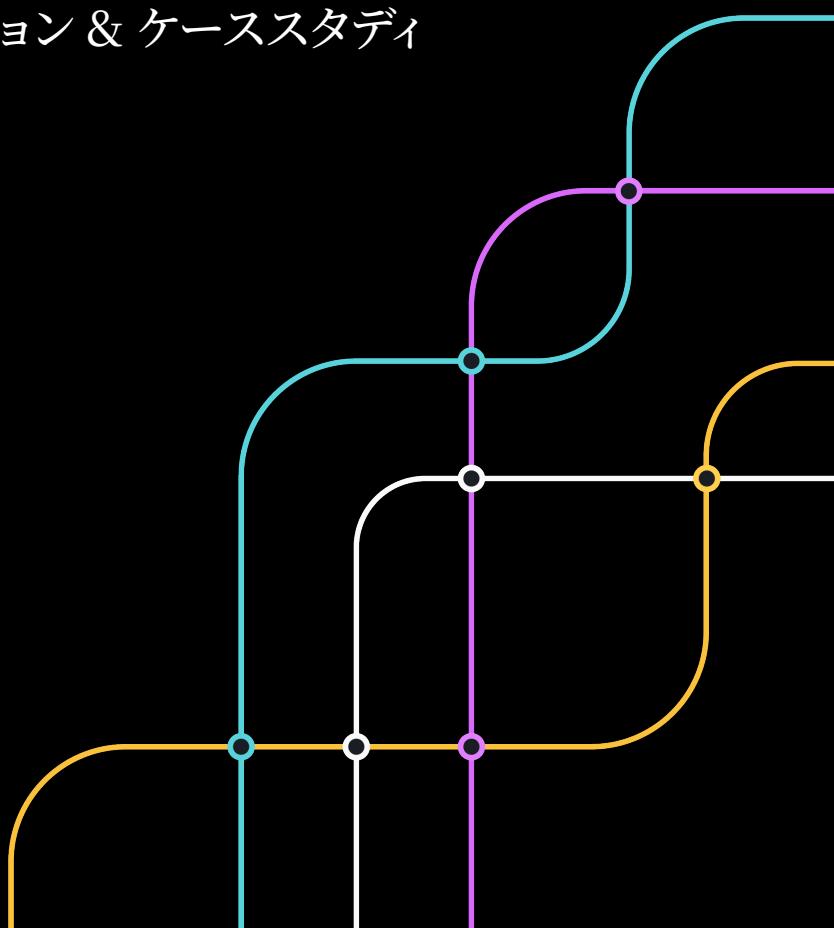




# 世界で信頼される MLOps / LLMOps プラットフォーム

ソリューション & ケーススタディ



# Weights & Biases:

MLOps/LLMopsプラットフォーム

Weights & Biasesは、GenAI業界を初め、あらゆるAI開発を支える実務者向け開発プラットフォームです。独自の基盤モデルの構築、既存 モデルの継続学習・ファインチューニング、LLM APIを使ったアプリケーションのデプロイまで、一貫して開発を管理。最先端のLLMを活用したユースケースを実現します。

世界をリードするAIチームがWeights & Biasesを信頼しています



W&Bの最新情報を発信! ぜひご登録ください



W&B YouTube



W&B Note



## W&B Models

W&B Modelsは、機械学習研究者・開発者がモデルを構築する上で避けては通れない実験的なプロセスを、もれなくトラッキングすることで、開発の再現性を担保し、より精度の高い結果に少ない工数で到達することを可能にします。

開発過程で使われたデータセットや出力されたモデルなど、あらゆるアーティファクトの関係性を記録し、全社レベルで管理するためのレジストリや、評価ジョブなどを自動化し、GPUキューに展開できるローンチ機能などを使い、開発からデプロイまでの核となるプロセスを大規模にスケールすることを可能にするMLOpsプラットフォームです。

## W&B Weave

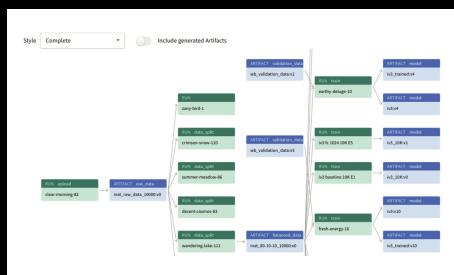
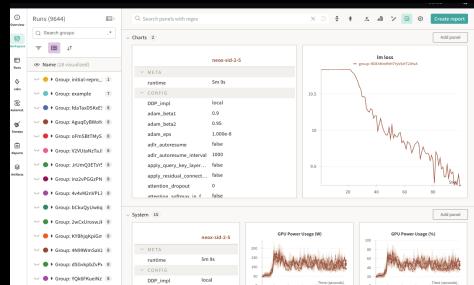
W&B Weaveは、APIなどで提供される既存の大規模言語モデル(LLM)をベースにしたAIアプリケーション開発におけるMLOpsを提供する、新しい製品です。検索拡張生成AI(RAG)などのように、プロンプト、データベース、エージェントなどを複雑に組み合わせたプロセスを可視化し、開発工数を削減します。

また、AIアプリケーションの評価手法を定義し、その結果を容易に比較可能にすることで、性能を量化します。ハイパーパラメーターの調整を通じて、より短時間で質の高いアプリケーションの構築を可能にするだけでなく、デプロイ後も継続的にモニタリングを行い、トレーサビリティーの高い運用を可能にします。

# WandB Modelsを使ったMLOpsの構築

## 基盤モデル開発・ ファインチューニングの実験管理

WandB Modelsを使用すると、LLMモデル学習の実験の進行状況や評価結果、GPUリソースの使用状況をリアルタイムに追跡・分析できます。大規模な分散学習や、複数の計算基盤を使った開発でも、一元化されたダッシュボードで容易に実験管理が可能になり、機械学習エンジニアの生産性を大幅に向上します。



## 学習データ、コード、モデルの関連 を見失わないリネージ追跡

Artifactsを使用することで、実験間の関係性を追跡できます。例えば多数の試行錯誤が発生するLLMファインチューニングにおいて、それぞれのモデルの学習に使用されたデータセットや前処理を一目で確認することができます。これにより、大規模な開発プロジェクトでも高い再現性とトレーサビリティを実現します。

## 画像タスクからの インサイドの抽出

Modelsは、マルチメディア対応テーブル機能を備え、深層学習の結果から迅速にインサイトを得ることができます。例えば、セグメンテーションや物体検出の予測結果と実測値をインタラクティブに可視化し、データの特定属性や予測性の低いサンプルをすぐに特定できるため、改善策の立案が容易になります。

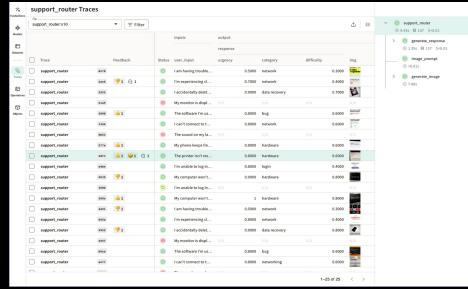


# WandB Weaveを使ったLLM Opsの構築

TraceによるLLMアプリケーションの全ての入出力の可視化

WeaveのTrace機能を使用すると、LLMアプリケーション開発の全プロセスを詳細に記録し、可視化することができます。API呼び出しから中間処理、最終出力まで、開発フレーム全体を追跡し、各ステップの入出力データや実行時間を自動的にキャプチャします。これにより、ボトルネックを特定し、効率的にデバッグを行うことを可能にすることで、Traceは迅速な問題解決を実現します。

Traceは迅速な問題解決と最適化を支援します。



## Models/Datasetsによるより体系的・構造的なLLMの追跡と管理

WeaveのModels/Datasets機能は、LLMのモデルとデータセットの保存と読み出し、バージョン管理を効率化します。LLMを構成するこれらの要素をより構造的・体系的に追跡・管理することで、イテレーションを加速させ、より高品質なアプリケーションの開発を可能にします。

```
extract.first_person_name.v1
name extract_first_person_name [2 versions]
version
Calls 1 calls

Code Use

v0 v1

1 import weave
2 import openai
3
4 #weave.openapi_key = "your_openai_api_key"
5
6 doc = weave.Definition(first_person_name_decl)
7 client = openai.Client()
8
9 prompt_template = "What is first person's name in the following text?"
```

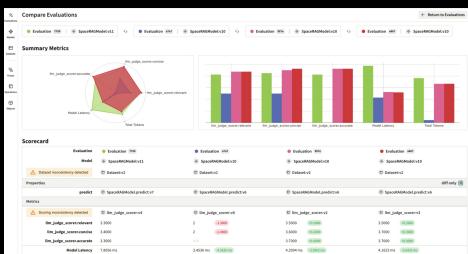
```
1 import weave
2 import openai
3
4 #weave.openapi_key = "your_openai_api_key"
5
6 doc = weave.Definition(first_person_name_decl)
7 client = openai.Client()
8
9 prompt_template = "What is first person's last name in the following text?"
```

```
response = client.chat_completions.create(
    model="gpt-3.5-turbo",
    messages=[{"role": "user", "content": prompt_template},
              {"role": "assistant", "content": ""}]
)
messages[1].role = "user"
messages[1].content = prompt_template
```

```
return response.choices[0].message.content
```

# EvaluationsによるLLM性能評価の自動化

WeaveのEvaluations機能は、LLMアプリケーションの性能評価を自動化し、継続的な改善を支援します。カスタム評価指標やスコアリング関数を定義し、モデルの出力を多角的に分析することができます。評価結果は自動的に記録され、バージョン間で比較可能な形式で保存されます。これにより、モデルの改善点を容易に特定し、データドリブンな意思決定を行うことが可能になります。





# 生成AIとW&Bで実現する 次世代の購入体験

Teo Narboneta Zosa 氏 | メルカリ

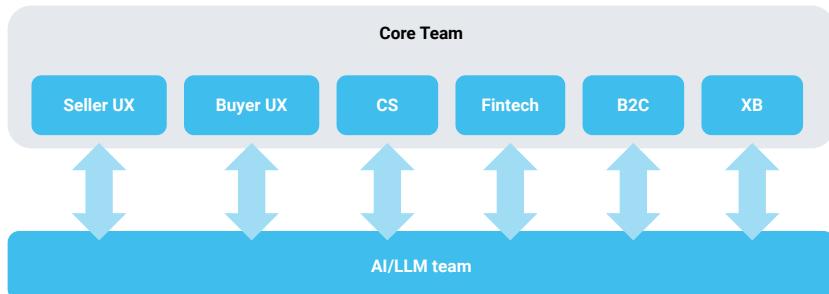
## メルカリでは専門チームによる 生成 AI 活用が急速に進む

「CtoC（個人間売買）のマーケットプレイスとして日本最大規模を誇るメルカリは、月間アクティブユーザーは約 2400 万人、年間総売上は 1 兆円を超え、数十億件もの商品リストが掲載されている」。ゾサ氏は冒頭、そう語ります。同アプリは 2013 年にスタート。近年では生成 AI を活用して、次世代の顧客体験を創出しようという取り組みが急速に進んでいるといいます。

2023 年 5 月、メルカリは AI と LLM（大規模言語モデル）の専門チームを設置。チームは、メルカリ市場に価値を追加することを目的として設立され、マーケットプレイスのコア機能に関わる他のチームと密接に連携し、生産性向上のため尽力しています。

その成果には、LLM を活用した「メルカリ AI アシスト」も含まれます。同機能は、利用者が出品物を撮影したあとにカテゴリーを選ぶと、AI が商品説明や状態、販売価格などの必要な情報を自動入力するというもの。内容を確認して、

# Getting Generative AI to Production



Planning, model selection, prompt engineering, product implementation, etc.

9

AI/LLM 専門チームは、マーケットプレイスのコアチームと密接に連携し、LLM を活用した機能「メルカリ AI アシスト」他、メルカリに様々な AI 機能を追加している

最終的なリストをタップするだけで商品が出品されるため、ユーザーの手間は大幅に軽減されます。

## W&B Models: 類似商品の推薦機能開発を W&B が加速

“画像から ML 機能を活用し、価値を生み出す”機能の実現のためには、数十億のエントリーを持つ広範な自社の製品データに合わせてファインチューニングされた埋め込みモデルの開発が必要だったとゾサ氏。言語と画像を結びつける CLIP の派生 AI モデル・SigLIP（シグリップ）を用いてチューニングを始めましたが「プロジェクトの成功を支えたのが、Weights & Biases の存在だった」といいます。

メルカリは、統一された埋め込みモデル実現の

ために WandB を使い、精度の高い推薦結果の実現を目指しました。ユーザーの検索行動を分析し、AI が正確に関連性の高い商品を提案できるよう、反復実験を繰り返しました。「数千のモデル、トレーニングラン、数百万のメトリクスの反復を、わずか 2 人のエンジニアが数週間で達成した。これはすべて WandB が追跡を容易にし、進行させること可能にしてくれたおかげです」

## W&B Weave: 検索精度向上への LLM の活用を支援

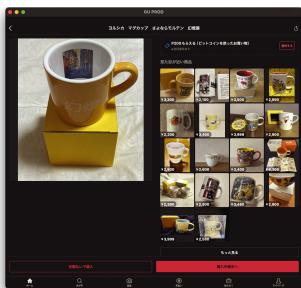
LLM を使うことで、検索精度に関しては、オープンソースの全文検索エンジン「ElasticSearch（エラスティックサーチ）」などの、従来型のテキストベースの検索エンジンでは対応しきれなかったクエリ（検索語句）や、ブランド名のあ

# Fine-tuned Image Embeddings

Solution/解決方法



MobileNet  
(Control)



Fine-tuned SigLIP

32

スマートフォンから写真を読み込むと、メルカリにある同商品や近い商品を検索するもので、これによりユーザーは求め る商品を容易に見つけられるようになった

いまいな表記にも対応できるようになったとゾサ氏は言います。

この開発の中では、WandB の LLM 評価・分析ツール「Weave」が大きな役割を果たしたといいます。文字通りたった 3 行のコードで、入出力のログ記録や可視化が可能になり、従来の構造化 JSON ログよりも格段に使いやすい環境を実現。クライアントのバージョン管理や、改善・改悪の追跡も容易になりました。

具体的には、与えられたクエリと商品画像に対する関連性スコアを算出するプロンプトの構築で Weave を活用しました。関連性スコアは「完全に関連がある」から「代替品として許容できる」、「補完的な商品」、そして「まったく関連がない」までの範囲で算出されます。例えば、iPhone を検索した時に iPhone アクセサリーが

出てくるようなケースが挙げられます。これには LLM にチェーン・オブ・ソートプロンプトやインターネット検索を組み合わせるアプローチを用いましたが、そのトラッキングと評価が課題でした。

Weave の評価機能を使用することで、数百件の事例に対する関連性ラベルの精度や、多岐にわたるスコアリング機能の効果を、レーダーチャートやヒストグラムを通じて視覚的に確認することが可能になりました。「プロンプトの微細な変更、例えばマークダウン形式の使用有無による精度への影響なども、容易に検証できるようになった」とゾサ氏は説明します。

メルカリは今後も生成 AI を使った革新を進め、より多くの価値をユーザーに提供し続けるゾサ氏は宣言します。

The screenshot shows the Tealfeed interface. On the left, a sidebar lists 'Projects' and 'Traces'. Under 'Traces', there's a section for 'run\_llm\_eval\_pipeline' with 100+ entries. A specific trace entry is selected, showing details like 'User', 'Called', 'Tokens', 'Cost', 'Latency', and 'Status'. On the right, a detailed view of the selected trace is shown, including sections for 'Inputs' (e.g., 'eva\_query' with path 'eva\_query'), 'Output' (e.g., 'eva\_listings' with path 'eva\_listings'), and 'query\_explanation' (with path 'query\_explanation').

モデルがどのように動作しているのか、リアルタイムでのデバッグと可視化ができ、開発者はモデルの挙動を瞬時に理解できる

The screenshot shows the 'Compare Evaluations' page. It displays two evaluations: 'Evaluation (SM)' and 'Evaluation (SM) | ExtractItemRelevanceJudgment:v246'. Below the evaluations, there are two charts: a 3D scatter plot titled 'Summary Metrics' comparing 'eva\_query\_retrieving\_extraction', 'eva\_fields\_similarity', and 'eva\_fields\_relevance', and a bar chart titled 'Scorecard' comparing various metrics across different models and configurations.

数百件の事例に対する関連性ラベルの精度や、多岐にわたるスコアリング機能の効果を、レーダーチャートやヒストグラムを通じて視覚的に確認することが可能に



# Nejumi LLM日本語評価と Nikkei AI Model Score



<http://nejumi.ai>



Nikkei AI  
Model Score

## ホワイトペーパー： LLM評価のベストプラクティス



W&B社のこれまでのLLMモデル評価に関するリサーチ結果を集結したホワイトペーパーです。言語能力・安全性・領域特化型評価を包含するLLM評価の全体像と各評価項目のタクソノミーを詳細に解説しています

また、日本のLLM開発をリードする研究者、エンジニア、事業開発者からのフィードバックを踏まえ、実践的な日本語評価のベストプラクティスをまとめています。



<https://wandb.me/jp-llm-eval-wp-dload>

Weights & Biasesでは2023年初頭よりLLMの日本語評価に取り組んでおり、言語能力と安全性において、多岐にわたる評価フレームワークを構築し、60を超える公開モデルを評価してきました。

評価結果はWandBレポートの形式でNejumi.aiにて公開をすると同時に、2024年8月からは日経新聞社と共同でNikkei AI Model Scoreを開発しました。

引き続き極めて激しい競争が続くLLM開発の状況を的確に把握するとともに、自社での開発に最適なモデル選択を行うためのツールとして多くの方にご利用いただいているいます。

弊社の評価フレームワークは、経産省・NEDOの推奨するGENIACプロジェクトやNIIの主宰するLLM-jp、そして多くの国内企業の社内モデル評価に採用されています。

# W&BとNVIDIAとのインテグレーション

Weights & BiasesはNVIDIA製品とのシームレスなインテグレーションを提供しています。データ管理や実験トラッキング、モデル最適化など、W&Bの提供する革新的なAI開発環境をNVIDIAユーザーはさらに深く活用できます。■NVIDIAのGPUハードウェアおよびNeMoや NVIDIA AI Enterpriseなどソフトウェアとのシームレスな統合により、■AI開発のプロセスを加速させます。

## W&B support for NVIDIA technologies

**Weights & Biases MLOps**

- Artifacts**  
Version and manage your ML pipelines
- Tables**  
Visualize and explore your ML data
- Experiments**  
Track and visualize your ML experiments
- Sweeps**  
Optimize your hyperparameters
- Reports**  
Document and share your ML insights
- Registry**  
Publish & share your ML models & datasets
- Launch**  
Package and run your ML workflow jobs

**NVIDIA tools & frameworks supported by W&B**

- NGC
- Base Command Platform
- NeMo/BioNeMo
- TAO
- MONAI
- Triton
- NVIDIA AI Enterprise

**Accelerated Mainstream Servers**

- NVIDIA GPU (DGX) - On-prem, Cloud, Hybrid, DGX Cloud Application Acceleration
- NVIDIA CERTIFIED**

Preferred Partner    INCEPTION PREMIER MEMBER    AI ACCELERATED

# Weights & Biasesについて

Weights & Biases, Inc. (W&B社) は、米国サンフランシスコを拠点とし、エンタープライズグレードのML実験管理およびエンドツーエンドMLOpsワークフローを包含する開発・運用者向けプラットフォームを提供しています。

WandBは、LLM開発や自動運転、創薬など幅広い深層学習ユースケースに対応し、NVIDIA、OpenAI、Toyotaなど、国内外で80万人以上の機械学習開発者に信頼されているAI開発の新たなベストプラクティスです。

W&B社日本語ウェブサイト: <https://wandb.jp>

W&B Japan 株式会社: 〒150-0002 東京都渋谷区2丁目24-12 渋谷スクランブルスクエア39F

ご連絡先: [contact-jp@wandb.com](mailto:contact-jp@wandb.com)

