

エッジ AI ソリューションの可能性と課題

萩原エレクトロニクス H-AI カンファレンス

宗像尚郎

hisao.munakata.vt(at)renesas.com

ルネサスエレクトロニクス株式会社
ハイパフォーマンスコンピューティングプロダクトグループ
SoC ソフトウエアイネーブルメント部 シニアダイレクタ

2025-12-15

自己紹介 (Who am I)

ルネサスで オープンソースを活用した SW ソリューション の開発に従事しています

■ ルネサスエレクトロニクス株式会社

- 日立、三菱、NEC の半導体事業を統合し、グローバル半導体メーカーに発展
- 自動車向け 大規模ソフトウェア基盤 (OS、プラットフォーム) 開発 に従事
- エッジ AI 開発プラットフォーム (SparrowHawk) を開発し 2026 年 1 月発売予定
- 社内のオープンソース開発活動 (Linux kernel 開発など) を管掌

■ オープンソース開発コミュニティ、コンソーシアム活動 (会社公認の社外活動)

- The Linux Foundation 理事 (2017~2022)
- AGL (Automotive Grade Linux) プロジェクト、yocto プロジェクト理事
- COVESA (Connected Vehicle System Alliance) 理事
- 社外講師、講演、インターフェース誌などへの投稿など多数

「オープンソースを活用した製品開発」で重要となる産学官の連携も支援しています

今 AI が全てを解決するという幻想が蔓延 (AI バブル)、それでも...

BUSINESS INSIDER

Subscribe | Newsletters

DOW JONES ▼ -0.45% NASDAQ ▼ -0.25% S&P 500 ▼ -0.35% AAPL ▼ -0.2% NVDA ▲ +2.36% MSFT ▼ -0.15% AMZN ▲ +0.02% META ▼ -0.34% TSLA ▼ -0.02%

AI

Bill Gates says we're in an AI bubble similar to the dot-com bubble

By Bryan Metzger [+ Follow](#)



"There are a ton of these investments that will be dead ends," Gates said. Bryan Bedder/Getty Images for Bloomberg Philanthropies

- Bill Gates says we're in the midst of an AI bubble.
- He says it's probably not like the "tulip mania" that took place in the Netherlands in the 1630s.
- It could be like the dot-com bubble, where some companies ended up overvalued.

<https://www.businessinsider.com/bill-gates-ai-bubble-similar-dot-com-bubble-2025-10>

2025 年は AI が普通の生活者のツールとなった 歴史的な転換点

「ChatGPT (LLM 技術)」が破壊的なブレークスルーをもたらした

- 組み込み機器にも実装可能な “認識 (Perception AI)” 系の技術は既に実用段階
 - 音声インターフェース (DNN) → アレクサ、Siri などのフロントエンド
 - 物体認識 / 画像判定 (CNN) → 良品 / 不良品の判定などに応用
 - 物体認識は 自動車の運転支援 (ADAS) に広く応用 (標識、車線、前车などの認識)
- 2025 年にコンピュータが 自然言語を “理解 (Cognitive AI)” するようになった
 - プログラムを書かずにコンピュータをフル活用 ⇒ コンピューティングの民主化
 - ネット上に散在している膨大なコンテンツを AI が検索・分析して回答
 - これまで人間がやってきたかなりの作業は AI の方が上手くできると分かった
 - (特定の領域では) AI が人間の能力を超える「シンギュラリティ」を実感

今後、先行する「認識系 AI」と「理解系 AI」が融合した全く新しい世界観が拓ける？

第一部：エッジ AI の現在と未来

認識型エッジAIの現状

現代の AI の二大潮流（認識型・理解型）

認識型の AI（Perception AI）

- アルゴリズム：CNN、DNN など
- モデル：ResNet, Yolo, MobileNet
- 処理内容：認識、検出
 - 標識認識、車線認識 → 運転支援
 - 音声認識 → アレクサ、Siri など
- 推論処理は電力効率に優れた組み込みプロセッサ（10TPOS 級）で実行可能
- ほぼリアルタイムで高精度に物体認識ができる（理解系 AI の不得意領域）

理解型の AI（Cognitive AI）

- アルゴリズム：Transformer が主流
- モデル：GPT, Llama, Stable Diffusion など
- 処理内容：生成、対話
 - 自然言語による対話
 - 画像生成
- 認識系 AI の 10 倍～100 倍の計算資源（CPU、メモリー）が必要になるので、現状はクラウドの計算パワーが不可欠

認識型 AI の推論処理は 10TOPS 程度で実行可能なので組み込み機器に搭載可能

AI の実行には「学習（モデル作成）」と「推論」の2段階が必要

学習（Training）

- 大量のデータを使ってモデルのパラメータを最適化すること
 - 教師あり学習（ラベル付きデータ）
 - 教師なし学習（強化学習）
- 学習には **大きな計算資源** が必要
 - CPU 能力、ストレージサイズなど
- AI 認識の信頼度は、**計算精度よりも学習したデータの量に比例** する

推論（Inference）

- 学習済みのモデルを利用して、対象データを分類すること
- 推論処理は **相対的に軽量** で、SOC に内蔵された HW 資源を有効活用できる
 - CPU（汎用）
 - GPU（ビデオカード）
 - DSP（積和演算器）
 - NPU（専用 AI エンジン）

推論処理にはクラウドの計算資源は必須でない ⇒ エッジだけで AI 処理完結が可能

AI 推論をエッジだけで実行する：その「アドバンテージ」と「課題」

積極的に「エッジ内で AI 認識処理を実行させたい場面」が増えている

■ アドバンテージ

- リアルタイム応答性：クラウド呼び出しが不要
- 認識精度の高さ
- データをアップロードしないので プライバシー／セキュリティが担保される
例、独居老人の見守りなど
- インターネット ネットが利用できない環境 でも実行可能
例、ドローン、ロボットへの適用など

■ 課題

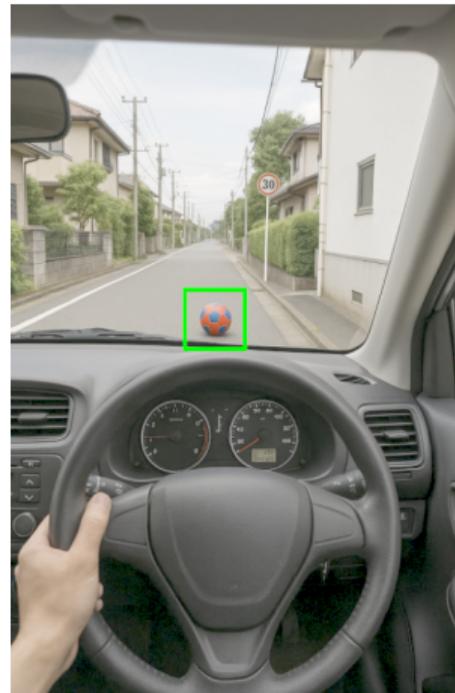
- 組み込みプロセッサの処理能力：⇒ NPU 内蔵の組み込みプロセッサが充実
⇒ エッジ側で AI 推論を実行できる環境が揃ってきた
- 電力効率 (TOPS/W) の最適化：クラウド AI の電力要求は青天井
- プログラム API：CUDA 以外の AI 実行環境の整備

認識型エッジAIの再定義

CNN (Convolutional Neural Network)

CNN は「画像に何が写っているかを識別」する

- 事前に大量の画像から「検出したい物体」のモデルを用意
- 画像の特徴量を抽出し、視覚情報の中からパターンを認識
- 識別結果は構造化された固定形式で出力される
 - ラベル、ボックス、マスクなど
- 代表的な用途
 - 物体分類 (Cat / Dog / Car)
 - 特徴量抽出 (Edges, Shapes)
 - 位置検出・領域分割
 - 顔認識・特徴比較
 - 道路標識・車線認識
- 学習していない物体は特定できない (例、道路で寝ている人)



VLM (Vision Language Model)、VLA (Vision Language Action)

VLM = 視覚情報を意味的に理解

- “画像の意味・関係性・意図”を推論
- マルチモーダル（画像 + テキスト）
- 認識精度や速度は CNN よりも劣る
- 特徴
 - 画像の意味を説明する
 - 画像に関する質問に応答する
 - 誰が何をしているかを推論
 - 画像から意図を推定（例：道路上のボールから子供の飛び出しを予測）
 - 見えていない事象の推測（場面補完）

VLA = 視覚情報から何をすべきか推定

- VLM（画像と言語の統合理解）をさらに進化させ、視覚・言語・行動（Action）を統合的に扱うモデル体系
- ロボティクスやエージェント AI において、「見る → 理解する → 行動する」を一貫して学習・推論できる 枠組み
- 特徴
 - 言語指示だけで複雑なタスクを遂行
 - 未学習のタスクにも適応できる
 - 環境と目的に応じた柔軟な行動生成

AI エージェント (Agentic AI)

それぞれが別々の役割を持つ AI が協調してタスクを解く最先端のアーキテクチャ

- 今までは **人間が AI に指示 (プロンプト)** → 今後は **AI 同士が自律連携** する
- 単一 AI モデルでは実世界の課題解決に役不足なので、役割をエージェントに分割
 - **認識エージェント**：外界を観測し意味的な状態を生成
 - **計画エージェント**：目的に基づき計画立案・推論・判断を行う。
 - **行動エージェント**：指示を実行し、環境に働きかける
 - **記憶エージェント**：長期・短期記憶、コンテキスト保存
- **AI エージェント間を密連携** されるためのプロトコル
 - **MCP (Model Context Protocol)**：エージェントの権限・能力管理
 - **A2A (Agent-to-Agent)**：エージェント間の協調

画像理解系 AI の実装では **認識エージェントの精度や速度** がボトルネックになる

認識系エッジ AI の再定義 = VLM / VLA の認識エージェント

これまでの認識型エッジ AI (eAI 1.0)

- 画像の中の **物体を識別** する
- 識別以上の **意味解釈はできない**
- 独立した **一つの AI** として機能する
- アプリとは **Rest API 等で疎連携** する
- 学習には **大量のデータと計算量** が必要
- 学習していない物体は検出できない

認識型エッジ AI の再定義 (eAI 2.0)

- エッジ AI が物体識別した結果直接 **後段の理解型 AI の入力** となる
- エッジ AI は **識別エージェント** としてリアルタイムで **高精度な識別** を担う
- **MCP、A2A プロトコル** に対応させて後段の **理解型エージェントと密連携**
- 未知の物体にも一定レベルの解釈

今後 **CNN** を搭載したエッジ AI デバイスが「**LLM の眼**」という新しい役割を担う

第二部：エッジ AI ケーススタディ：予知保全

予知保全（Predictive Maintenance）

予防保全（Preventive Maintenance）

- 使用時間を基準に一律に部品交換
 - エンジンオイルの交換（半年毎）
 - ブレーキフルードの交換（2年毎）
 - タイミングベルト（100,000km 毎）
- 実際の消耗度は使用状況に大きく依存するので必要ない部品も交換対象に
 - 過剰整備 = コスト高
 - 過剰品質 = 競争力低下
 - 工場の稼働率低下、待ち時間

予知保全（Predictive Maintenance）

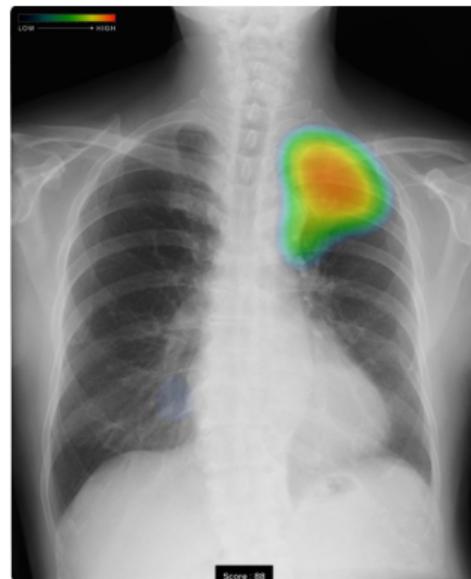
- 目に見えない予兆を捉えて部品交換
 - ブレーキパッド、タイヤ残量 → 可視
 - 振動パターン、電圧変動 → 不可視
- 必要な部品だけを必要な時に交換
- 整備入庫のタイミングの適切化
- 利用状況によって故障の兆候は様々、何から予兆を捉えるかの定義が課題（これまでは職人のカンに頼ってきた）

現在は「予防保全」が主流だが、「予知保全」が実現できればメリットは大きい

認識型 AI を活用した アノマリー検知

アノマリー = (まだ) 理論的には説明できていない兆候

- 認識型 AI を利用して **アノマリー (微小変化)** を検出する
- 1つの測定データからアノマリーを検出するのは困難
 - 大量の過去データとの比較からアノマリーを検出する
 - アノマリーは人間には判定出来ないような微細な特異値として出現する
- AI に **膨大な過去データを「学習」** させ、対象データが該当しているかの確率を「推論」する
- **認識型 AI によるアノマリー検出** の応用例
 - (医療) レントゲン写真画像から癌の兆候を捉える
 - (工業) 振動パターンからジェットエンジンの故障を予知



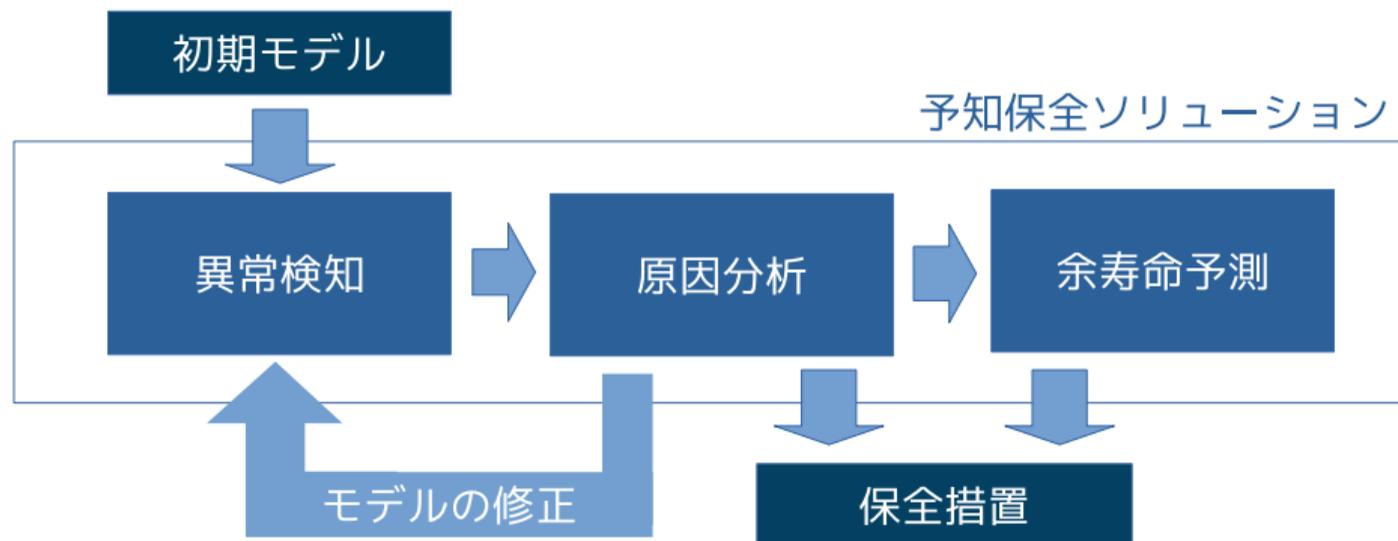
アノマリー検出はすでに「認識型 AI (eAI1.0) の得意領域」として確立されている

ルールベースの検出 vs. 認識型 AI によるアノマリー検出

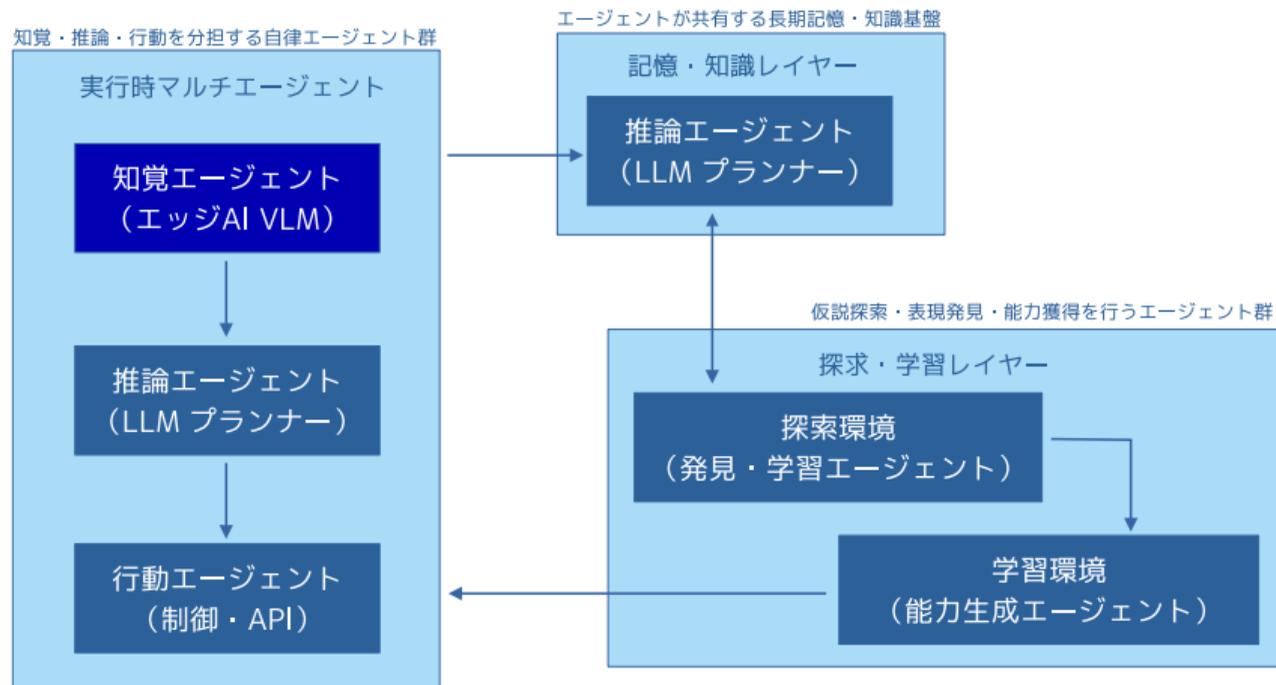
ルールベース（従来）	項目	AIベース（アノマリー検出）
人手で定義 （例：平均値・分散・温度閾値等）	特徴抽出	自動抽出 （CNN/DNN がパターンを自律学習）
if/then ルール・閾値比較	判断ロジック	統計的分布、潜在表現の異常度算出
正常／異常を人間が定義可能	前提	正常データのみから「正常性」を学習可能
単変量 or 数式ベース	検知方式	多次元・時系列・非線形関係を考慮
新たな条件に弱い（ルール再設計）	柔軟性	モデル再学習で自動対応が可能

認識型 AI のアノマリー検出は「データ分布そのものを学習」して正常／異常を識別

予知保全ソリューションの構成要素



予知保全 AI エージェント（イメージ）



第三部：AI を巡る共創開発のトレンド

なぜ今「AI のオープン性」が緊急課題になっているのか？

もし AI 技術が特定の企業や団体に独占されてしまったら....

- 生活のあらゆる場面を支える基盤技術インフラとなった AI が抱える課題
 - 学習データセット が不明
 - モデルウェイト がブラックボックス
 - ユーザーが AI が内在する **バイアス・知財・法的問題** を検証できない
- **説明責任** を果たすには **第三者による検証（＝モデル挙動の再現性）** が不可欠
- 主な AI 企業の **データセットやモデルの透明性** の現状は....
 - GPTx (OpenAI) 非公開、今後 Open Wait 化を表明
 - Lama2 (Meta) 非公開、利用はオープン（無償）だがモデルは非公開
 - DeepSeek (中国) 公開

今後 AI には「信頼性」だけでなく「第三者による検証性」の確保が求められる

クローズドな技術には（説明責任以外にも）多くのリスクがある

現在の AI は「交渉力の非対称性問題」が顕著である

- AI 処理用の HW については NVIDIA が事実上独占している状況
 - 各社が NVIDIA GPU の物量確保に躍起となっている
 - 中国は DeepSeek で NVIDIA GPU に依存しない AI を実現
- LLM モデル開発 では OpenAI (ChatGPT) が大きく先行している
 - Google (Gemini) が急迫している状況
 - OSS ベースの AI モデル性能が先行ソリューションに迫ってきている
- クローズドな AI が抱える 潜在リスク
 - サスティナビリティ（サービス終了、バージョンアップ強制、過去モデルの利用制約）
 - ビジネス性（API 価格の一方的改定、利用規約変更、データ持ち出し不可）
 - 政策・地政学（国際情勢による供給停止、規制・輸出管理の影響）

特定の先行企業による独占に対抗する新興勢力（含む OSS）が台頭している

オープン化の取り組み（1）

AI の中立性を確保を目指したコンソーシアム活動の例

■ LF AI & Data

- AI/ML/データ領域の OSS プロジェクトを統合、相互運用性のある AI 基盤を推進
- ONNX、PyTorch、Acumos、Kubeflow などを含む中立的なホスト
- AI の倫理性・透明性・ガバナンスに注力、OSS による AI エコシステム発展を支援
- <https://lfaidata.foundation/>

■ UXL (Unified Acceleration Foundation)

- Intel・Arm・Qualcomm らが中心となり、AI/データ処理向け共通プログラミングモデル (oneAPI 互換) を推進
- LF AI & Data 等との連携を視野に、AI 推論や HPC 統合の業界標準化を目指す
- <https://uxlfoundation.org/>

オープンな AI モデル、汎用性のある AI プログラミング言語を目指して連携している

オープン化の取り組み（2）：MCP が Linux Foundation 傘下に

ANNOUNCEMENT

Linux Foundation
Announces the
Formation of the
Agentic AI Foundation

Anchored by New Project
Contributions including MCP,
goose, and AGENTS.md



**Agentic AI
Foundation**

<https://www.linuxfoundation.org/press/linux-foundation-announces-the-formation-of-the-agentic-ai-foundation>

何故多くの企業が **共創開発**（=オープンな R&D 活動）に取り組むのか

多くの先端企業が「make or buy」から「make or buy or collaborate」に転換

- make 戦略 = 差異化技術については自社内でクローズに技術開発
- buy 戦略 = 非差異化領域の技術、時間が求められる技術は社外から技術導入
- collaborate 戦略 = 他社と連携して差異化技術のベースを共同開発
 - 開発規模が大きく自社だけでは必要なリソース（資金、開発者等）を集められない
 - オープンな活動とすることで、技術の中立性担保や開発スピードアップを図る
 - 企業間の連携による マジョリティ確保でデファクトを獲得する
 - 中長期的にメンテナンス可能な サスティナブルなソリューションの構築
 - 企業同士で 似て非なる技術を重複開発するリスクを回避する

技術進化の速度、規模の拡大、グローバル化のトレンドを反映した技術の新潮流である

コンソーシアム活動内での 振る舞いの作法

自社利害だけを声高に主張するのは禁じ手 である

- コンソーシアムへ参加する各企業の代表は、自社の利益代表として振る舞うのではなく **コンソーシアム全体としての成果を上げることを第一義に考えるべき** である
- 特に参加企業が **自社内で開発した成果をコンソーシアムに持ち込む場合** には
 - 提供した技術が他のメンバー企業に使ってもらえるような **丁寧なサポート（ドキュメント等の整備、Q&A 対応等）** を行う
 - **開発ロードマップの共有** と、コンソーシアム内での **開発のリーダーシップ** を取る（コードの投げっぱなしはダメ、それでは何も起こらない）
 - サイバーセキュリティ対応など **必要な SW メンテナンス** を長期的に提供する
- 車輪の再発明を回避するために、**既存の類似の活動（OSS コミュニティ、他の産業コンソーシアム、技術標準化活動）との整合** を図るべきである

活動が成功するか否かはコンソーシアムのリーダーの指導力・調整力に大きく依存する

本日のまとめ

- 画像の中にある **物体の識別** に特化した **CNN アルゴリズム** を利用した「**認識型 AI**」は自動車の運転支援システムなど既に多くのソリューションで採用実績がある。エッジ AI とは認識型 AI の **推論処理** を組み込み機器に搭載したもので、**プライバシー確保** や **ネット非依存** などの **メリット** が注目されている。
- 2025 年には ChatGPT に代表される **大規模言語モデル (LLM)** を利用した「**理解型の AI**」が登場し、誰でも **普通の文章** 使って AI の力を利用するようになった。更に LLM の発展系として **画像の意味を理解** する **VLM** や、**画像からとるべき行動** をアドバイスする **VLA** といった新しい AI アルゴリズムも登場している。
- 役割の異なる複数の AI が自律的に連携する「**AI エージェント**」の実用化が進んでおり、AI エージェント同士がやりとりするための **MCP** や **A2A プロトコル** が定義された。画像認識系のエッジ AI は高速で高精度な「**LLM の目**」の役割を持った認識エージェントとして再定義され、高度な AI 処理の実現に極めて重要な役割を担う。
- エッジ AI の応用例として **予知保全** への適用を考察した。予知保全では **故障の兆候** となる **アノマリーの検出** が課題であったが、AI エージェントの活用によって高精度で故障を予知できる **可能性** がある。
- AI の適用範囲が大きく拡大に伴って **AI 関連技術のオープン性** の重要性が注目されている。先行する特定の企業に AI 関連の基盤技術を独占支配されないように、**OSS ベースの AI 技術基盤の整備** が進められている。