

DATASTACK
One-click RAG @ Scale

スケーラブルな RAGアプリをワンクリックで

Welcome

08 Nov, Friday
Tokyo

ご来場ありがとうございます

11月8日 (金) 東京



Revolutionising AI Development

Accelerating Innovation and Scale with Ease

AIアプリ開発の革新
イノベーションを加速して容易にスケール





Agenda

イベント概要

- ▶ Welcome Address **ごあいさつ**
- ▶ Keynote: Revolutionising AI Development: Accelerating Innovation and Scale with Ease
基調講演「イノベーションを加速し、容易にスケールできる AI開発の革新」
- ▶ Leveraging DataStax AI PaaS for Rapid Scalable Gen AI Application Development
「イノベーションを加速し、容易にスケールできる AI開発の革新」
- ▶ Afternoon Tea Break
- ▶ Staying Ahead of the Evolving AI Landscape - Google Cloud
- ▶ Demonstration: How to Build a RAG App in under 20 Minutes
- ▶ Networking

“Generative AI is the single most significant platform transition in computing history.”

Jensen Huang - CEO NVIDIA

**“生成AIは
コンピューティングの歴史で
最も重要なプラットフォームの
トランジションだ ”**

Jensen Huang - NVIDIA CEO

A.I. TURNS THIS SINGLE
BULLET POINT INTO A
LONG EMAIL I CAN
PRETEND I WROTE.

AIが、この1つの箇条書きを、私が書いた
ような長いメールに変換してるよ



A.I. MAKES A SINGLE
BULLET POINT OUT OF
THIS LONG EMAIL I CAN
PRETEND I READ.

AIが、私が読んだふりができるように長
いメールから1つの箇条書きを
作成してくれたよ



› Top 7 Use Cases for GenAI

生成AIの トップ7 ユースケース

CIO.com - July 2024

Advanced Chatbot (Chatbot with GenAI)
高機能なチャットボット

Digital Assistants (Doc summary, meeting notes etc)
文章要約や会議記録などのデジタルアシスタント

Coding Assistant
プログラミングの支援

Marketing Support/Content Creation
マーケティングサポートとコンテンツ制作

Drug Discovery (Specialised)
医薬品を探すこと

Cybersecurity and Fraud (ML with Content Creation)
サイバー犯罪とネット詐欺の検出

Business Process Augmentation (Evaluate Applications)
ビジネスプロセスの拡張

› Top 7 Use Cases for GenAI

生成AIの
トップ
ユースケース

Wave 1 GenAI 生成AIの第一波

CIO.com - July 2024

Advanced Chatbot (Chatbot with GenAI)
高機能なチャットボット

Digital Assistants (Doc summary, meeting notes etc)
文章要約や会議記録などのデジタルアシスタント

Coding Assistant
プログラミングの支援

Marketing Support/Content Creation
マーケティングサポートとコンテンツ制作

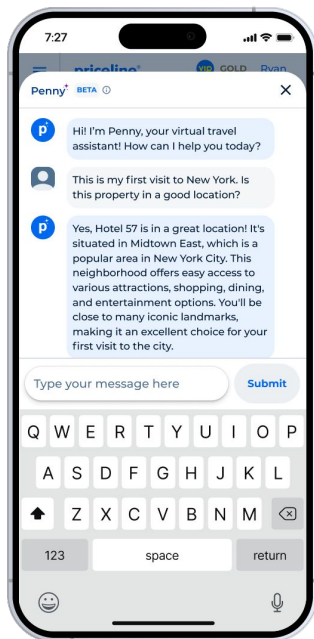
Drug Discovery (Specialized)
医薬品を探索

Cybersecurity and Fraud (ML with Content Creation)
サイバー犯罪とネット詐欺の検出

Business Process Augmentation (Evaluate Applications)
ビジネスプロセスの拡張

The Business Transformation With AI, Starts With Transforming The Business

AIによるビジネスの革新は
「ビジネスそのものを革新する」
ことから始まる



Transitioned to a scalable, cost-effective architecture that enhances user experience while reducing LLM-related expenses
 スケーラブルでコスト効果の高いアーキテクチャに移行し
 ユーザー体験をLLMで向上

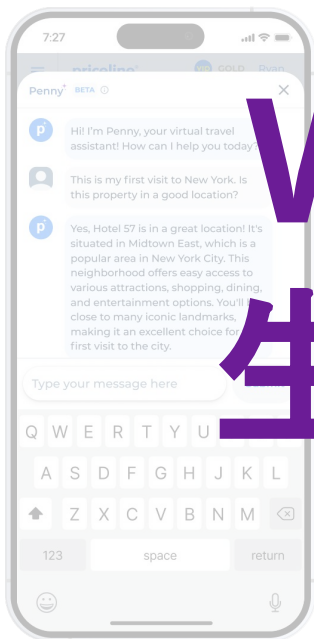
25% Increase in bookings
 in 2 weeks
 2週間で25%予約増加

Challenges 課題

- **Scalability Limits:** Growing data needs hit token limits and slow performance on GPT-3.5
 スケーラビリティの限界: 増え続けるデータで限界に達してGPT3.5のパフォーマンスが低下
- **Data Complexity:** Instant retrieval of highly specific details required more advanced search mechanisms
 データの複雑性: 詳細情報を取得するには高度な検索メカニズムが必要

Solution 解決策

- **Vector Search:** Astra DB for scalable, vector-based search optimizing performance
 ベクトル検索: Astra DBのスケラブルなベクトル検索による最適化された性能
- **Flexible LLM Selection:** Ability to explore different LLMs
 LLMを選択可能: 異なるLLMを検討できる柔軟性
- **Improved Search Accuracy:** More relevant results from millions of records
 検索の正確性の改善: 数百万のデータから高い回答精度を実現



Wave 2 GenAI 生成AIの第二波

Transitioned to a scalable, cost-effective architecture that enhances user experience while reducing LLM-related expenses
 スケーラブルでコスト効果の高いアーキテクチャに移行し
 ユーザー体験をLLMで向上

25% Increase in bookings in 2 weeks
 2週間で25%予約増加

Challenges 課題

Scalability limits: Scaling up with GPT-3.5 performance on GPT-3.5
 スケーラビリティの限界: 増強は継続的に限界に達してGPT3.5の性能が低下
 Complexity: Handling highly specific details required more advanced search mechanisms

データの複雑性: 詳細情報を取得するには高度な検索メカニズムが必要

Solution 解決策

- **Vector Search:** Astra DB for scalable, vector-based search optimizing performance
 ベクトル検索: Astra DBのスケラブルなベクトル検索による最適化された性能
- **Flexible LLM Selection:** Ability to explore different LLMs
 LLMを選択可能: 異なるLLMを検討できる柔軟性
- **Improved Search Accuracy:** More relevant results from millions of records
 検索の正確性の改善: 数百万のデータから高い回答精度を実現

What is slowing us down?

何が私たちを停滞させているのか？

▶ 11 Barriers To Effective AI Adoption

Forbes.com - Bernard Marr

AIの活用を阻害する 11の障壁

1. Leadership Inertia リーダーシップの怠惰
2. Fear of the unknown 未知への恐れ
3. Lack of Understanding of AI's Potential
AIの可能性が理解できない
4. **Data Availability and Quality**
データがない あるいは データが低クオリティ
5. **Skills Shortage**
スキル不足
6. **Integration Challenges with Legacy Systems**
レガシーシステムの統合が困難
7. Ethical and Legal Considerations 倫理/法的な配慮
8. **Costs** コスト
9. Lack of a Strategic Approach 戦略的アプローチの欠如
10. **Difficulty Scaling AI Initiatives**
AIイニシアティブをスケールさせられないという困難
11. Lack of Innovation Culture イノベーションカルチャーの欠如

<https://www.forbes.com/sites/bernardmarr/2024/05/10/11-barriers-to-effective-ai-adoption-and-how-to-overcome-them/>

› Hurdles of GenAI App Development

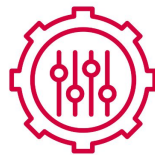
生成AIアプリ開発のハードル



Inconsistent Quality

一貫性のない品質

GenAI models and data availability can produce unpredictable and subpar outputs that fail to meet business needs



Skills

スキル

Developers need to research, identify, learn and test an array of new tools when building apps

開発者はさまざまなツールを探し、学び、テストする

必要がある



High Development Costs

開発コスト

Building and maintaining GenAI applications is expensive and time-consuming

AIアプリの開発と保守は

コストが高く 時間がかかる

Revolutionising AI Development

AI開発を革新する



2024 - Production AI 2024 - AIを実行環境へ

Energy



Financial Services



Transportation / Logistics



Media / Telecommunications



Retail / Food & Beverage



Technology & IT Services



THE FORRESTER WAVE™

The Forrester Wave™: Vector Databases

Q3 2024



Forrester Wave: Vector Databases

ベクトルDBベンダー
の評価

**The ONLY Hybrid
Leader.**

**(DataStaxは)
唯一の
ハイブリッドリーダー**

DataStax AI Platform as a Service

ORCHESTRATION

Agents Prompts Tools Models Vector Stores Embeddings Helpers



DATA PLATFORM

Vector Database



Hybrid Search



Data Streaming



INTEGRATED SERVICES

Ingestion



Agents



LLMs



RUNTIME DEPLOYMENT

nvidia

aws



Google Cloud

docker

kubernetes



› AI PaaS: The Catalyst for AI > AIを普及させる触媒

Rapidly Develop & Deploy AI Apps 迅速な開発とデプロイ

Fastest path to build,
deploy RAG

最速のRAGアプリ開発

Shorter dev cycle with test
& iterate from data

ingestion to production
データを活用した高速な

開発サイクルで

プロダクション環境へ

Enhance Agility for Innovation イノベーションへの 俊敏性を拡張

Developers can pivot to
evolving needs fast

ソフトウェア開発者が
革新を素早く取り込める

Stay ahead of market shifts
by prioritizing agility

マーケットの変化に対応
できる俊敏性を優先

Tailor for Seamless Scale, Performance シームレスなスケーラビ リティと性能

Easy integration of AI
ecosystem services without
complexity

複雑性を排除しながら
AIエコシステムとの容易な
統合を

Deploy AI applications and
models at scale

AIアプリ/モデルをスケール

DataStax AI Platform as a Service

We make it **10x easier** for developers
to build AI apps with **20% higher
relevance** and unlimited **scale**

DataStaxはAIアプリ開発を
10倍カンタンにできるようにして
20%以上回答精度を高め
無限のスケールビリティを実現します



DATASTAX

Thank you!
ありがとうございました！

