2 次世代 AI 研究プロジェクト

次世代 AI コンピューティングに向けて

CD 研の次世代 AI 研究プロジェクト（以下、C 次 P）では、既存の機械学習の枠組みを刷新し、自ら成長して社会に適応する AI の実現により、既存技術では解決が困難で、多様な価値観を反映した様々な実社会の課題の解決に貢献する研究を行っている。C 次 P で目指す次世代 AI コンピューティングの世界観と開発した次世代メディア処理 AI 技術「MediaGnosis™」を説明する。

次世代 AI コンピューティングの世界観

2010 年代よりディープラーニング (深層学習) 技術が普及しだし、我々の社会は本格的に AI 機能を獲得したといえる。現在、データとデジタル技術を活用して既存の仕組みを変革することで新しい価値を創造するデジタルトランスフォーメーション（DX）が様々な分野で成果をあげている。C 次 P では、次世代 AI 研究として、現状の AI を活用・応用して、さらなる DX を実現する「AI の社会適応」を推進するとともに、その限界を突破する「AI 機能の革新」に取り組んでいる。その目指す先に IOWN の一要素として次世代 AI コンピューティングという世界観を想定している（図 1）。次世代 AI コンピューティングでは、都市・街・家などのネットワーク化された世界がセンタ AI とエッジ AI により統合され、あらゆる場所で人々を見守りつつ、社会の Well-Being 拡大に貢献する（図 2）。

従来型の AI 機能は、それぞれ独立に技術開発された上で構築されており、人間に例えるならば、「それぞれの知覚機能ごとに 1 つずつ脳が準備されている」ことに相当して動作している。現状では下記の課題がある。

①学習に大量の教師データや計算資源が必要である。

②学習範囲外の状況に弱く、実世界で起こる想定外の状況への臨機応変な対応ができない。

③パターン処理には強いが、意味理解や説明などの高次処理はできていない。

一方、次世代 AI コンピューティングを実現するためには、社会に深く浸透できる AI が求められる。「AI 機能の革新」では、これらの①～③の課題を解決した汎用的でかつ軽量な AI の開発が重要である。人間は 1 つの脳に五感で得られた知識を蓄積することができ、その知識を共通的に参照しながら様々な情報を処理することが可能である。人間はこのような情報処理機構を持つからこそ、効率的に知識を蓄えて成長することができ、その知識を活用して世の中の様々な情報を正確に理解することができる。

NTT コンピュータ＆データサイエンス研究所
次世代 AI 研究プロジェクト
プロジェクト マネージャ　**草地 良規**氏

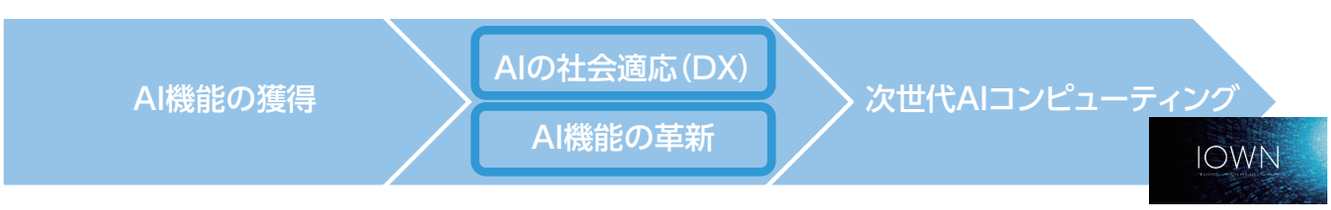

図 1　次世代 AI の位置づけ

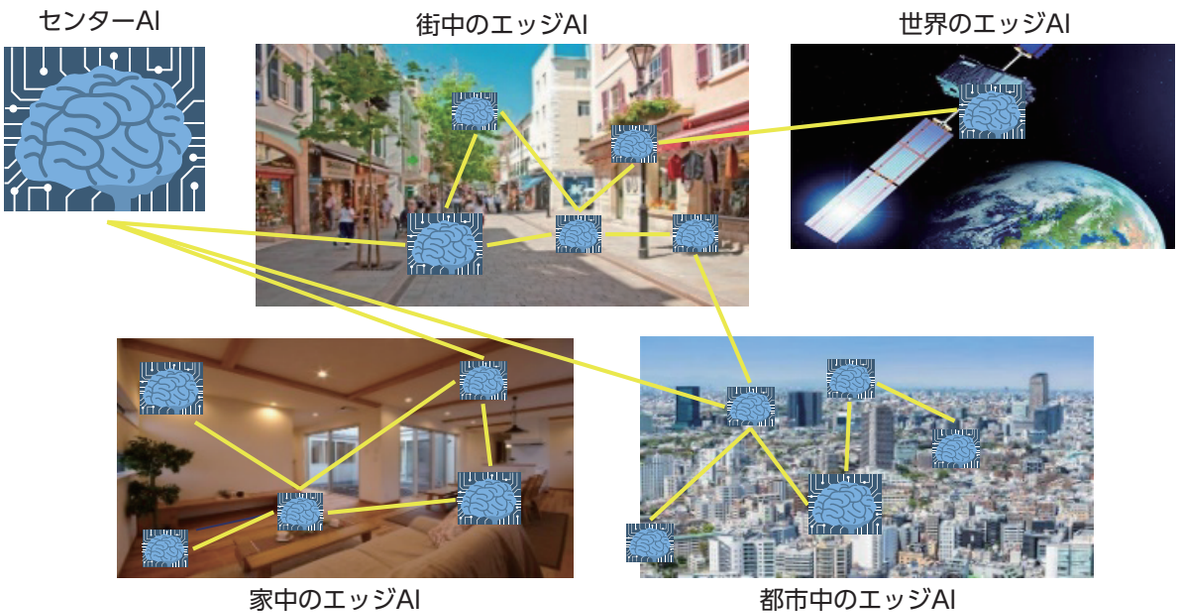

図２　次世代 AI コンピューティングの概要

C 次 P では、AI にも人間と同様の情報処理機構を持たせることで、人間のような効率的な「学習」やマルチモーダルの同時処理による高精度かつ総合的な「推論」が可能になると期待している。汎用的なメディア処理 AI の初期バージョンとして「MediaGnosis™」（第 1 版）の研究開発を進めてきた。

「MediaGnosis」の技術概要

「MediaGnosis」の名称は、「あらゆる Media（情報の記録）を人間のように統合的に Gnosis（知識）として扱い、それをもとに Diagnosis（判断）する」ということに由来している。

「MediaGnosis」（第 1 版）では、音声音響処理・画像映像処理・自然言語処理といったマルチメディアの情報処理を統合的にオールインワンで扱いこれまでよりも効率的な「学習」と、高精度かつ総合的な「推論」を実現している。

「MediaGnosis」の特長は、「知識統合型アーキテクチャ」により、人間のように 1 つの脳（AI においてはモデル）の中で、各種メディア処理 AI 機能を統合的にサポートするモデリングを行うことができる点である。

具体的には、「知識統合型アーキテクチャ」は入力モーダルおよび出力対象（数値・ベクトル・ラベル・テキストなど）の両者に対して、複数の機能間で同様の役割を持つ情報処理機構（理解部）を複数の機能間で共通化した（図 3）。これにより、1 つのモデルの中に様々な AI 処理機能を統合することが可能となった。

なお「Media Gnosis」には各種メディア処理 AI 機能に関する NTT の最新の研究成果が内包されている。本技術により、①これまでよりも少ないデータ量で人のように効率的な知識獲得が可能な「学習」が実現できる。例えば、音声認識の学習を行うことで、画像認識や機械翻訳の性能改善につなげることが可能とな

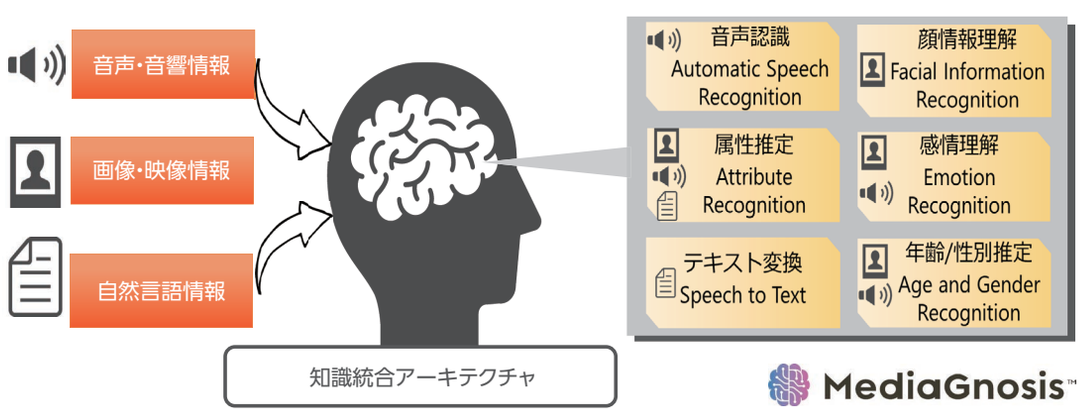

図３　MediaGnosis の概要

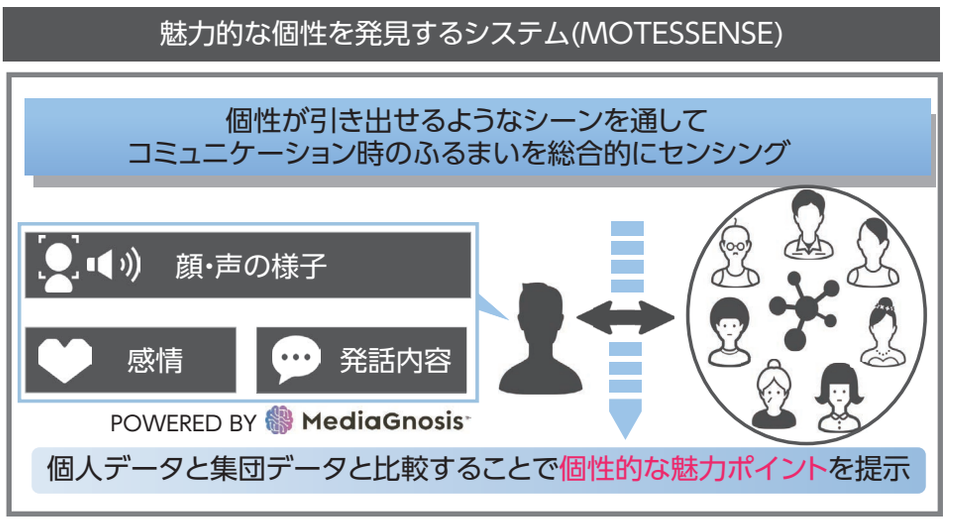

図4　MOTESSENSE の概要

る。②複数の機能を同時に駆動した「推論」が実現できる。例えば、声質、その言葉の内容、表情を同時に考慮して人の感情を推論することが可能となる。

ユースケースの例として、リモート会議の円滑化を狙うシステムが考えられる。相互の感情や発言量を分析することで会議を円滑に進めることをサポートできる。

「魅力的な個性」の発見を助ける「MOTESSENSE™」を開発

「多様性」の時代に個人の Well-Being を形成することを目指して、誰もが持っている「魅力的な個性」の発見を助けるアプリケーション「MOTESSENSE」を開発した（図4）。

「MOTESSENSE」では、他人と比較して個性的な部分を個人の魅力であると定義し、シチュエーションに応じたロールプレイの様子をカメラとマイクからセンシングすることにより、話し方、表情や仕草、言葉の内容などを総合的に AI で診断することで「モテ因子」（「魅力的な個性」に関する因子）を見つけ出す。初期バージョンでは、5 種類の観点（表現力・思考・味・雰囲気・アクティビティ）に対して独自に「モテ因子」を定義した。

「多様性」の尊重は、互いを受け入れ、個性や価値観の違いを受け入れるだけでなく、それらを融合させ、高め合うことで、社会を前進させるための鍵になるとも考えられている。

「MediaGnosis」により、多様な情報を人のように 1 つの脳で統合的に理解することが実現され、精緻な診断を可能としている。本アプリケーションを利用することで、利用者は自身の「魅力」を客観的に見つめ直すことができ、様々な観点で平均的な人との違いや個性的な点を知ることができる。なお、本アプリケーションによる診断は、他人の嗜好に依存した処理は含まれないとともに、良し悪しを診断するものではない。

IOWN 構想の実現による 2030 年の世界と次世代 AI コンピューティングの今後の課題

IOWN 構想が実現される 2030 年ころは、自分だけのパーソナル AI を個人の端末に保持し、開示情報を制御しながら、都市に遍在する AI エッジとパーソナル AI との連携技術が確立され、人の気持ちと意図が自然に周囲に伝わり、高度な情報サービスを意識せずに受けられる世界を実現している。

「MediaGnosis」の一部の機能（知識集約型アーキテクチャ上でモーダル独立に処理する機能）は、実用化に向けた検討が進んでおり、NTT Resonant より良質なコミュニケーションを API 連携によって支援する「AI suite」、NTT-TX より音声データ入力からテキスト出力までをオールインワンで実現する音声認識ソフトウェア「SpeechRec Server」として商用展開されている。その他の機能（複数モーダルを同時に処理する機能）については、実用に向けたフィージビリティの検証を現在進めており、2022 年度中に商用展開予定である。本技術を展開することで、より幅広い場面で AI による DX を推進するとともに、NTT が掲げる IOWN 構想で中核的な役割を担うデジタルツインコンピューティングの実現を加速する。さらに、汎用 AI の実現に向けて、より「人間に近い情報処理機構」を追求することで AI 技術の革新を推進したい。具体的には、より軽量な AI 技術、より少量データでの学習技術、人間に必須の記憶や知識のハンドリング技術の実現を行うとともに、様々な場所に AI 機能が存在・自律連動する AI ネットワーク（センタ AI、エッジ AI）の構築を実現したい。