7th Grand Canvas : AI品質の未来を共に描く
～AI品質マネジメントネットワーキングシンポジウム～

https://www.digiarc.aist.go.jp/event/7th_grand_canvas/

IBMのAIエージェントとLLMのISO/IEC 42001 認証

2026 年 2月 3日

日本IBM エコシステムテクニカルリーダーシップ
平山 毅　HIRATSU@jp.ibm.com

IBM

7th Grand Canvas：AI品質の未来を共に描く
~ AI品質マネジメントネットワーキングシンポジウム ~

開催日時：2026年2月3日(火) 13:00-17:30
開催場所：AP日本橋 6F Room G
開催形式：ハイブリッド形式(会場参加／オンライン参加)
参 加 費：無料

主催：国立研究開発法人産業技術総合研究所
共催：国立研究開発法人 新エネルギー・産業技術総合開発機構(NEDO)
後援：AIセーフティ・インスティテュート(AISI)

産総研 ともに挑む。つぎを創る。
NEDO
AISI Japan AI Safety Institute

15:30–17:25

一般募集した企業登壇者によるショートトークセッション

- ファシリテータ：妹尾 義樹（国立研究開発法人産業技術総合研究所）
- コメンテータ：三島 浩一 氏 (三菱電機株式会社)
- 応募登壇者：（1件5分から15分の講演：登壇者募集中）
  - 松木 晋祐 氏（株式会社ベリサーブ）
  - 小島 潤 氏（EY新日本有限責任監査法人）
  - 小川 隆一 氏（情報処理推進機構）
  - 小尾 和美 氏（株式会社キャリア・マム）
  - 藤井 涼 氏（株式会社AI共創総研）
  - 多賀 太 氏（株式会社AItegrity）
  - 仲津 由希子 氏 （株式会社電通総研）
  - 平山 毅 氏（日本アイ・ビー・エム株式会社）

自己紹介

日本IBM株式会社　テクノロジー事業本部 エコシステム共創本部
エコシステムテクニカルリーダーシップ　平山 毅
https://www.linkedin.com/in/tsuyoshihirayama/

テクノロジー事業本部エコシステム共創本部でエコシステムテクニカルリーダーシップを担当。

IBMでは、クラウド事業、Red Hat アライアンス事業、Data AI事業、クライアントエンジニアリング事業金融部門、直近まではエコシステムエンジニアリングの日本での立ち上げをリード。それ以前は、

アマゾンウェブサービスでソリューションアーキテクト、プロフェッショナルコンサルタント、

野村総合研究所でファイナンシャルエンジニア、

東京証券取引所でITサービスのマネージャー、経営企画、ITアーキテクト、金融派生商品の開発者。

Linux Foundationの日本チャプターのリード、情報処理学会デジタルプラクティス編集委員、

総務省Beyond 5G新経営事務局委員、（官民学連携）。左記の通り。技術書籍を8冊出版。

国際学会、国際カンファレンスなど登壇多数。金融やITの認定資格多数保有。

神奈川大学、早稲田大学大学院、事業構想大学院大学、東京都立産業技術大学院大学で講義も担当。

東京都立日比谷高等学校卒業、東京理科大学理工学部卒業、

早稲田大学大学院経営管理研究科MBAファイナンス修了。

北陸先端科学技術大学院大学、長岡技術科学大学の博士課程在学。

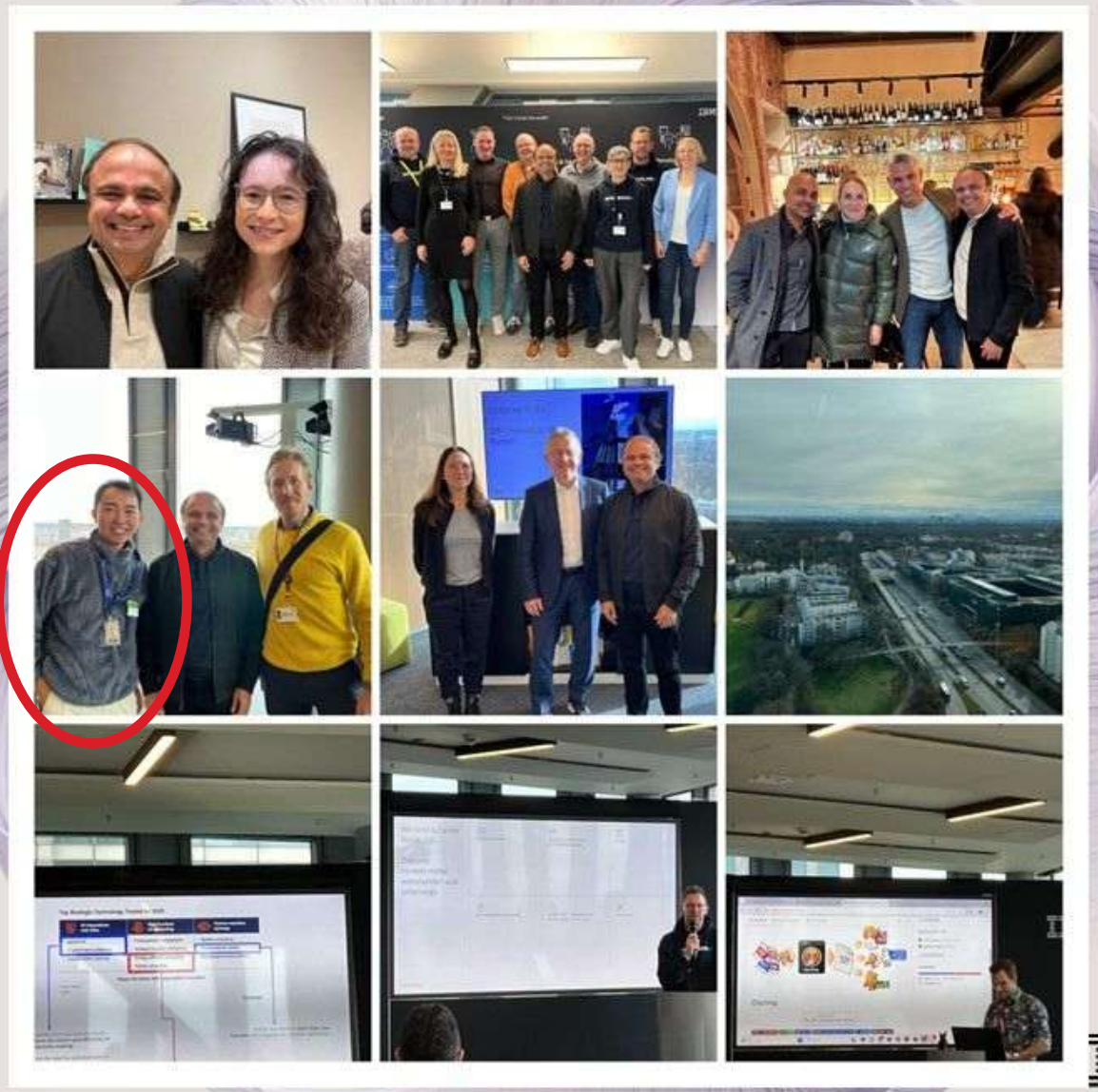
NEDO AI品質マネジメント講座5期生。Beyond 5G新経営戦略センター リーダーズフォーラム4期生。

前回 6th Grand Canvas オープンコミュニティで進める品質とガバナンス を講演

watsonx Build Partner insights from Amsterdam & Munich

~

* Specialized Models
* Holistic Governance
* Enterprise-scale RAG

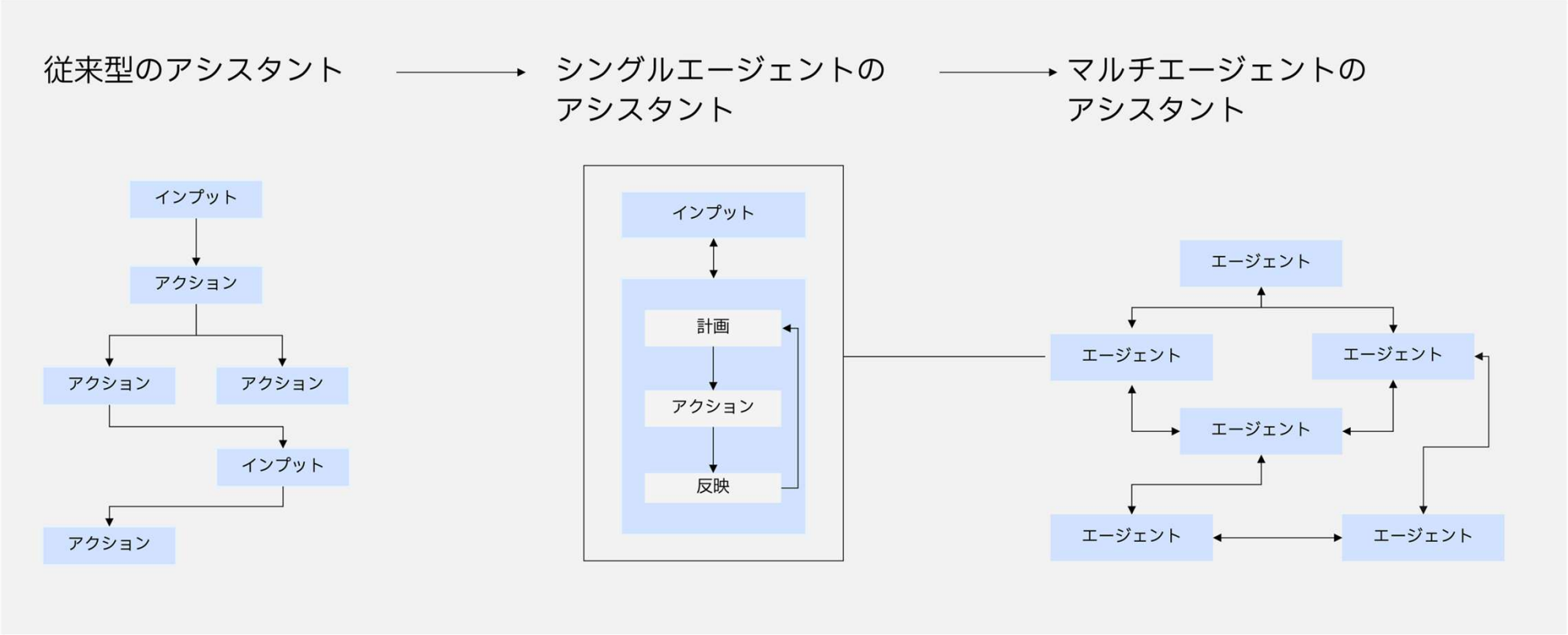


https://www.linkedin.com/posts/omkar-nimbalkar-741b423_watsonx-ai-automotive-activity-7266461148374347777-7Xf1?utm_source=share&utm_medium=member_desktop

IBM

AIエージェントとは

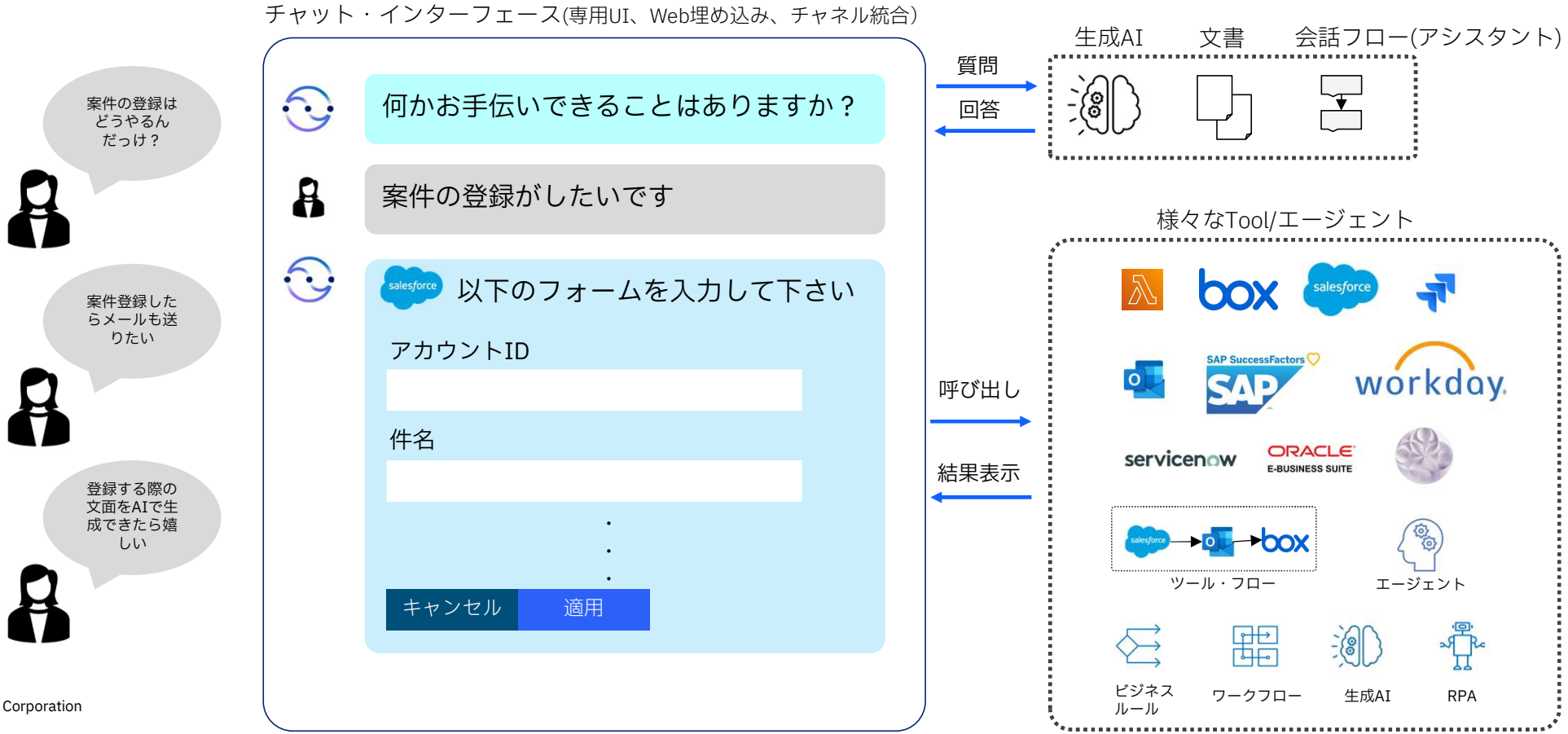
従来型アシスタントはルールベースで単一のタスクをこなすのに対し、**AIエージェントは与えられたインプットからどのような作業をすればよいかを自律的に計画し実行する。**将来的にはマルチエージェントで相互にやり取りする姿へ発展



出典: https://www.ibm.com/jp-ja/think/topics/ai-agents

watsonx Orchestrate

watsonx OrchestrateはAIエージェント/アシスタント構築のためのプラットフォームです。生成AIやRAGによる回答生成から、Toolの呼び出し、複数Toolを組み合わせた業務シナリオの自動化を可能にします。また、ワークフローや、ビジネス・ルール、RPAといった生成AIを補完する機能も含めて単一のプラットフォームとして提供し、生成AIを活用した業務の自動化を迅速に実現します。



watsonx Orchestrateの特徴

事前定義されたAgentやToolの活用

短期間で導入可能

- 人事、調達、販売の各領域で利用可能な定義済みの100個以上のAgentと400個以上のToolを提供
- カタログに登録されたAgentやToolを組み合わせて業務ユーザーでもローコードでクイックにAgentを構築可能

単なる生成AIワークフローではカバーできない

業務利用に必要な機能を提供

- 生成AIだけでは対応が難しい、ルール・ベースの処理や、事前定義された会話フローなどを柔軟に組み合わせて処理を実装可能
- 様々な認証方式やSSO、外部チャネル接続に対応

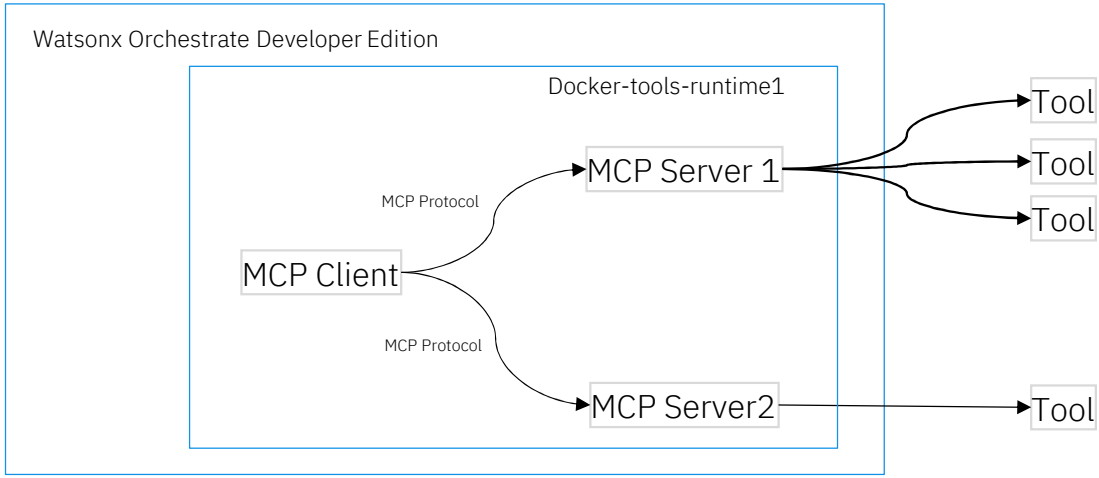
開発キットや標準対応による

高い開発生産性と拡張性

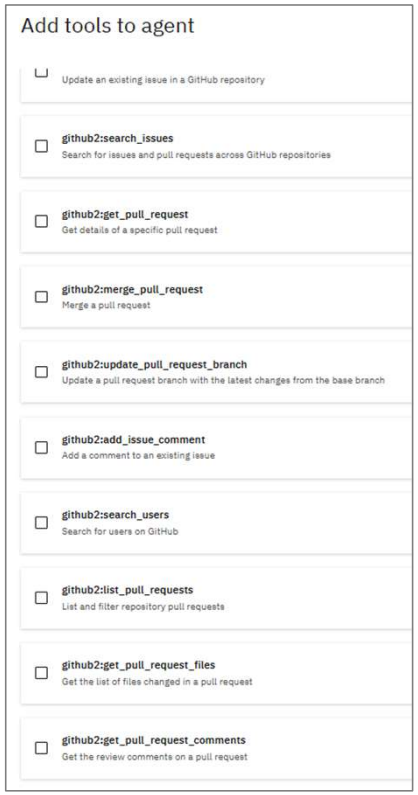
- 開発者はADK（Agent Development Kit）とPC上で動作するwatsonx Orchestrate Developer Editionを組み合わせ開発が可能。
- MCP(Model Context Protocol)等の標準プロトコルに対応し、外部のToolやAgentと柔軟に連携が可能。
- 外部LLMの使用やモニタリング機能も提供。

watsonx OrchestrateにおけるMCP(Model Context Protocol)対応

MCPサーバーをコマンドでインストールすることにより、watsonx Orchestrate上で動作させ、複数のToolを
MCPサーバー経由で利用することが可能になります。



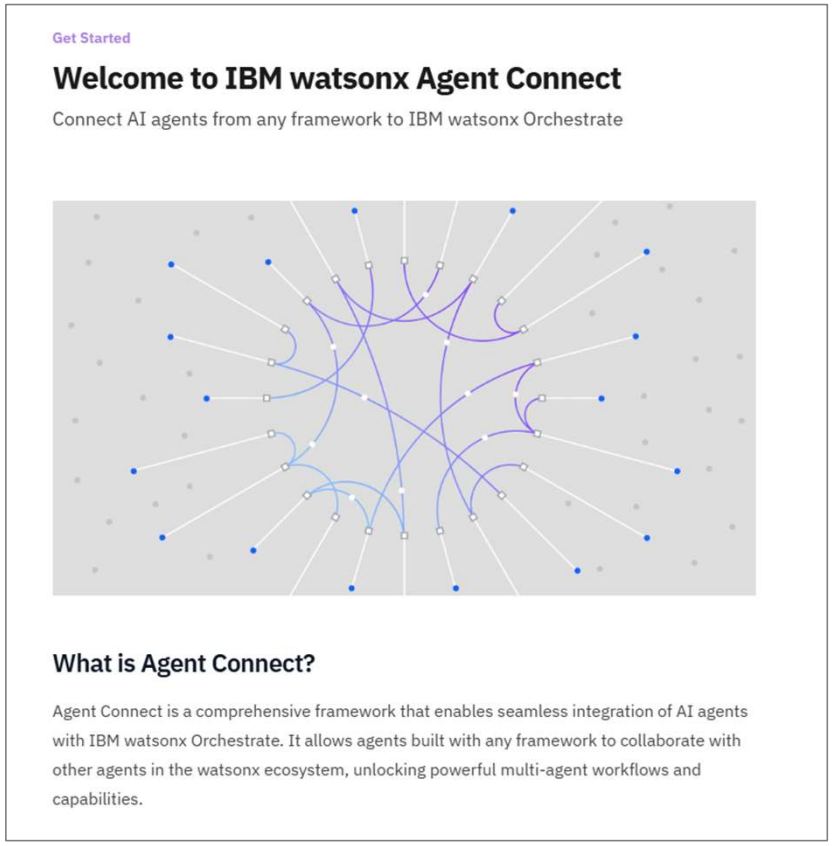
※現時点では、node/pythonのStanddard I/Oのみサポート、
今後リモートのMCPServerへの対応、watsonx Orchestrate上で作成したToolのMCP経由での公開なども予定



GitHub MCP Serveと連携した場合の例

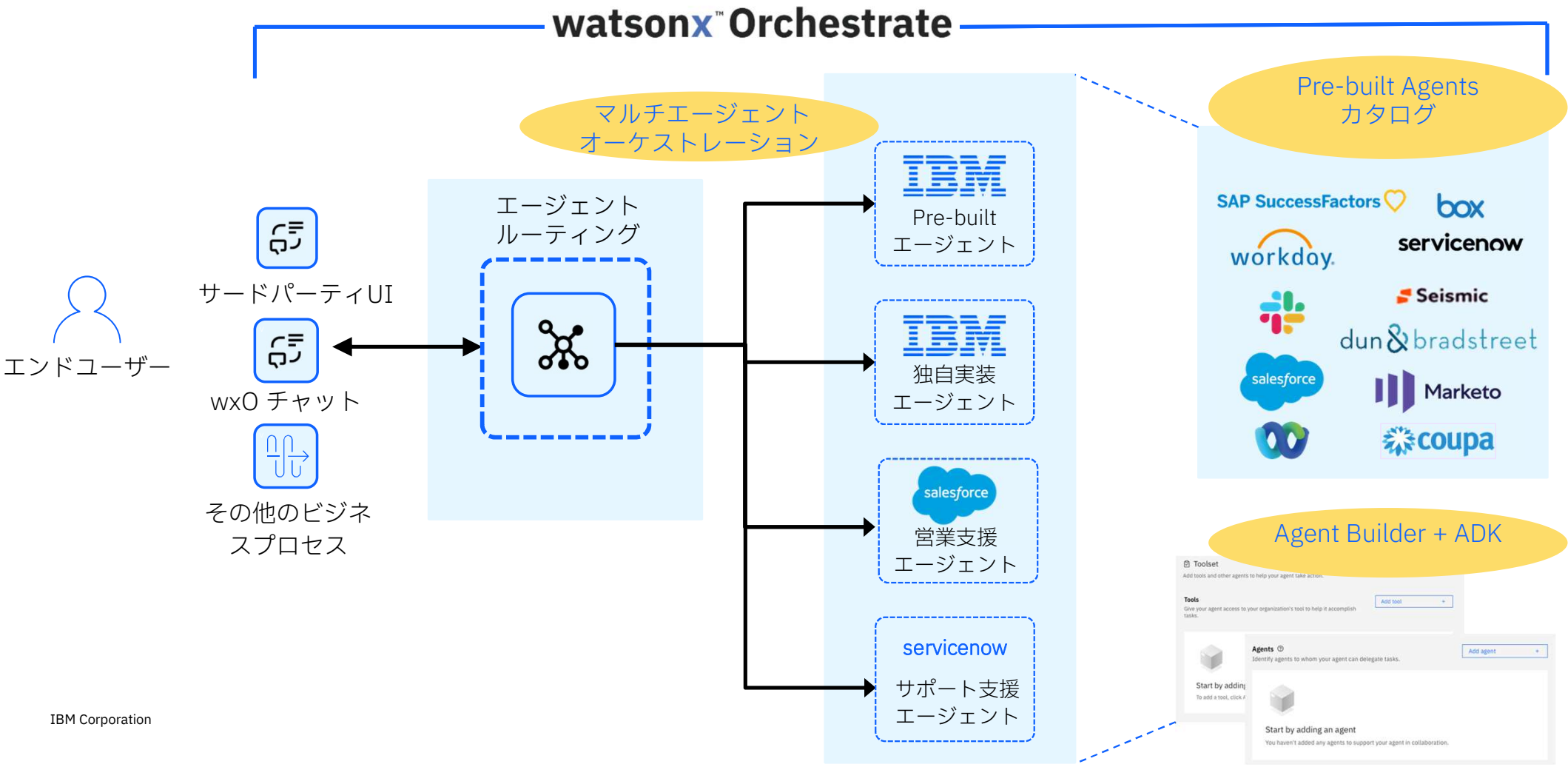
IBM Agent Connect

AI Agentのエコシステム
Agent実装に依存しないAgent間の連携フレームワーク(ACF)、パートナー・プログラムを提供



box
lyzr
salesforce
Seismic
servicenow
SIRION
causaLens
(h[s])
HYPERSCIENCE
MavenAGI
netomi
symplistic.ai

マルチエージェント・オーケストレーションの実現



OSSベースのIBM独自の大規模言語モデルGranite

「デコーダー」アーキテクチャを採用した**IBM独自の大規模言語モデル**（LLM）として、Granite.13b.instructおよびGranite.13b.chatを2023年9月に提供開始

- 要約、質問応答、分類などビジネス領域のタスクで優れた性能を発揮しつつ、コンテンツ生成、洞察抽出、RAG（Retrieval-Augmented Generation）、固有表現抽出といった他の自然言語処理（NLP）タスクもサポート

- V100-32GBのシングルGPUにフィットし、**他社の巨大なモデルに比べて効率的**であり、**環境負荷を抑えることが可能**

- **ビジネス・ユーザーのニーズをターゲット**として以下の領域のデータを使用してモデルを学習
  - **インターネット：**インターネットから取得した一般的な非構造化言語データ
  - **学術：**科学技術に特化した技術的な非構造化言語データ
  - **コード：**さまざまなプログラミング言語をカバーする非構造化コードデータ
  - **法務:** 法律意見書やその他の公的提出書類から取得した企業関連の非構造化言語データ
  - **財務：**一般に公開された財務文書や報告書から取得した企業関連の非構造化言語データ

IBM独自基盤モデル 技術仕様の公開

Graniteモデルの公開にあわせ技術仕様に関する詳細を公開し、
透明性と責任あるAIへのコミットメントを示す

オリジナル版：https://ibm.biz/techpaper-granite-13b
和訳版：https://ibm.biz/techpaper-jp-granite-13b

公開している技術仕様

■ データソース
- 学習データの一覧

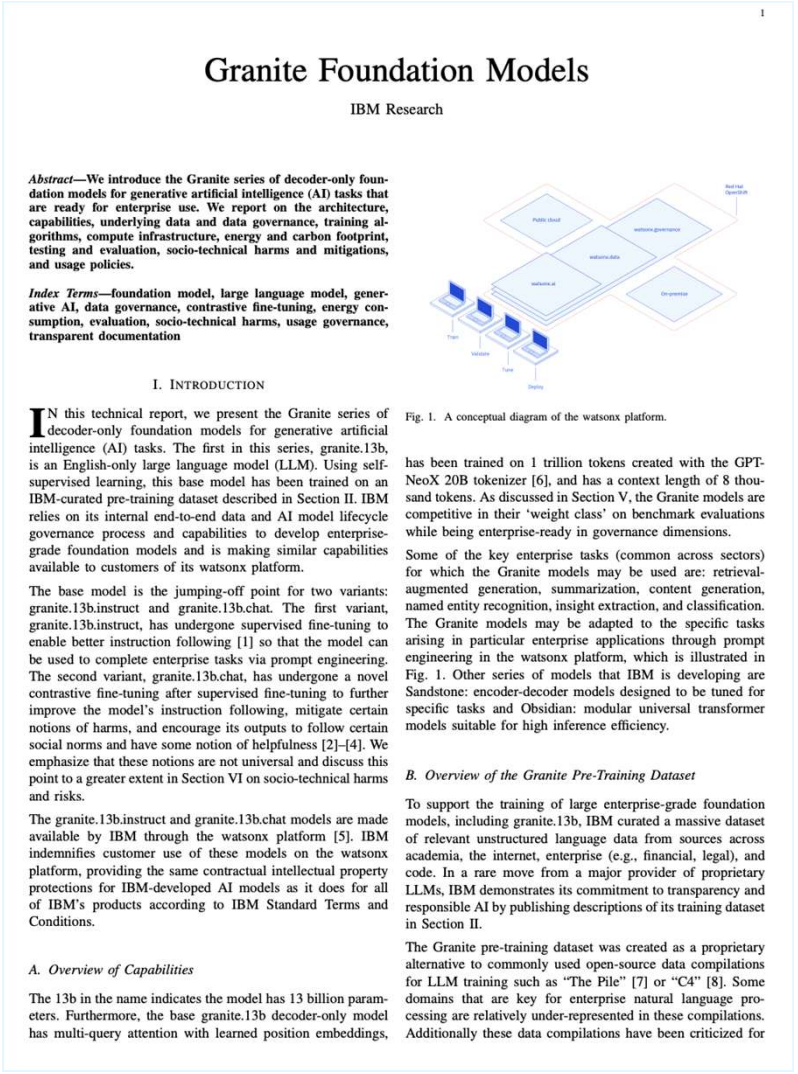
■ データ・ガバナンス
- データのクリアランスと収集
- 前処理パイプライン
- トークナイゼーション

■ 学習
- アルゴリズムの詳細
- 計算
- エネルギー消費と炭素排出量

■ テストと評価
- 基盤モデルの評価フレームワーク
- Granite モデルの評価とベンチマーク

■ 社会技術的弊害とリスク

■ 利用ポリシー



1

Granite Foundation Models

IBM Research

***Abstract*—We introduce the Granite series of decoder-only foundation models for generative artificial intelligence (AI) tasks that are ready for enterprise use. We report on the architecture, capabilities, underlying data and data governance, training algorithms, compute infrastructure, energy and carbon footprint, testing and evaluation, socio-technical harms and mitigations, and usage policies.**

***Index Terms*—foundation model, large language model, generative AI, data governance, contrastive fine-tuning, energy consumption, evaluation, socio-technical harms, usage governance, transparent documentation**

I. Introduction

IN this technical report, we present the Granite series of decoder-only foundation models for generative artificial intelligence (AI) tasks. The first in this series, granite.13b, is an English-only large language model (LLM). Using self-supervised learning, this base model has been trained on an IBM-curated pre-training dataset described in Section II. IBM relies on its internal end-to-end data and AI model lifecycle governance process and capabilities to develop enterprise-grade foundation models and is making similar capabilities available to customers of its watsonx platform.

The base model is the jumping-off point for two variants: granite.13b.instruct and granite.13b.chat. The first variant, granite.13b.instruct, has undergone supervised fine-tuning to enable better instruction following [1] so that the model can be used to complete enterprise tasks via prompt engineering. The second variant, granite.13b.chat, has undergone a novel contrastive fine-tuning after supervised fine-tuning to further improve the model's instruction following, mitigate certain notions of harms, and encourage its outputs to follow certain social norms and have some notion of helpfulness [2]–[4]. We emphasize that these notions are not universal and discuss this point to a greater extent in Section VI on socio-technical harms and risks.

The granite.13b.instruct and granite.13b.chat models are made available by IBM through the watsonx platform [5]. IBM indemnifies customer use of these models on the watsonx platform, providing the same contractual intellectual property protections for IBM-developed AI models as it does for all of IBM's products according to IBM Standard Terms and Conditions.

A. Overview of Capabilities

The 13b in the name indicates the model has 13 billion parameters. Furthermore, the base granite.13b decoder-only model has multi-query attention with learned position embeddings,

Fig. 1. A conceptual diagram of the watsonx platform.

has been trained on 1 trillion tokens created with the GPT-NeoX 20B tokenizer [6], and has a context length of 8 thousand tokens. As discussed in Section V, the Granite models are competitive in their 'weight class' on benchmark evaluations while being enterprise-ready in governance dimensions.

Some of the key enterprise tasks (common across sectors) for which the Granite models may be used are: retrieval-augmented generation, summarization, content generation, named entity recognition, insight extraction, and classification. The Granite models may be adapted to the specific tasks arising in particular enterprise applications through prompt engineering in the watsonx platform, which is illustrated in Fig. 1. Other series of models that IBM is developing are Sandstone: encoder-decoder models designed to be tuned for specific tasks and Obsidian: modular universal transformer models suitable for high inference efficiency.

B. Overview of the Granite Pre-Training Dataset

To support the training of large enterprise-grade foundation models, including granite.13b, IBM curated a massive dataset of relevant unstructured language data from sources across academia, the internet, enterprise (e.g., financial, legal), and code. In a rare move from a major provider of proprietary LLMs, IBM demonstrates its commitment to transparency and responsible AI by publishing descriptions of its training dataset in Section II.

The Granite pre-training dataset was created as a proprietary alternative to commonly used open-source data compilations for LLM training such as "The Pile" [7] or "C4" [8]. Some domains that are key for enterprise natural language processing are relatively under-represented in these compilations. Additionally these data compilations have been criticized for

IBM独自基盤モデル Granite のISO/IEC 42001認定

ISO/IEC 42001:2023:
組織がAIに関するポリシー、プロセス、および安全策を確立する方法の指針として策定。

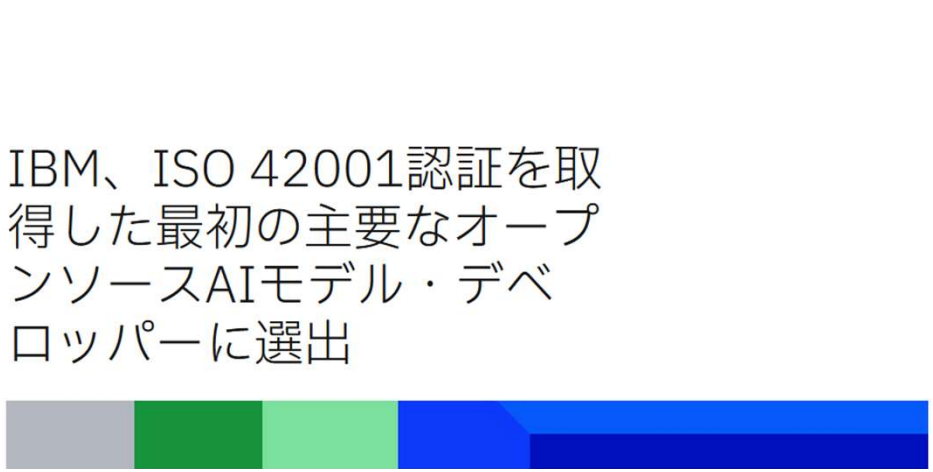
（AIが責任を持って構築、デプロイ、維持されているかを外部から監査するための、
世界的に認められた初のフレームワークとして誕生。）

※IBM GraniteのAI管理システム（AIMS）は、ISO/IEC 42001:2023に基づく認定を取得。
↓
1.安全で責任あるAIのための国際的に認められたベスト・プラクティスに沿っている
2.AI管理プロセスが最高レベルの精査を満たしている

標準が責任あるAIが、AIモデル開発者にとって、実際にどのようなものであるかを示し明瞭化。

また、IBMにとっては、AIモデルの開示に関する業界トップの透明性に加え、AIのセキュリティー、安全性、ガバナンスの実践に対する長年の投資の妥当性を検証。

https://www.ibm.com/jp-ja/new/announcements/ibm-granite-iso-42001



ISO認証による信頼と企業向けへの生成AIエコシステム適用

信頼されたLLMは企業向け適用に有効

↓

企業向けにおいては、ベースのLLMに加え、業界ごとのSLM、日本語LLMも組み合わせたマルチモーダルも有効活用。

↓

この信頼できる基盤モデル群をベースにして、AIエージェントに活用。

エコシステム型で汎用的に導入へ

https://www.ibm.com/jp-ja/think/insights/banking-financial-dx-consulting-series-12-genai-for-financial-industry

Let‘s Collaborate AI Project with IBM
Let’s Try watsonx orchestrate for AI Agent
on Granite certified with ISO/IEC 42001