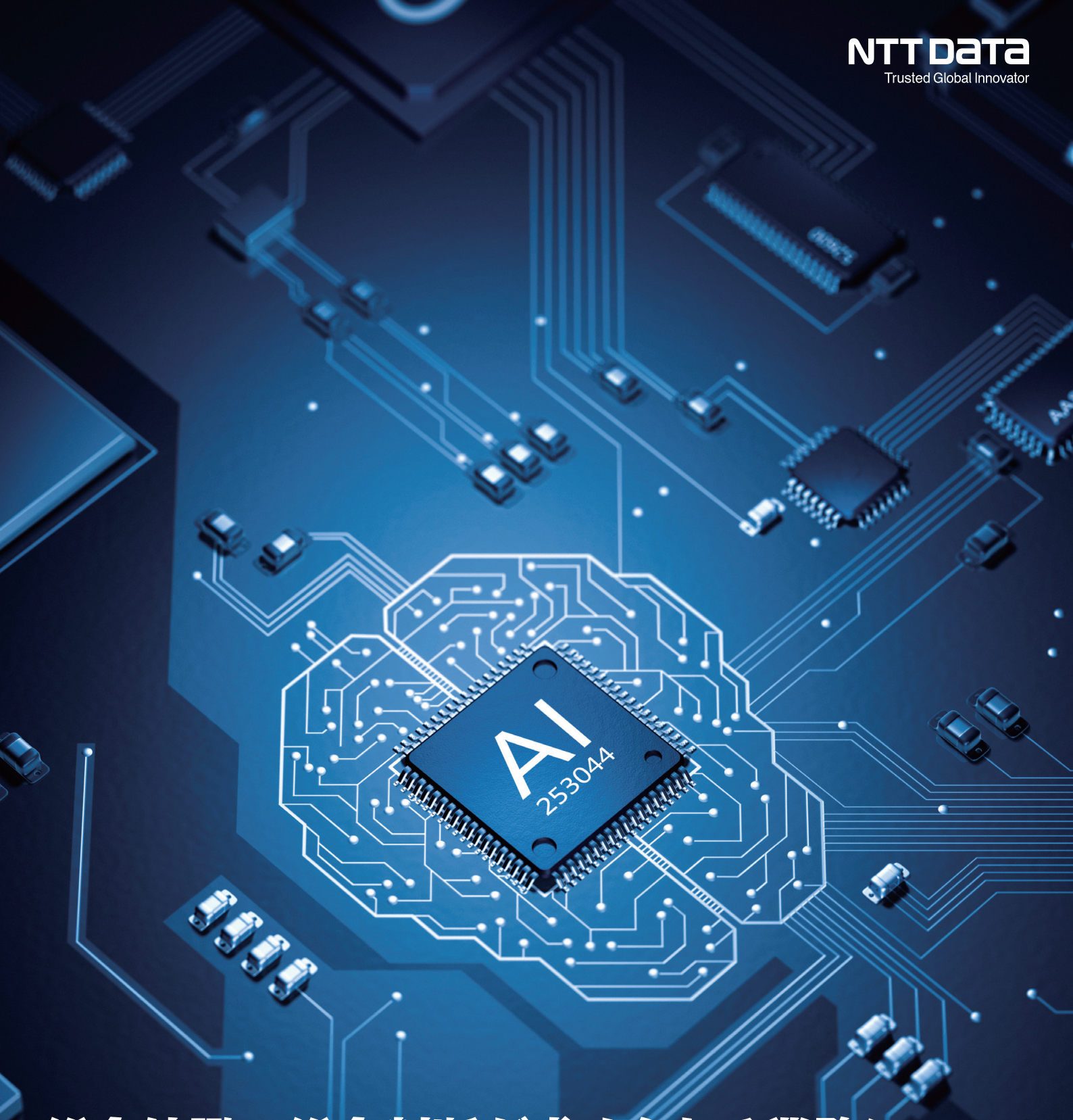
NTT DATA
Trusted Global Innovator



総合俯瞰、総合判断が求められる業務にAIを適用するためのキーテクノロジー

～複数のインプット情報形態を扱う「マルチモーダルAI」～

#マルチモーダルAI　#人的判断　#柔軟なオペレーション　#感覚を代替するAI　#拡張知能

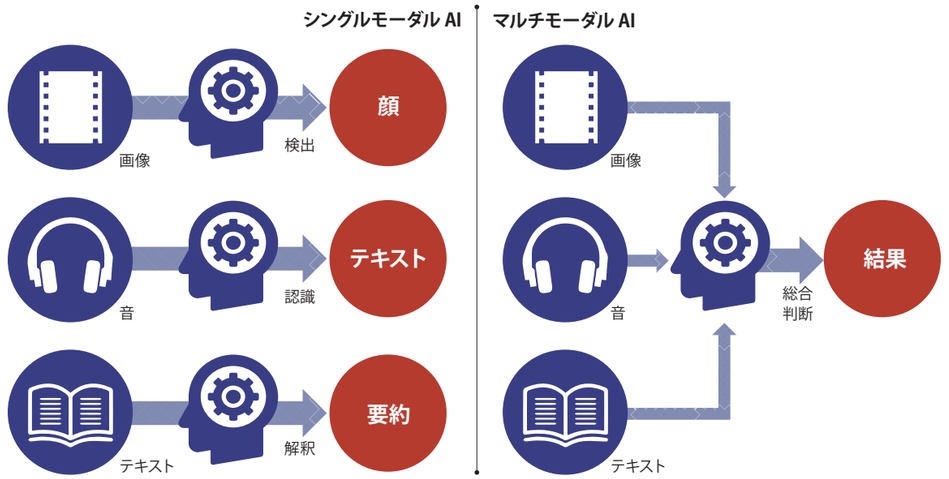
C O N T E N T S

CHAPTER.1
はじめに

オフィス業務は AI により効率化・自動化が進んでいます。特にルールが固定された基準審査などのオペレーションは判断ロジックも単純で、扱う情報形態も文字や数値などに限定すると、AI を導入しやすい傾向にあります。一方で、類似審査や異常審査のようなシステムオペレーションの合間にあり、多様な情報から総合的な俯瞰と判断を行う業務は、柔軟なオペレーションが求められるため、AI の適用が難しい状況が続いています。これは、人間が複数の感覚を使って取得した情報を俯瞰・統合して判断する行動であり、画像から物体を検出する AI、音声を認識する AI、テキストを解釈する AI といった、単純な AI では実現できないことを意味します。

そこで、人間の感覚に近い判断を AI で扱えるようインプット情報形態の種類を増やすことで、総合的な俯瞰と判断が必要な業務の AI 化を目指す新たなアプローチが、「マルチモーダル AI」です。

Figure.1 シングルモーダル AI とマルチモーダル AI の違い

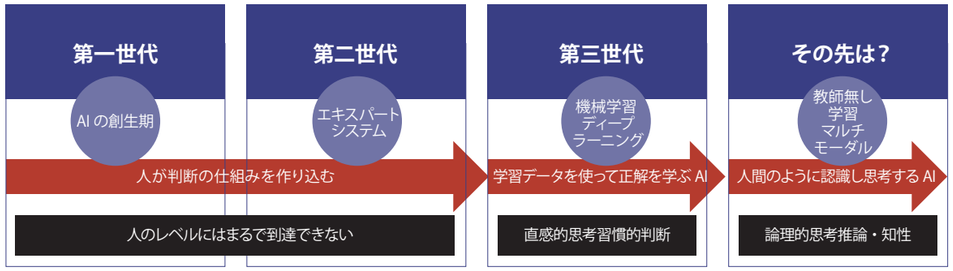


CHAPTER.2
次世代の AI は人間のように
認識し思考する方向へ

AI は世代を重ね進化を続けてきました。初期の AI は判断のパターンを作り込むことから始まり、決まった判定ロジックに基づく対応を実施するものでした。その後ディープラーニング技術の登場により、大量のインプットデータを元に AI が自ら学習し、直感的・習慣的に正解となる判断を実施できるようになりました。これにより、特定の画像を検出するなど単純な業務の自動化において、人間には難しい膨大な分量を AI で処理する使い道が主流となりました。そして、この次に来る AI の進化は、

決められたルールを学び判定するだけでなく、人間のように認識し思考する柔軟な AI だと言われています。例えば、「元気です」と口では言っていても顔色から具合が悪いのでは?と察するようなことが、AI にも可能になるということです。こうした論理的思考を実現するには、人間の五感に相当するような複数のインプット情報形態を扱い、人間の知性に近い学習方法を用いる必要があります。「マルチモーダル AI」や「教師無し学習」はこの部分に斬り込む技術であり、AI の次なる進化・知的活動への進出を支える有望な要素となっています。

Figure.2 AI の進化



CHAPTER.3
マルチモーダル AI の実現方法

マルチモーダル AI は 2 段階のステップで構成されます。

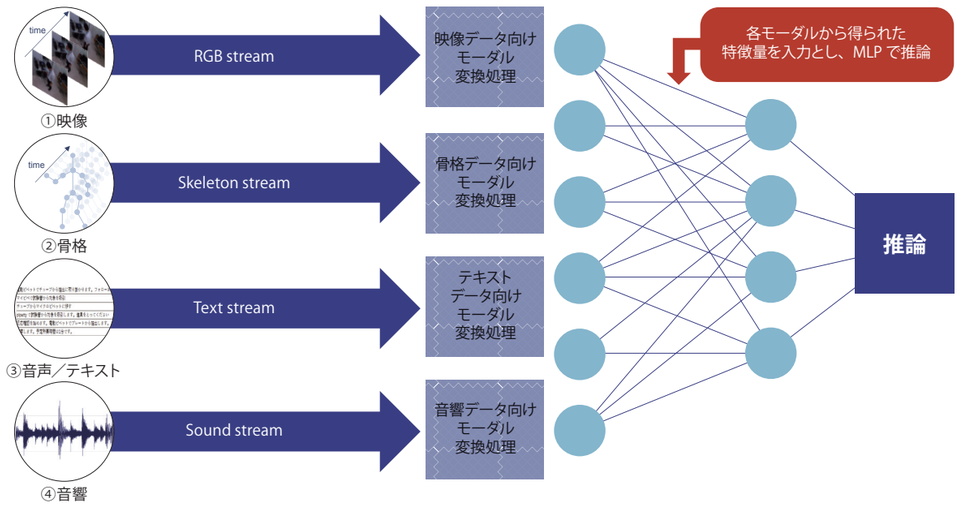
1．インプットを取得しモーダルに変換する手段

まずはセンサを用いて複数種類のインプットを獲得できる環境が必要です。例えば「職員が不正な行動をしていないか監視するケース」を想定した場合、カメラ映像／音／振動／匂いなどに相当するセンサデータを取得します。各センサデータは形式が多種多様であるため、特徴を分析可能な情報形態（モーダル）に変換します。

2．複数のモーダルを統合して扱う仕組みの導入

次に、各モーダルの隠れた関係性を学習し、特定の業務に利用するための仕組みを用意します。NTT データでは、複数種類のインプット情報を同一の枠組みの中で扱い、統合的に処理できるデータマネジメント方式を開発しました。これにより、例えば類似の物を探す場合に、テキストと画像を統合して扱うことで、従来よりも高精度の絞り込み検索が可能となります。

Figure.3 マルチモーダル AI の実現方法



CHAPTER.4
「マルチモーダル AI」PoC 事例

NTT データは、大規模な社会インフラシステムの開発を多数実施してきました。その中で様々なお客様の業務に深く入り込み、システム化における課題やノウハウに触れてきた経験を通じて、業務に対してどのモーダルをどのように使えば良いか、幅広い知見を有しています。既に具体的なユースケースでお客様との PoC を始めています。

Figure.4 実験行動を判別し、記録ノートを自動生成

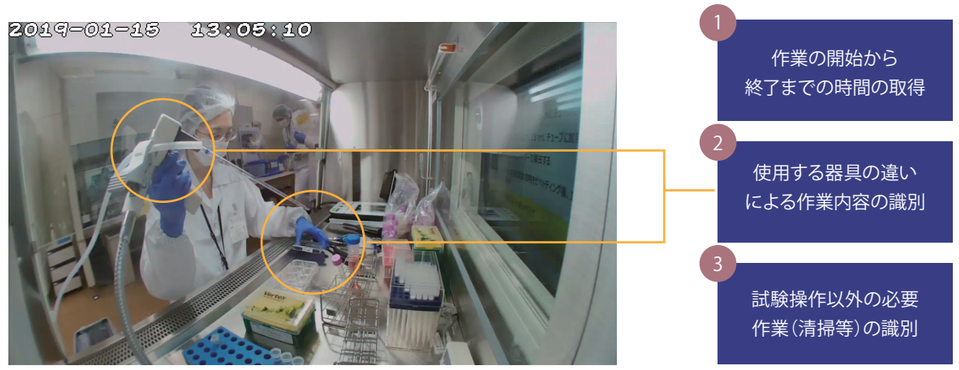


Figure.5 マルチモーダル AI とシングルモーダル AI の認識精度比較

行動認識精度の比較

モーダル	全体	低難易度	高難易度
静止画	△	△	△
動画	○	○	○
骨格	△	△	△
テキスト	△	△	△
音声	×	×	×
MM-AI	◎	◎	◎

【事例 1】ライフサイエンス研究の記録

ライフサイエンス研究の分野で必須となる、実験ノート作成の高度化に適用した事例。研究員が行う実験作業を映像と発話を基に行動を判別し、実験記録ノートを自動で作成します。研究者の発する言葉だけでなく行動を視覚的に捉え、時には機器操作音などの音も利用し、統合的に判断します。

なお、当事例では AI による行動認識の精度を数値化・比較しており、「一般的な行動（例：座る）」と「実験特有の作業（例：電動ピペットとプレー

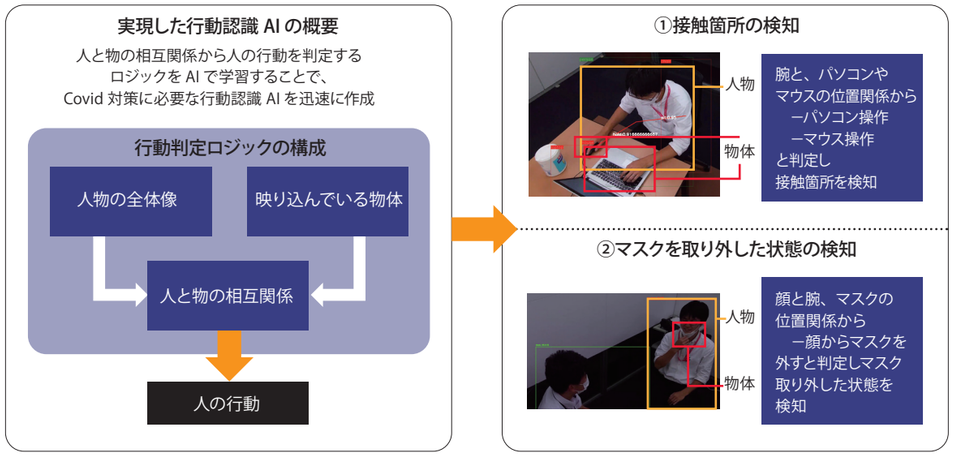
トの同時操作）」といった、認識すべき行動の詳細レベルが異なる場合であっても、マルチモーダル AI はシングルモーダル AI よりも高い認識精度を達成しました。

【事例 2】接触個所の検知

COVID-19を意識した事例。飲食店やオフィス等で、清掃作業が必要になる人が接触した箇所や、濃厚接触に該当するようなマスクを取り外した状態をそれぞれ検知します。検知した情報を活用することで、清掃作業の作業もれや、発病者との濃厚接触者の特定の迅速化につながります。

Figure.6 接触個所の検知

人の動きの「全体像」と映っている「物体」のモーダルを活用することで、
清掃作業の必要な①接触箇所と、濃厚接触に該当し得る②マスクを取り外した状態を検知



【事例 3】迷惑行為や不正行為の監視

防犯分野に適用した事例。マンション敷地内で迷惑行為につながる行動がないかを監視し、画像や音など複数の条件を掛け合わせて判断します。迷惑行為には、マンション・エントランス内でスケート・ボードを乗り回すといった行動もあれば、言い争いをしているような行動もあります。スケート・ボードの乗り回しは映像だけでも判断できますが、言い争いの場合は動作が小さくても声の大きさが迷惑行為にあたるため、映像と音量に基づいた統合判断を行う必要があります。

Figure.7 迷惑行為や不正行為の監視

監視カメラの「映像」と、音量マイクの「音情報」のモーダルを活用することで
マンション内での禁止行為を自動検知し、警備員の業務支援への適用可能性を検証

CHAPTER.5
AI が総合的な俯瞰と判断を行う時代へ

■AI が総合的な俯瞰と判断を行う時代へ

「マルチモーダル AI」が確立すれば、AI は人間の五感に比肩する認識機能を持ち、俯瞰的な状況把握や柔軟な対処が可能になります。AI の適用領域は特化型から汎用型に向け一気に拡大し、例えば「人間味を感じられる見守りロボット」や「心情を理解した議事録が書ける AI」など、現在の AI では人間が満足できる結果を出せていない領域に踏み込んだ新たな用途が開拓されるでしょう。NTT データでは、一刻も早い現場適用を目指して「マルチモーダル AI」の研究開発／ PoC を推進しています。

　問い合わせ先：NTT データ　技術開発本部　rdhkouhou@kits.nttdata.co.jp