64点というFP2学科試験の合格ラインを達成しました。一方 tsuzumi では、初期の段階で60点には及ばないものの54点を獲得しました。200問の予習をした段階で70点をクリアし、合格圏内に入りました。そういう意味では約10倍の学習量を削減しながらも合格ラインを達成したという、非常に効率的な結果が出ました。

3 番目の特長は「低コスト・高セキュアの維持」です。例えば、DeepSeek 700B を実際に動かす場合、NVIDIA の H100 GPU を16基ほど使用する必要があり、ハードウェアの単純コストだけでも約1億円かかるといわれています。一方 tsuzumi では、必要なハードウェアコストは約500万円で済みます。3 世代前の GPU でも動作可能であり、この結果は、従来比で10～20倍のコスト優位性を実現しています。

最後にとても重要なのが「国産 AI (ソブリン AI)」であるということです。国産の AI として、いかに AI をキープしていくか

という観点においてもメリットがあります。まず、①言語・文化をいかに言語モデルとして守っていくかという主権、そして②学習データとして著作権違反をしたデータを学んでいないかという、学習データを自分たちで決められるという主権、③開発のスケジュールや開発手法、これ自身を決めることができるという主権、さらには④ライセンスの主権です。現在、オープンソースの LLM が多く存在しますが、いつ課金が始まるか、あるいはいつ使用できなくなるかは予測が難しいため、こうしたライセンスの管理権も重要です。最後に、⑤技術の主権についてです。AI は今後、半導体やスマートフォン OS と比較しても遜色ないほど、IT 基盤として重要な存在になると考えられます。

こうした状況に対して、日本が技術的に主導できる立場を維持し、新しい技術を開発できる、あるいは少なくとも理解できる体制を持つことが重要です。そのため、NTT は技術的な主権を維持する観点から、自ら LLM を開発しています。

■生成 AI 応用

従来の AI 音声は「トランシーバ型」といい、人間が話し終わった後に AI が話し始めるため、間があいたり、言葉がかぶってしまったときにテンポが悪くなったりしていました。また、口調も固いことから不自然さがありました。しかしこの課題に対して「Full-Duplex 型 (全二重型)」という技術を開発しました (図 9)。

これは、相手の言葉を音声認識でテキスト化し、言語モデルが学習して、その回答

を音声合成するやり方ではなく、一気に話している言葉を学習データとして入力し、そのときの対話の状況を教師データとして学習します。この手法は Speech to Speech ともいいます。この手法を使うことで、会話が重なりながらもテンポよく、相手の言葉とかぶったり遮ったりせずに自然に話ができます。実はこの技術の開発は約半年間で行いました。開発当初は日本語を話してはいるけれども違和感のある音声になってしまうようなところからスタートしました。また、Speech to Speech は制御も難しく、失敗の事例も多く存在します。例えば、同じ質問を 2 度行ってしまったり、質問の間に「失礼しました」といった言葉が入らなかったり、カード番号が16桁必要な場面で、15桁しか言われていなくてもそのまま処理を進めてしまうような失敗もあります。これは会話を会話のパターンとして学習しているので、正確なところが制御できていないという点では技術的な課題となっています。さらに、コールセンタでは、非常に怒ったお客さまからの電話がかかってくることもあります。その際、どのように対応するかが重要です。プロのオペレータは、怒り口調に対してトーンを落として対応しますが、AI がどこまでこのような対応を正確に再現できるかは、まだ多くの研究課題が残されています。

また、ネットワークへの応用として、ネットワーク障害復旧エージェントを紹介します (図10)。ネットワーク装置のどこかが故障しているという状況において、検査、診断を行い実際に復旧し、最後には復旧レ

人どうしのような自然な対話。コールセンタやパーソナル AI サービスへの適用をめざす

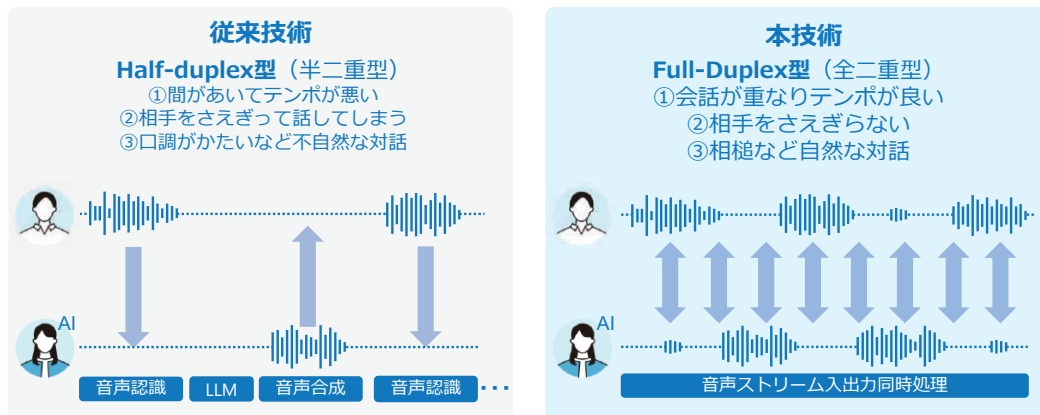


図 9 Full-Duplex 音声対話 AI

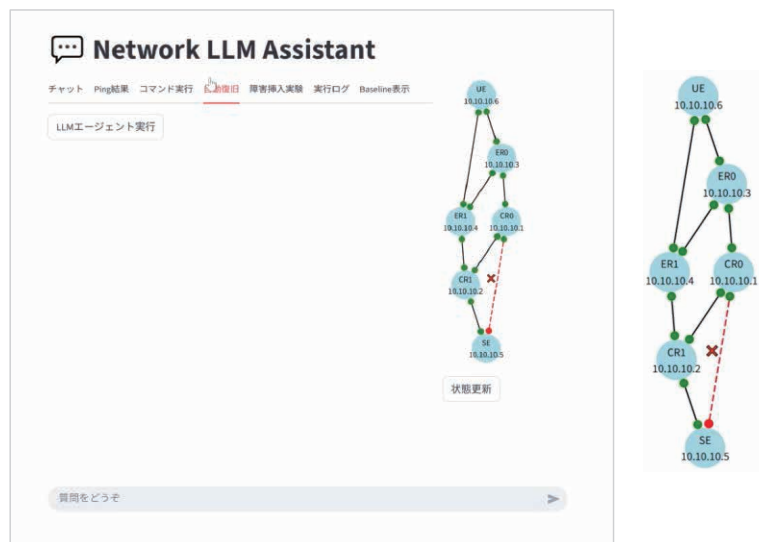


図10 ネットワーク障害復旧エージェント

①異常検知

1. 全ルータ間の疎通確認

②要因推定

1. 全ルータのIF確認
2. 全ルータのルート確認
3. トポロジ情報の確認
4. 過去の故障履歴との比較

③復旧

1. 復旧方針の立案
2. 原因IFのshutdown
3. OSPF経路再計算
4. 全ルータ間の疎通確認
5. 障害復旧レポート作成

ポートを上げるところまで、すべてAIが自動化してくれるものです。①異常検知、②要因推定、③復旧といった流れでAIエージェントが連携しながらやり遂げます。

■言語以外の大規模モデル

私たちは大規模行動モデルを開発しています。言語モデルは、ある単語が入力された後に次の単語が出現する確率を予測するモデルですが、行動モデルも同様の考え方で、ある行動の後に別の行動が起こる確率を予測する行動確率モデルです。このモデルをマーケティングや販促に応用しています。

例えば、過去にニュースを閲覧し、広告を見て、ポイントを付与したユーザに対して、①何もなかった場合、②DMを送った場合、③電話をかけた場合に何が起こるかをすべて確率として算出します。何もなかった場合に、料金プランを変更する確率は7%、スマホを購入する確率は1%、動画のオプションを追加する確率は4%、といったかたちで予測できます。この結果から、もっとも効果的な施策が電話であることが分かりました。そして実際、NTTドコモが電話にて販促を行ったところ、受注率は2倍に向上したという実験結果が出ました。

もう1つ、ワールドモデルという視覚、あるいはビジョンから飛び出て世界を認識するというAI（フィジカルAI）が流行しています。

現在、トヨタと一緒にモビリティAIの

開発を行っています。自動運転は車の自立した運転から、人、あるいはインフラと協調した三位一体性が重要であるといわれています。NTTでは人の行動をいかに予測できるかという研究をしています。例えば、横断歩道を渡ろうとしている人がいる場合を考えます。1人は車が来たことに気付き立ち止まることができましたが、もう1人は不注意でそのまま渡ってしまい、ひかれかけるといふ事象が起きたとします。この2人の行動の違いを、AIが予測できるかが課題となっています。従来はNaïve Modelといって物理的運動が継続したことによって起こる結果を予測しますが、今回の例題では予測に失敗します。一步進んだ社会行動モデルになると、目的地からの引力と障害、例えば赤信号や車からの斥力が釣り合うと停止するという予測モデルがあります。しかしこれも予測に失敗してしまいます。

NTTでは、社会行動モデルと主観空間モデルを組み合わせ活用しています（図11）。具体的には、手前の人は前を向いているため信号が見えていますが、後ろの人は視線の向きから信号の方向を向いておらず、上方しか見えていない、という差を空間認知モデルの違いとして利用することで、行動の予測を行うことに成功しています。人を単なる物体として扱うのではなく、その人の行動や社会的状況を考慮して行動を予測しています。

■AGI/ASIに向けた研究

1番目がマインド・キャプショニングです。これは映像を見たり想起した際の脳波を測ることによって、その内容を言語化させる技術です。動画を見ている状態で、MRIが脳波を計測し、分析をしています。分析をしながら推測を高めていき、徐々に正解なデータに近づいていきます。

「脳からの生成文と人手による正解文を用いた動画同定成績」のグラフから興味深い結果が得られました。言語野を含む全脳と言語野を含まない全脳とで比較した際、高い精度で知覚内容のテキスト化に成功し、言語野を除外しても大きな低下がみられませんでした。これは、言語情報ではなく非言語的意味情報ととらえて言語化していることを示唆しています。これは言語野を損傷している方や、まだ言語の獲得ができていない乳幼児、あるいは動物までも脳波を測ると何を見ているか言語化できる可能性があるということです。

2番目がマシンアンラーニングです（図12）。この研究は一度覚えたデータをどのように忘却させるかという研究です。どのようなときに役立つかというと、著作権違反のデータを学習した際に、その違反した部分だけを忘れることができます。通常、一度学習してしまったものを忘れようとする、学習データをすべてリセットした状態にし、最初からすべて学習し直すという作業を行わなければなりません。マシンアンラーニングは、出来上がった知識に対

人の主観的空間を考慮したSocial Force Model（社会行動+主観空間モデル）

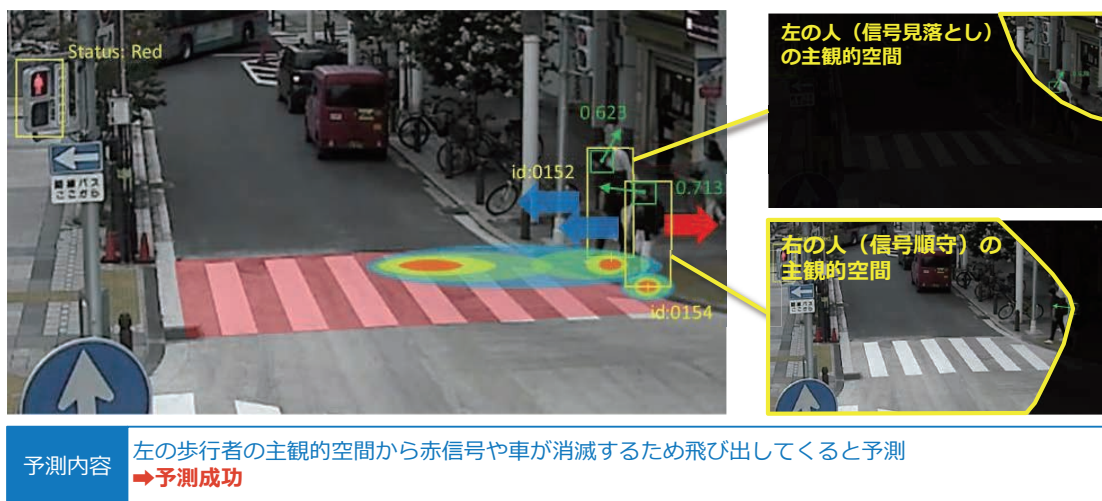


図11 ワールドモデル

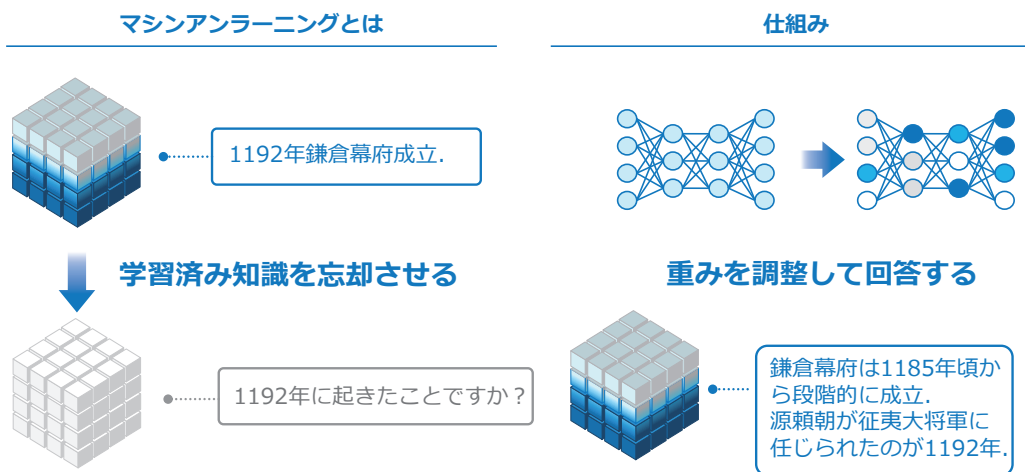


図12 マシンアンラーニング

して1日の知識だけを効率良く削除していくことができます。これが実現されれば著作権問題などの対応が容易になります。

3番目がLLMの心理的メカニズムです。従来のLLMはハルシネーションが多い、嘘をつくなどコントロールできないといわれてきました。しかし、研究を重ねるうちにLLMの中に例えば、嘘、優しさ、忸度などに強く反応するニューロンの「組」が存在することが分かってきました。これらを活用することにより、LLMの嘘発見器をつくるのが可能となったのです。嘘を指摘すると嘘ニューロンの活性度が下がります。LLMの嘘発見器をみると、このAIが嘘をついているのかついていないのか、ニューロンを見れば分かるようになります。これ

らを活用すると嘘をつかないように調整できる可能性があり、またさまざまなハルシネーションに応用できるのではないかとわれています。

最後に量子AIです。現在のAIは、莫大なエネルギーを使い稼働しています。そこで注目されるのが量子AIです。光の重ね合わせと干渉を活かし、より少ないエネルギーでより深い学習を可能にする研究が進んでいます。そしてその進化は一方向ではありません。量子がAIを高め、AIが量子を磨きます。知が響き合い加速する時代が始まっています。この進化により、量子AIは世界そのもののパターンや見えない関係性をとらえようとしています。量子は現在の古典のデジタルと比較してもアナログ計

算機に近く、ノイズが多いといわれています。しかし、私たち人間の脳もノイズが多く、言語データにもノイズが多いといわれている中で、ノイズをうまく活用することによって効率的なものができないかという研究です。

この量子AIを含め、NTT研究所としては、冒頭で最初に述べましたIOWN, tsuzumi, 光量子コンピュータ、量子AIを含むアプローチで、古典と量子の情報処理を交えながら、研究開発を進めてまいります。