

## 経済産業大臣賞

# 「映像認識 AI×LLM」に関する研究開発と産業適用

1 日本電気株式会社 ビジュアルインテリジェンス研究所

2 日本電気株式会社 データ基盤サービス統括部

劉 健全<sup>1</sup> 船田 純一<sup>1</sup> 鈴木 哲明<sup>1</sup> 山崎 智史<sup>1</sup>  
藤若 雅也<sup>1</sup> 平川 康史<sup>1</sup> 佐々木 洋平<sup>2</sup> 進藤 直樹<sup>2</sup>

## 1. 緒言

現代社会では、防犯カメラ、車載カメラ、スマートフォン等により膨大な映像データが日々生成されている。これらの映像は事故の記録、業務プロセスの確認、顧客行動分析など計り知れない価値を潜在的に持つが、その多くは未活用のまま埋没しているのが現状である。産業デジタルトランスフォーメーション(DX)推進において映像利活用は生産性向上や価値創出の鍵であるが、データ量の爆発的增加に対し、人手による確認・分析は限界に達しており、自動化・効率化が社会的な需要となっている。

既存の映像認識 AI は物体検出等の要素技術では進歩したが、映像全体の文脈やイベント間の因果関係、即ち映像が内包する「物語」の解読には限界があった。一方、大規模言語モデル(LLM)はテキスト処理能力が高いものの、動的で複雑な動画データへの直接適用は計算コストや時間制約、さらにハルシネーションのリスクという課題を抱える。結果として、長時間映像から真に有用な情報を効率的かつ正確に抽出する決定的な技術は無かった。

本稿で提案する「映像認識 AI × LLM」は、これらの課題解決を目指し開発された世界初の技術である。NEC の高度な映像認識 AI 群と最新 LLM を独創的なアーキテクチャで融合し、長時間動画の内容を深く理解、利用者の自然言語指示に応じ関連シーンを短縮動画で抽出し、内容を「物語」として説明する文章を自動生成する。社会科学の物語概念に着想を得た「ユーザー視点の物語生成」という斬新なアプローチを取り入れることで、本技術を世界で初めて開発した。本稿では、まず背景と技術課題を述べ(2章)、提案技術詳細を解説(3章)、次に製品化と応用事例を紹介し実用性、潜在的可能性、市場性を示し(4章)、最後に意義と展望をまとめる(5章)。

## 2. 研究開発の背景と技術課題

デジタル化の波は社会の隅々にまで及び、映像データはかつてない規模で生成・蓄積されている。交通モニタリング、工場、医療現場、店舗、個人デバイス等から生成されるデータは価値の源泉だが、その活用は既存技術の限界により阻まれていた。

第一に、従来の映像認識 AI の限界である。要素技術、例えば物体検出や行動認識の精度は向上したが、それらは映像を構成する部品の認識に過ぎない。部品間の関係性、時間的な文脈、背景にある状況や因果関係といった、映像全体の意味構造を捉える能力は依然として低い。例えば交通事故映像から単に「車」や「人」を検出するだけでなく、「なぜ事故が発生したのか」という因果関係まで理解することは、従来の技術では困難であった。また、長時間にわたる映像データから特定の重要なシーンを自動的に発見するスケーラビリティも欠如していた。

第二に、LLM の動画データへの適用限界である。LLM はテキスト処理において驚異的な能力を発揮するが、情報密度が高く時空間構造を持つ複雑な動画への直接適用はハードルが高い。膨大なトークン長となるため計算コストが爆発し、現行アーキテクチャでは動画特有の動的な変化や長期的な時間依存性を捉えるのが不得意である。さらに、LLM が学習データに含まれない事象や誤った知識に基づいて、もっともらしい虚偽の情報(ハルシネーション)を生成するリスクは、特に客観的な事実性が求められる映像分析においては致命的な欠点となりうる。

第三に、これら技術的限界を補う人手分析の限界である。結局、多くの場合、映像データの最終的な確認・判断は人間に委ねられてきた。しかし、増大し続ける映像データを人間がすべて目視で確認することは、時間的にもコスト的にも現実的ではない。長時間にわたる確認作業は、集中力の低下を招き、重要な情報を見逃すリスクを高める。また、分析者の経験や主観によって解釈がばらつき、客観性や一貫性を担保することも難しい。

これらの問題が複合的に絡み合い、映像データという価値ある情報源を有効活用できない状況を生んでいた。この膠着状態を打破すべく、我々は映像解析へのアプローチを根本から問い直した。社会科学や認知科学の知見から、「人間は世界を物語として理解する」という根源的な事実に着目した。そして、映像解析の問題を、「利用者の知りたいことに応じて、映像という生のデータから関連する事実群を抽出し、それらを意味のある『物語』として再構成し、分かりやすく伝達する」という独創的な問題設定へと転換した。

この発想に基づき、映像認識 AI と LLM の最適な役割分担と連携による独創的な融合フレームワークを考案した。すなわち、映像認識 AI に映像内の客観的な「事実」抽出を、LLM に利用者の「意図」理解と、事実に基づく「物語」の構成・説明を担わせる。これにより、映像認識 AI の文脈理解不足と LLM の信頼性問題を同時に解決する道筋をつけた。提案技術は、この思想を具現化するものであり、事実データベース(グラフ構造 DB) と対話型インターフェース(LLM)を通じて、利用者が膨大な映像データの中から必要な物語を自在に引き出すことを可能にする。次章でその詳細を述べる。

### 3. 提案技術：映像認識 AI × LLM

本研究で開発した技術は、映像認識 AI と LLM を高度に連携させた世界初の生成 AI フレームワークである(図1, 図2)。「ユーザー視点の物語がある映像要約」をコンセプトとし、長時間動画から利用者の意図に合致した情報を効率的に抽出し、理解しやすい「物語」として提示する。その独創性は、両 AI 技術の強みを活かし弱点を補完する相乗効果を生むアーキテクチャにある。

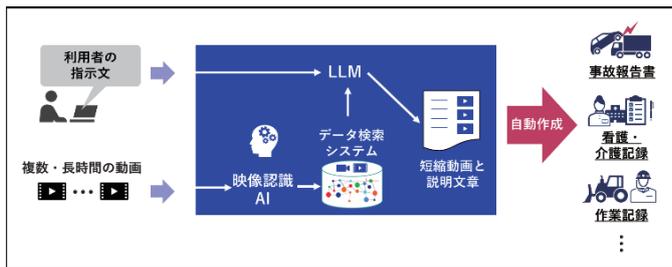


図1：開発した生成 AI フレームワーク「映像認識 AI × LLM」



図2：本技術の実演デモ画面

#### 3.1 フレームワーク概要

処理は大きく二段階構成である。第一段階(前処理：オフライン)では、入力映像に対し、100種類以上の NEC 独自映像認識 AI 群(物体、人物、行動、環境等を認識)を適用し、多様な要素とその時間変化を詳細に認識する。これらの認識結果は単なるタグではなく、要素間の関係性や時間的連続性を示す独自のグラフ構造として表現され、高速検索可能なデータベースに格納される。データは圧縮されつつ意味構造が保持され、効率的な分析基盤が構築

される。第二段階(オンライン問い合わせ処理)では、利用者が自然言語で指示を入力する。LLMが指示文の意図を深く理解し、検索クエリに変換する。データ検索システムがグラフ構造DBから合致する部分グラフ(シーンに対応)を高速に抽出する。抽出部分グラフに基づき、関連映像セグメントから短縮動画が生成されると同時に、LLMが部分グラフ内の事実情報から状況や因果関係を説明する物語性のある要約文章を自動生成する。必要に応じ、報告書フォーマットに整形される。これにより、利用者は専門知識なく自然言語でAIと対話し、映像から必要な情報を物語として引き出せる。

### 3.2 構成コンポーネント詳細

1. **映像認識 AI 群**：物体検出・追跡、人物検出・属性推定・顔認識、行動認識(歩行、転倒等)、車両認識、標識・信号認識、環境認識(天気、昼夜等)、シーン分類など、100種以上のエンジン群。NECの長年の研究開発の蓄積であり、OSSも活用。多様な映像への網羅性と高精度認識を実現し、客観的な「事実」抽出の基盤となる。
2. **グラフ構造データベース**：映像認識結果を、ノード(オブジェクト、イベント)とエッジ(関係、時間遷移)で構成されるグラフ構造として格納。意味的関連性や時間順序を保持しつつデータをコンパクト化。NEC独自のDB技術により、大規模データに対する高速な類似検索、複雑なイベントパターン分析を可能にする。独創的なデータ表現・管理技術である。
3. **大規模言語モデル(LLM)**：自然言語理解と生成を担当。利用者の曖昧な指示からも意図を解釈し、検索条件を生成。抽出されたグラフ構造(事実情報)に基づき、文脈、因果関係、時間経過を考慮した一貫性ある説明文(物語)を生成。NEC開発の軽量・高性能LLMを活用し、GPU1基搭載PC等でも動作可能とし、オンプレミスやエッジ運用を容易にする。特定ドメイン知識を学習させるファインチューニングにより専門性の高い要求に対応。重要な点として、生成プロセスをグラフ構造DBの事実情報に強く制約する(事実ベース生成)ことで、LLMの弱点であるハルシネーションを効果的に抑制し、出力の信頼性を担保する。
4. **データ検索システム**：LLM生成クエリに基づき、大規模グラフ構造DBから関連性の高い部分グラフを高速・高精度に検索・照合。テキスト情報とグラフ構造間の高度な類似度計算(ベクトル検索等)を用いる。応答性の高さが実用性を支える。
5. **生成・出力モジュール**：抽出部分グラフに対応する短縮動画を生成、LLM生成の説明文と同期させて提示。多様な業務報告書フォーマットへの自動マッピング機能も持つ。

### 3.3 提案技術の特徴・世界初の機能と性能

本技術は以下の点で卓越し、高い独創性、新規性、実用性、潜在的可能性を示す。

- **世界初の統合技術(独創性・新規性)**：映像認識AIとLLMを実用レベルで連携させ、動画からの物語生成を実現した点は世界初。コンセプト自体もACM Multimedia 2022で受賞[1]して、国際的に新規性が認められている。
- **高効率・高速シーン特定(実用性)**：人手による目視確認不要となり、分析時間を大幅に短縮(例：損保業務で半減)。事前に映像処理を実施しておくことで、1時間超の映像からの物語生成も数秒で処理可能。

- **高信頼・高品質な要約・報告書(実用性)**：事実ベース生成によるハルシネーション抑制。ファインチューニングによる専門知識獲得。専門家レベルの品質を実現。
- **低コスト・柔軟な導入(実用性・市場性)**：GPU1基で動作可能。クラウド、オンプレミス、エッジに対応。導入・運用コストを低減。
- **容易な操作性(実用性)**：自然言語による対話的インターフェース。専門スキル不要。
- **広範な応用可能性(潜在的可能性・市場性)**：フレームワークの汎用性が高く、多様な産業・ユースケースに適応可能。CEATEC AWARD 2024受賞[2]もその汎用性を評価。

これらの特徴により、本技術は映像データ活用のブレークスルーとなり、社会の様々な場面で価値を提供する。

## 4. 製品化と幅広い応用

### 4.1 製品化アプローチ

本技術の潜在的可能性を現実価値へ転換し社会実装を加速するため、実用性を最重視したアジャイルな製品化アプローチを採っている。技術を早期に顧客が評価し、現場ニーズを迅速に反映させることを目的とし、2024年3月より「NEC Multimedia OLAP 映像認識 AI×LLM 有償 PoC サービス」として試用版提供を開始した[3]。これにより、損害保険会社や自動車メーカー等の顧客は、自社データを用いて本技術の効果を検証し、具体的な導入検討を進めることが可能となった。我々はこの試用版提供を通じ、現場の生の声(課題、要望、活用アイデア等)を直接収集する。フィードバックを迅速に技術改良やソリューション機能開発に反映させるサイクルを繰り返し、真に顧客ニーズに合致した、実用性の高い製品・サービスを顧客と共に創り上げる「顧客共創型」開発プロセスを実践している。

### 4.2 幅広い応用(潜在的可能性・実用性・市場性)

「映像認識 AI×LLM」は、その汎用性の高さから、極めて広範な産業分野への応用が期待される。映像データが生成され、その内容理解や効率的な活用が求められるあらゆる場面で価値を発揮しうる。保険、交通、航空、医療・介護、製造、建設、小売、放送、スポーツ&エンタメ、自動車開発、モビリティサービス等の多様な領域において、従来人手に頼っていた映像確認・分析業務を自動化・高度化し、業務効率の大幅向上、ヒューマンエラー削減、安全管理レベル強化、新たな知見発見、革新的サービス創出に貢献する。これらの応用分野における市場規模は巨大であり、本技術が秘める市場性は計り知れない。以下に主要な応用事例を挙げ、その実用性、潜在的可能性、市場性を基礎情報に基づき記述する。

- **損保領域**

自動車保険の事故調査は、本技術の実用性が実証されている主要応用分野である。事故時のドライブレコーダー映像は原因究明や過失割合算定の重要証拠だが、長時間映像の確認は調査員に多大な負担を強いる。本技術は、映像認識 AIが事故の瞬間や関連状況(信号の色、車両の動き、歩行者有無等)を詳細に認識する。LLM がこれらの客観的事実を統合・分析し、「赤信号を無視したトラックが交差点に進入し、青信号で直進してきた乗用車と衝突」(図2)といった具体的な事故状況と推定原因を、報告書形式の自然言語

で記述し、対応する短縮動画と共に自動生成する。検証では、従来数時間～数日要していた映像確認・報告書作成時間を半減できることを確認済みである。この実用性により、査定業務の迅速化・効率化、判断の客観性向上、不正請求抑止への貢献が期待される。2024年3月より試用版を提供開始し、複数の損保会社等が導入検討を進めている。関連市場規模は約8000億円[13]と推計され、高い市場性が見込まれる。

● 航空領域

航空需要回復[4]に伴い、グランドハンドリング業務の効率化と安全性向上が急務だが、人手不足が課題である[5]。本技術は、駐機エリアのカメラ映像を解析し、航空機準備状況(ランプオペレーション：航空機誘導、貨物搬入、燃料補給、ケータリング準備等)を把握・要約する(図3)。LLMが各作業の状況を自然言語で報告し、遅延予測や人員配置最適化を支援、定時運航率向上に繋げる。技術的には、駐機エリア映像に対し映像認識と生成AIを適用し、航空機準備状況を把握・要約する。今後の発展として、現場コミュニケーションや作業記録等、蓄積データを組合せ分析することで、業務進捗解釈とアラート発出、過去データ比較による運用状況要約等、業務種別に合わせた情報要約の検討を進めている。国際航空宇宙展[6]やCEATEC2024[7]でのデモ展示では、顧客から予想を超える良い反響を得ており、その潜在的可能性は大きい。グランドハンドリング市場は約4.75兆円規模[14]とされ、高い市場性を持つ。



図3：航空機の準備状況の把握／要約デモンストレーション

● スポーツ&エンタメ領域

スポーツ中継等からハイライトシーンを迅速に生成・配信することは、視聴者エンゲージメント向上に重要である。本技術は、映像認識と生成AIを活用し、「ハイライトシーンのリアルタイム自動抽出と要約文章の自動生成および投稿」を行う。従来、ハイライト・要約は試合終了後にまとめて配信されることが多いが、本技術ではラウンド間のイン

ターバルに直前ラウンドのハイライト抽出と要約文生成を行い、配信先媒体(SNS等)仕様に合わせて配信する技術を開発した。2024年9月のプロボクシング世界戦での実用性実証実験[8]では、各ラウンドのハイライト・要約文がインターバル時にSNSへ配信されることを確認した(図4)。配信映像の複数データを分析するマルチモーダル分析により、ラウンド毎の特徴的シーン抽出、ダイジェスト文及び各選手の動きに焦点を当てたハイライト・要約文生成を実現。臨場感や特定選手描写など多様な視点のストーリーが感じられる要約を可能にした。このリアルタイム性と物語生成能力に新規性がある。放送映像向け特定コンテンツ収集(市場規模4200億円[15])やナレーション原稿作成への応用も期待され、一定の市場性が見込める。



図4：SNS へのハイライトシーン・要約文章配信例

- 自動車開発領域

ADAS(先進運転支援システム)や自動運転機能の高度化に伴い、開発時の設計・評価・デバッグには膨大な手間がかかる。特に実走行データ(映像、センサー情報等)から不具合や性能問題シーンを抽出する作業は、従来人手による目視確認が必要で時間とコストを要した。本技術により、収集した大量データから必要なシーン抽出を効率化できる。映像認識AIが車両・歩行者・路面・信号等を認識しメタデータとしてDBに蓄積する。利用者が自然言語で必要なシーンを指示すると、DBから合致シーンを探し、状況の要約文と該当映像を出力する。特に、想定外の稀な交通状況(コーナーケース)での不具合特定に有効である。多様な対象に対する映像認識AIのメタデータに対し、LLMで状況を要約することで、複数の多様なシーンの組み合わせを自然言語で指定し対話的に検索可能となり、効率的な不具合データ収集とデバッグが可能となる(図5)。自動運転用AI学習データの収集・説明文作成も効率化する。高い実用性があり、約7500億円規模[16]の関連市場で高い市場性が期待される。

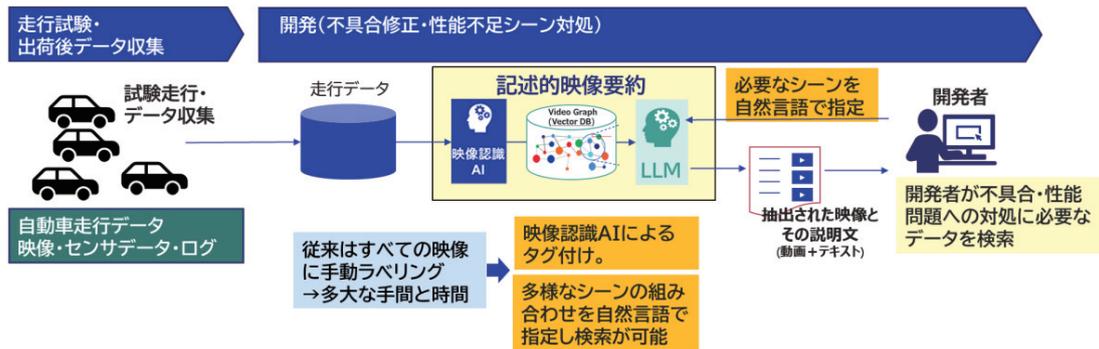


図5：映像認識 AI と LLM を活用した走行データ検索・要約の例

・ 医療領域

医師の働き方改革に伴う労働時間短縮要請[9]により、業務効率化が急務である。本技術は、医師業務の一つである手術記録作成支援に貢献できる。近年、手術映像は術野カメラ、術場カメラ、内視鏡カメラ等で撮影されている。これらの映像を分析することで、どの医師がいつどのような作業を実施したかを認識し、文字と画像を用いた手術記録の素案を自動生成できる。手術を担当した医師は、AI作成の素案を確認・修正・追記するだけでよくなり、記録作成の効率化が期待できる(図6)。これにより医師の負担軽減、記録の客観性・網羅性向上、医療安全強化に繋がる。看護・介護記録作成支援への応用も有望である。既存の手術映像インフラを活用できる点も実用性を高める。医療 DX 需要増大を背景に、約13.8兆円規模[17]のヘルスケア IT 市場における本技術の潜在的可能性と市場性は極めて大きい。

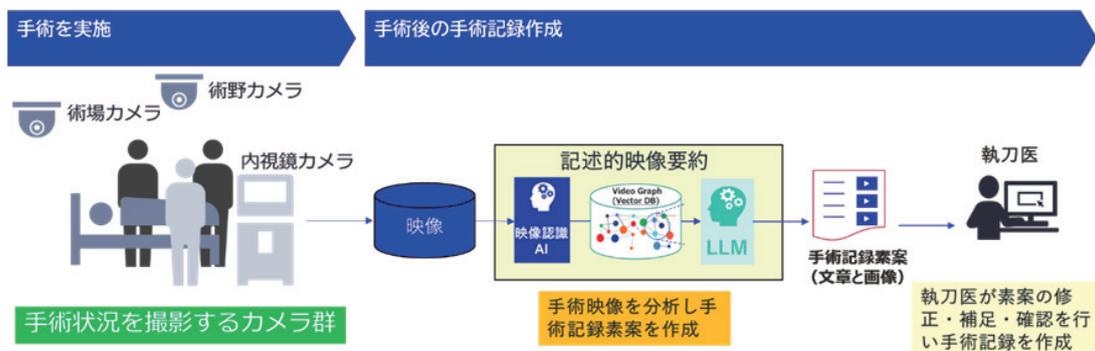


図6：映像認識 AI と LLM を活用した手術記録作成支援のイメージ図

・ モビリティ領域(運転支援・診断)

ドライブレコーダー映像は事故証拠だけでなく、安全運転促進にも活用され始めているが、膨大データの効率的・効果的分析と適切なフィードバック手法が課題であった。本技術では、運転映像からドライバーの運転操作や周囲の走行環境を AI が理解し、安全・省燃費運転の実践度を総合的に診断、運転習慣改善を促す具体的なフィードバック文章を生成可能となる(図7, 図8)。フリート管理システムでは、GPS や加速度センサーデー

タに加え映像情報を活用し、危険操作(急ブレーキ等)発生時の状況を把握、より適切なフィードバックを提供し、人手による映像確認コストを削減できる。テレマティクス保険では、映像活用により運転診断精度を高め、より公平な保険料算定と効果的な行動変容促進に繋げる。自動車メーカーの運転診断サービスでは、車両センサーデータと映像情報を組み合わせ、運転操作タイミングを含めた詳細分析と効果的なフィードバック提供を可能にする。映像情報を活用した高精度診断とフィードバック提供という点で実用性があり、Mobility as a Service(MaaS)やコネクテッドカー市場の発展と共に市場性も期待される。

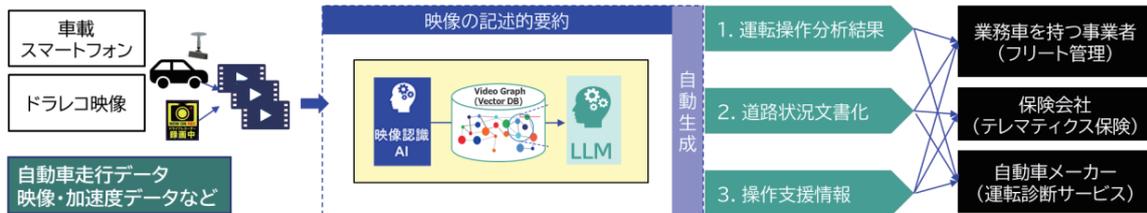


図7：映像認識 AI と LLM により運転操作や道路状況及び操作支援情報を生成



図8：映像認識 AI と LLM により出力された運転操作およびフィードバック文章の例

- 製造・物流・建設領域

生産人口減少に伴い、これらの現場では作業効率化や安全管理支援の必要性が高まっている。本技術はこれらの現場支援に貢献できる。製造・物流業では、NECは映像解析を用いた現場作業状況分析に取り組んでおり[10]、その分析結果に本技術を適用することで、現場作業の課題洗い出しと業務改善案作成が可能になると考えている(図9)。建設業では、ICT 土工等が進められ生産性向上が図られている[11]。NECも現場カメラ映像分析による業務フロー効率化やリソース最適化の検証・実用化を進めている[12]。これらの映像分析結果に本技術を適用することで、現場の詳細な作業内容を含む作業報告書の自動生成が実現可能と考えている。これにより、生産性向上、品質安定化、安全管理強化、技能伝承支援、トレーサビリティ確保に貢献する。スマートファクトリー/ロジスティクス/コンストラクション実現に不可欠な技術として高い実用性があり、建設プロジェクト管理ソフトウェア市場(約2500億円[18])を含む巨大な市場性を持つ。

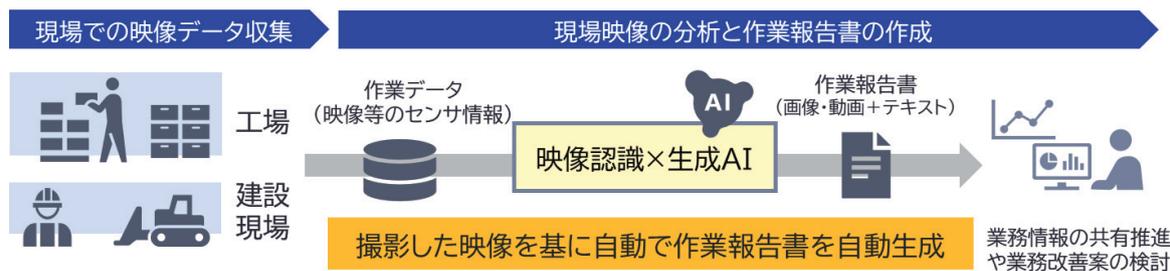


図9：製造／建設現場での映像認識×生成 AI を活用した業務効率化例

これらの多様な応用事例は、本技術が持つ卓越した汎用性、深い潜在的可能性、そして確かな実用性と市場性を具体的に示している。

## 5. 結 言

本稿では、映像認識 AI と LLM を世界で初めて融合した革新的技術「映像認識 AI × LLM」を提案した。本技術は、社会科学の物語概念に着想を得た独創的アプローチに基づき、長時間動画から利用者の意図を汲み取り、状況や因果関係を物語として記述する説明文章と関連短縮動画を自動生成する。その新規性は ACM Multimedia 2022 最優秀 BNI 論文賞、汎用性と潜在的可能性は CEATEC AWARD 2024 25 周年特別賞により国際的・国内的に評価されている。損保業務での事故報告書作成時間半減など、多様な産業での実証を通じて高い実用性が示されている。業務効率化、安全性向上、コスト削減、新たな価値創出に貢献し、巨大な市場性を持つ。試用版提供によるアジャイルな顧客共創で社会実装を加速しており、映像データ利活用のパラダイムシフトを促す基盤技術として産業 DX を推進する。今後も技術改良と適用領域拡大を進め、安全・安心で効率的な社会の実現に貢献していく所存である。

## 参考文献

- [1] 共同研究先 (NUS：シンガポール国立大学) の Web サイト掲載 (2023/2/2).  
<https://www.comp.nus.edu.sg/news/2023-bni-award/>
- [2] CEATEC AWARD 2024 オフィシャルサイト発表. <https://www.ceatec.com/ja/award/>
- [3] NEC Multimedia OLAP 映像認識 AI × LLM の有償 PoC サービス (試用版) を提供開始.  
[https://jpn.nec.com/data/multimedia\\_olap/service002.html](https://jpn.nec.com/data/multimedia_olap/service002.html)
- [4] 国土交通省、“航空輸送統計速報(令和6年(2024年)分)”.  
[https://www.mlit.go.jp/report/press/joho05\\_hh\\_000838.html](https://www.mlit.go.jp/report/press/joho05_hh_000838.html)
- [5] 国土交通省、“業務(グラハン・保安検査)の現状と課題”.  
<https://www.mlit.go.jp/koku/content/001588945.pdf>
- [6] 2024 国際航空宇宙展、<https://www.japanaerospace.jp/jp/>
- [7] CEATEC、<https://www.ceatec.com/ja/>
- [8] NEC、井上尚弥選手のボクシング世界タイトルマッチにおいて、映像認識 AI × LLM を活用したハイライトシーンのリアルタイム SNS 配信の実証実験を実施。  
[https://jpn.nec.com/press/202409/20240905\\_01.html](https://jpn.nec.com/press/202409/20240905_01.html)

- [9] 厚生労働省、“医師の働き方改革”. <https://www.mhlw.go.jp/content/10800000/001129457.pdf>
- [10] NEC, “NEC Digital Twin 現場可視化・分析サービス”.  
<https://jpn.nec.com/iot/platform/optimization-va/index.html>
- [11] 国土交通省、i-Construction、<https://www.mlit.go.jp/tec/i-construction/index.html>
- [12] NEC, “映像から複数の人物の多種多様な作業内容を高精度に認識する技術を開発”.  
[https://jpn.nec.com/press/202211/20221128\\_01.html](https://jpn.nec.com/press/202211/20221128_01.html)、2022年11月
- [13] Global Insurance Market Report (GIMAR). <https://www.iais.org/activities-topics/financial-stability/gimar/>
- [14] Allied Analytics LLP: Airport Ground Handling Market in 2025 - Competitive Analysis and Industry Forecast. February 12, 2025.  
<https://www.einpresswire.com/article/785403231/airport-ground-handling-market-in-2025-competitive-analysis-and-industry-forecast-at-a-9-2-cagr-from-2023-to-2032>
- [15] GEM Partners「動画配信 (VOD) 市場5年間予測 (2024-2028年) レポート」.  
<https://www.gem-standard.com/columns/789>
- [16] GRAND VIEW RESEARCH: Driving Simulator Market Size & Share, Industry Report, 2030.  
<https://www.grandviewresearch.com/industry-analysis/driving-simulator-market>
- [17] GII 市場調査レポート「医療 IT の世界市場：用途別・地域別の展望、業界分析、将来予測 (2021年～2027年)」. <https://prtimes.jp/main/html/rd/p/000000730.000071640.html>
- [18] PPL (Market.us) 市場調査レポート「建設プロジェクト管理ソフトウェアの市場規模は、2028年までに約2億2,580万」. <https://prtimes.jp/main/html/rd/p/000000058.000076785.html>