

週刊生成AI with AWS – 2025/1/27週

by Masato Kobayashi | on 03 2月 2025 | in General, Generative AI | [Permalink](#) | [Share](#)

みなさん、こんにちは。AWS ソリューションアーキテクトの小林です。

先週はDeepSeekの話題がホットな一週間でした。私自身もいろいろなお客さんから、DeepSeekに関してお問い合わせやご相談をいただきました。AWSとしては用途に応じて最適な精度・コスト・レイテンシを備えたモデルを選択して利用できたり、時には自分で開発・調達したモデルをデプロイして利用できることが大事だと考えており、さっそく[DeepSeekモデルについても選択肢のひとつに加わりました](#)。

それでは、1 月 27 日週の生成AI with AWS界隈のニュースを見ていきましょう。

さまざまなニュース

◦ [ブログ記事「DeepSeek-R1 models now available on AWS」を公開](#)

英語版の記事そのままですが、速報ということで取り上げます。世の中で注目されているDeepSeekモデルですが、AWSのAmazon BedrockやAmazon SageMaker AIで動作させることが可能になっています。昨年のre:inventで発表されたBedrock Marketplaceの仕組みを使ってDeepSeek-R1を組み込んだアプリケーションをホストしたり、SageMaker JumpStartで学習済みのDeepSeek-R1モデルを動かしてハンズオンの検証を行うなどができるようになっていきますので、興味のある方はぜひトライしてみてください。

◦ [ブログ記事「Amazon Bedrock で DeepSeek-R1 Distilled Llama モデルをデプロイする」を公開](#)

こちらはDeepSeekに関する和訳済みのブログ記事です。DeepSeek-R1をベースとして、Meta LlamaやQwenのアーキテクチャに基づく15-700億パラメータの蒸留モデルが公開されています。DeepSeek-R1-Distill-Llama-8BとDeepSeek-R1-Distill-Llama-70BについてはAmazon Bedrock Custom Model Import機能でインポートして利用できますので、その方法をご紹介しますブログ記事です。

◦ [AWS生成AI国内事例ブログ: 東京海上日動システムズ株式会社様、LLMによるアプリケーションモダナイゼーションへの挑戦](#)

東京海上日動システムズ様では、多くの基幹系システムをAWSに移行済みですが、一部はオンプレミスでの運用を継続しており、移行済みのシステムの多くはリフト&シフトによるレガシーなアプリケーション構造のままとなっています。オンプレミスのサーバやEC2で稼働しているJavaアプリケーションをサーバレスアーキテクチャにモダナイズするために、生成AIを活用し効率化することにチャレンジされています。AWS Prototyping Programを活用することで素早く小規模なアプリケーションでの検証を実施、95%は生成AIによるコードで動作し、エラーの多くが単純な修正で解消できることが確認されました。次のチャレンジはひとつのパッケージで完結しないアプリケーションでのバリデーションチェックやエラーハンドリングとのことです。生成AIによるアプリケーションモダナイゼーションは興味深い分野ですので、同様の課題感をお持ちの方はぜひご一読ください。

◦ [ブログ記事「GraphRAG Toolkit の紹介」を公開](#)

検索拡張生成(RAG)の精度や、質問への適合性を高めるため、グラフDBによる情報間の関係性を利用するGraphRAGというテクニックが知られています。この記事はグラフDBを活用したRAGワークフローの構築を容易にするPythonライブラリであるGraphRAG Toolkitの意義と使い方をご紹介しますものです。

◦ [ブログ記事「AWSで実現する安全な生成 AI アプリケーション – OWASP Top 10 for LLM Applications 2025 の活用例」を公開](#)

生成AIによるアプリケーションの安全性は、様々な企業や組織にとって重要な課題です。このブログ記事で

はOWASP(Open Worldwide Application Security Project)が提唱する、LLMを組み込んだアプリケーションにおける主要な10のセキュリティ脅威をまとめたOWASP Top10 for LLM ApplicationについてAWSでアプリケーションを設計・開発する方が考慮すべきポイントやリスクシナリオを概説しています。

- **ブログ記事「金融業界における生成AI活用動向」**

様々な業界で生成AIの可能性への期待が高まる中で、2024年は業務での実用を検討・開始する年となりました。この記事では、AWSで金融領域の事業開発担当者からみたAIの活用動向について、インタビュー形式でご紹介するものです。読み物として気軽に読めるようになっていきますので、金融業界と関わりの深い方も、そうでない方も、ぜひご覧ください。

- **ブログ記事「デジタル庁主催の AI ハッカソンに参加しました」を公開**

2024年11月にデジタル庁主催で「AIハッカソン、アイデアソン」が開催されました。AWSのエンジニアチームとしてもこの取り組みに参加させていただきましたので、考案したソリューションについてご紹介するブログ記事です。

- **ブログ記事「生成AIとデータによる小売体験の刷新」**

他業界でもそうですが、小売業や消費財業界ではデジタルトランスフォーメーションの重要性が一段と高く叫ばれています。この記事では、生成AIによってどういった変革が可能になるのかを紹介しています。

サービスアップデート

- **Amazon SageMaker Unified Studioのプレビュー可能リージョンを7箇所追加**

Amazon SageMaker Unified Studioはデータ・アナリティクス・AIに関するコラボレーションやデータを扱う処理の素早い構築を可能にする統合環境です。今回、新たに7つのリージョン(ソウル、シンガポール、シドニー、フランクフルト、ロンドン、サンパウロ、カナダ(中央))でプレビューが可能になりました。

- **Amazon Q in QuickSightのDashboard Q&A機能を発表**

Amazon Q in QuickSightでDashboard Q&A機能がご利用いただけるようになりました。ダッシュボードにおいて、データに関するQ&Aに応答する機能をワンクリックで追加でき、ダッシュボードのユーザがデータに関する疑問を持った際にセルフサービスで解決するために役立ちます。

- **Amazon Q Developer Agentが生成したコードに対するビルドとテストのリアルタイム実行に対応**

Amazon Q Developer Agentがアップデートされ、生成したコードを開発者がレビューする前にビルドやテストを行うスクリプトを実行できるようになりました。Amazon Qが生成したコードを開発者がチェックする前に、指定されたビルドやテストを自動実行しそれにパスしたものだけを開発者に提示することで、開発者に対してより精度の高いコードが提示される可能性が高まる機能です。

- **Amazon Q Developer Pro Tierで新規登録ユーザに対する通知メールの自動送信に対応**

Amazon Q Developer Pro Tierで新規登録されたユーザに対して自動的にメール通知が行われるようになりました。このメールは24時間以内に送信され、開発者がAmazon Q Developerを利用する上で重要な情報が含まれており、管理者の手間を省くことにつながります。

著者について

TAGS: AWSサービスアップデートまとめ, 週刊AWS