

Google Cloud Next Tokyo 26

ZEAL

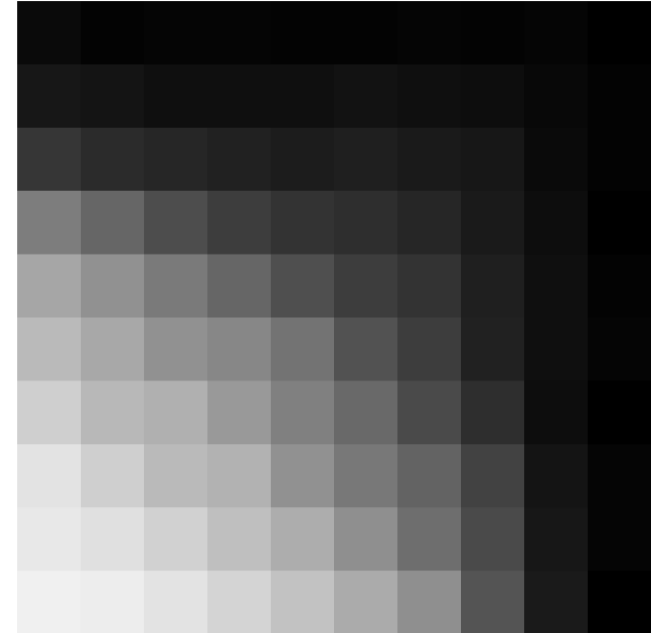
ZEAL CORPORATION

www.zdh.co.jp

Google Cloud で作る AI-Ready な データレイクハウス

株式会社ジール
辻 修寛

2026.07.31



自己紹介

辻 修寛／Nobuhiro Tsuji

株式会社ジール

データ & AI CoE 本部
テクノロジープリセールス部

略 歴

小売業での店舗マネジメントを経験後、IT業界へ転身。

インフラエンジニアとして金融機関のミッションクリティカルなシステム基盤構築に従事する。現在はプリセールスエンジニアとして、Google Cloudをはじめとするモダンデータスタックを活用。

お客様の課題に寄り添い、データおよび AI を実務に活かすための支援を幅広く展開している。

好きなGoogle Cloud のサービス：BigQuery, Looker



｜ ジールのビジネス

ジールは、データを最大限に活用し
お客様のDX・AI活用を支援する
DX推進事業を展開しています

データ分析・AI／BIコンサルティング

データ分析基盤・AI／BIデータ活用環境構築

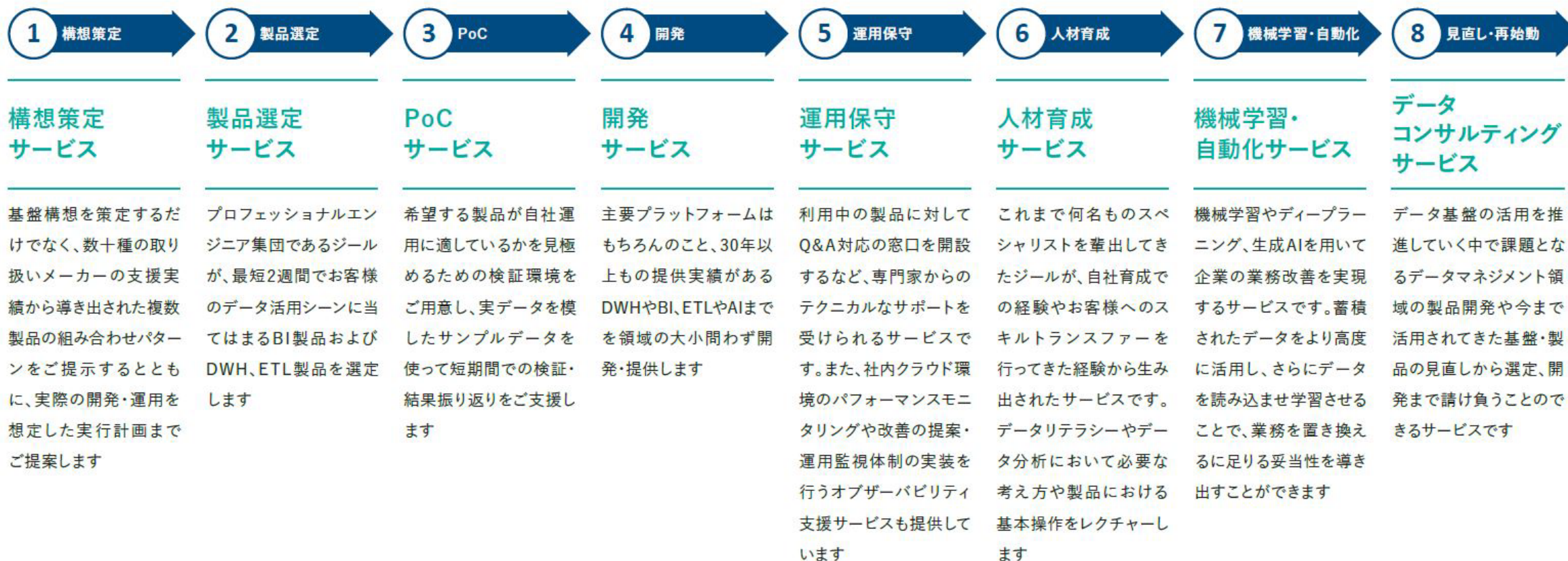
データドリブン人材育成

AI系ソフトウェア・IP開発

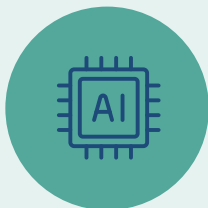
クラウドネイティブ化・DevOps支援

データ利活用プロジェクトの各フェーズに応じたサービス

データ活用プロジェクト推進を熟知しており、各フェーズに応じ最適なサービスを提供しています。



企業の AI 活用の成否は非構造化データにある



AI は驚くほどに賢いが、社内のことは知らない
自社データを使うことで AI 利用価値を最大化できる

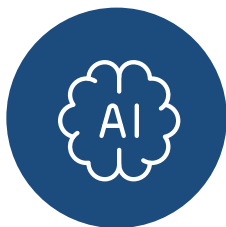


社内ナレッジの約70～90%、PDF、メール、画像や音声
などの”非構造化データ”である
ユーザーが求める情報の大半も”非構造化データ”となる



構造化・非構造化データを蓄積できるレイクハウスが
必要となる

AI 導入後に顕在化した理想と現実のギャップ



期待した精度が出ない

- ✓ “もっともらしく” 事実と異なる回答（ハルシネーション）を生成する
- ✓ PDF 内の「表データ」の行や列を読み違える
- ✓ AI の回答が信用できず、最終的にソースファイルを確認している

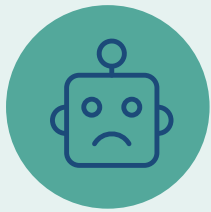


想像以上にコストが高い

- ✓ ユーザーが質問のたびにPDFを丸ごとLLMに読ませるなどの非効率な設計により、API 課金が高騰する
- ✓ 検索用インフラの維持・管理や、個別に開発したデータパイプラインの運用・保守コストがかさむ

想定していたより人の作業が楽にならない
ROI（費用対効果）が出ない

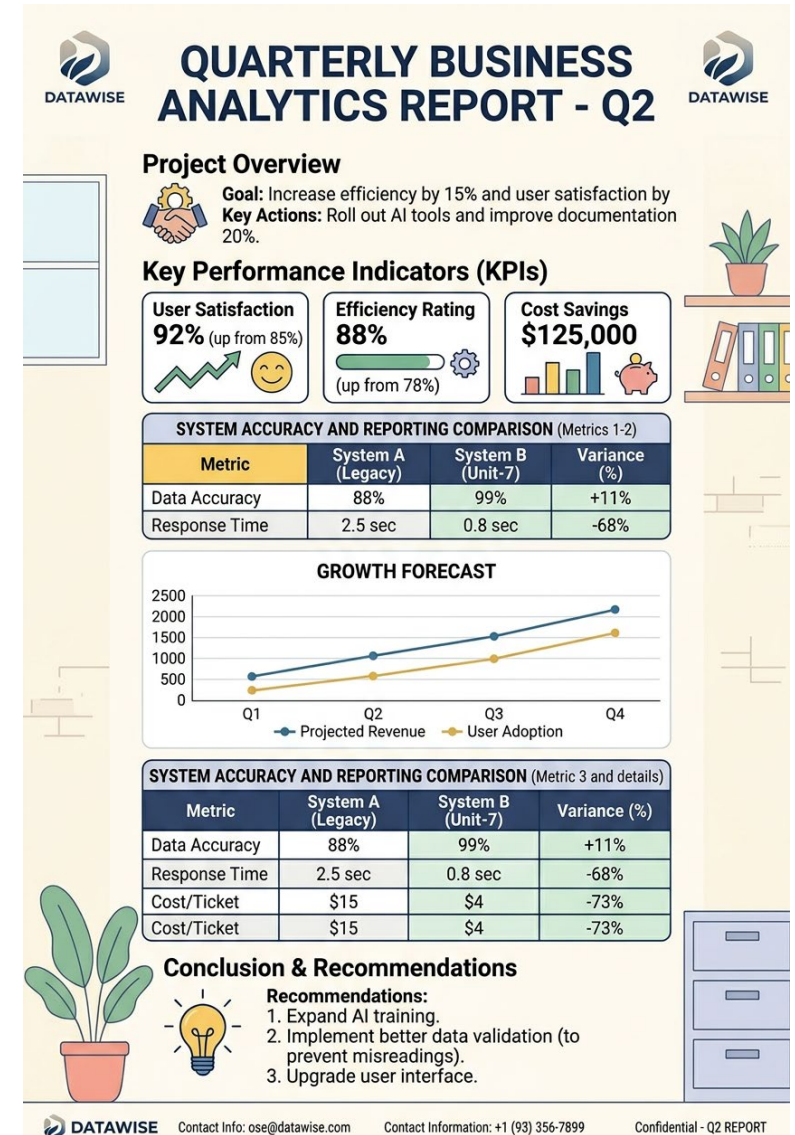
なぜ AI は期待通りに動かないのか？



問題原因は、データのフォーマットにある
「人間が読みやすい見た目」 は、AI にとっては
ノイズや誤読の原因



非構造化データを、あらかじめ AI が理解しやすい
『AI-Ready』 な形式に前処理することが鍵となる



AI が正しく解釈できるオープンフォーマット

「Markdown」

見出し、リスト、表の構造を明示。AI の「チャンク分割（ぶつ切り処理）時の文脈喪失」を防ぐ

→ 「Mermaid」

業務フローやシステム構成を「テキストコード」で表現。AI に因果関係や接続ロジックを正確に理解させる

2. データ分類と承認ルール

機密性区分	データ例	承認者	有効期限
極秘	顧客個人情報、未公開財務情報	データガバナンス委員会	30日
社外秘	設計書、ソースコード	部門長	90日
社内公開	社内マニュアル	自動承認	無期限



2. データ分類と承認ルール

```
| 機密性区分 | データ例 | 承認者 | 有効期限 |  
| :--- | :--- | :--- | :---: |  
| 極秘 | 顧客個人情報、未公開財務情報 | データガバナンス委員会 | 30日 |  
| 社外秘 | 設計書、ソースコード | 部門長 | 90日 |  
| 社内公開 | 社内マニュアル | 自動承認 | 無期限 |
```

AI に構造化されたテキストを読ませることで精度が劇的に向上する

フローチャート、棒グラフも表現可能

3. アクセス権申請フロー

```
```mermaid
graph TD
 A[アクセス権申請] --> B{機密区分}

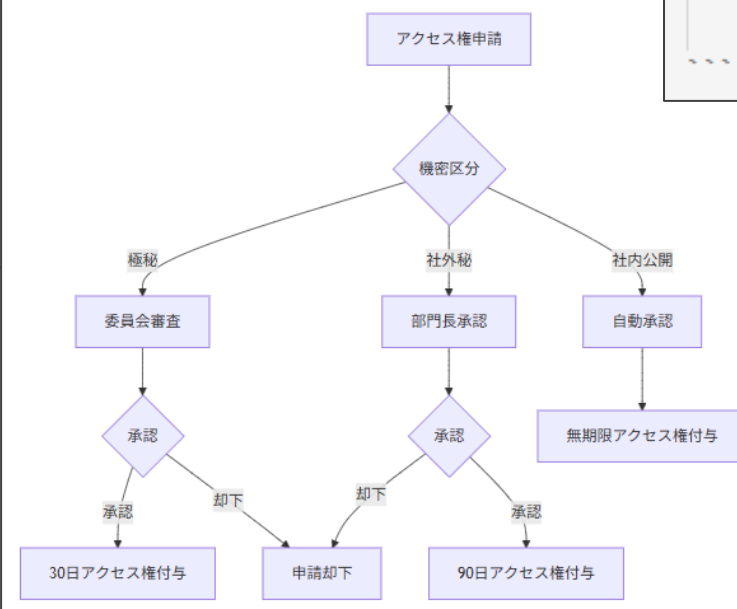
 B -->|極秘| C[委員会審査]
 B -->|社外秘| D[部門長承認]
 B -->|社内公開| E[自動承認]

 C --> F{承認}
 D --> G{承認}

 F -->|承認| H[30日アクセス権付与]
 F -->|却下| I[申請却下]
 G -->|承認| J[90日アクセス権付与]
 G -->|却下| I

 E --> K[無期限アクセス権付与]
```

## 3. アクセス権申請フロー



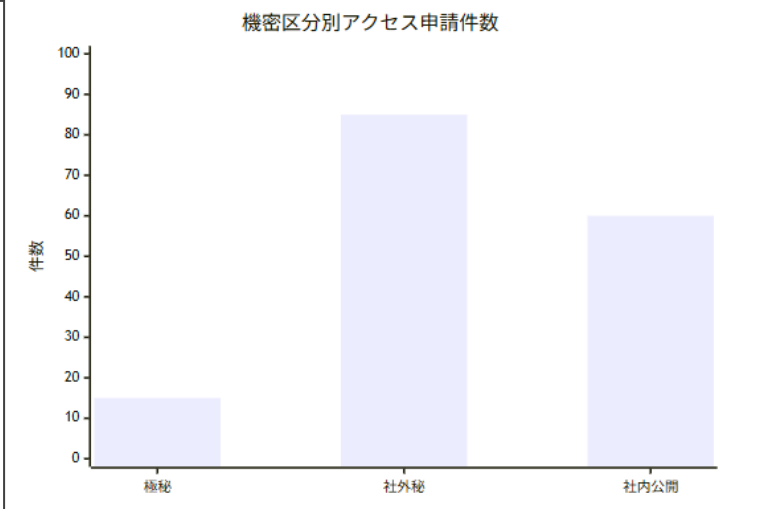
## ## 4. 機密区分別申請件数

申請件数の大半は社外秘データ向けであり、  
極秘データへのアクセス申請は限定的である。

```
```mermaid
xychart-beta
    title "機密区分別アクセス申請件数"
    x-axis ["極秘", "社外秘", "社内公開"]
    y-axis "件数" 0 --> 100
    bar [15, 85, 60]
```

4. 機密区分別申請件数

申請件数の大半は社外秘データ向けであり、
極秘データへのアクセス申請は限定的である。



検証：PDFとMarkdown化データをGeminiで比較

◎検証内容

総務省が公開する人口推計のレポートを使用。

Gemini-2.5-flashに次を指示

①プロンプトとPDF

②プロンプトとPDFをマークダウン化したデータ

◎プロンプト

総人口の表（直近1年）の、6月の総人口の数値を正確に教えてください

◎PDFの対象部分

年月	2025年							2026年					
	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	5月	6月
総人口 (千人)	123,371	123,367	123,268	123,192	123,219	123,203	123,165	122,980	<i>P</i> 122,860	<i>P</i> 122,850	<i>P</i> 122,860	<i>P</i> 122,810	<i>P</i> 122,850
前年同 月比(%)	-0.49	-0.49	-0.50	-0.47	-0.47	-0.47	-0.47	-0.46	<i>P</i> -0.47	<i>P</i> -0.46	<i>P</i> -0.44	<i>P</i> -0.43	<i>P</i> -0.43

出典：[2026年（令和8年）6月報](#)

P：概算値

結果：PDFとMarkdown化データをGeminiで比較

①プロンプト+PDFそのまま

{"text":"5ページ目の総人口の表（直近1年）に登場する「6月」は以下の通りです。
2025年6月: 123,371千人
2026年6月: 122,850千人"}

ジョブ情報	結果	可視化	JSON	実行の詳細	実行グラフ
行	answer				
1	{\"candidates\": [{\"avg_logprobs\": -0.061092438905135445, \"content\": {\"parts\": [{\"text\": \"5ページ目の総人口の表（直近1年）に登場する「6月」は以下の通りです。\\n\\n* **2025年6月**:\\n123,371千人\\n* **2026年6月**:\\n122,850千人\"}], \"role\": \"model\"}, \"finish_reason\": \"STOP\", \"score\": -4.215378284454346}], \"create_time\": \"2026-07-20T18:38:41.244084Z\", \"model_version\": \"gemini-2.5-flash\", \"response_id\": \"MWteavTvDoeC\"}				

②プロンプト+マークダウン化したデータ

{"text":"5ページ目の総人口の表（直近1年）に登場する「6月」は以下の通りです。
2025年 6月: 総人口 123,371千人
2026年 6月: 総人口 **P** 122,850千人"}

ジョブ情報	結果	可視化	JSON	実行の詳細	実行グラフ
行	answer				
1	{\"candidates\": [{\"avg_logprobs\": -0.05797215511924342, \"content\": {\"parts\": [{\"text\": \"5ページ目の総人口の表（直近1年）に登場する「6月」は以下の通りです。\\n\\n* **2025年 6月**: 総人口 123,371千人\\n* **2026年 6月**: 総人口 P 122,850千人\"}], \"role\": \"model\"}, \"finish_reason\": \"STOP\", \"score\": -4.4058837890625}], \"create_time\": \"2026-07-20T18:38:52.140392Z\", \"model_version\": \"gemini-2.5-flash\", \"response_id\": \"PGteaujlCOWznMwP862T2\"}				

PDFそのままのケースでは情報が削り落とされていた

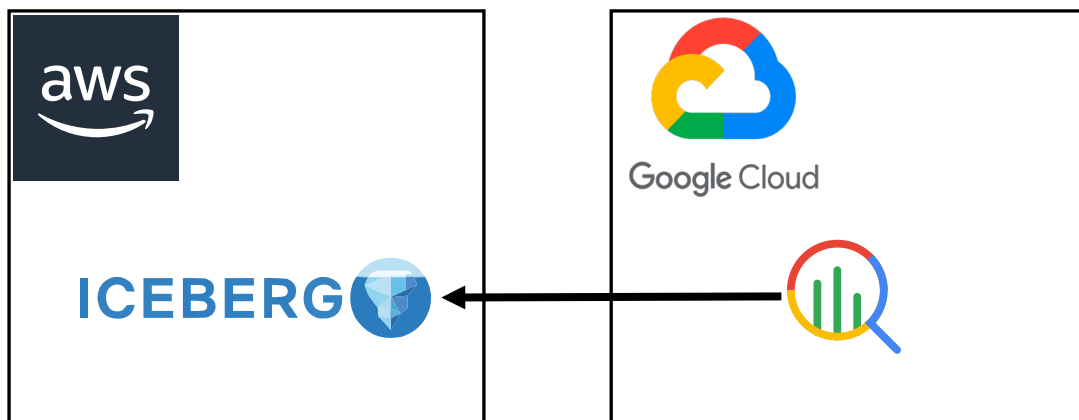
AI-Ready なデータは”使い回せる”ことが重要

✕ 特定のデータベース

- AI やツールを変えるたびにデータ移行が必要
- データコピーが増え、管理が複雑化
- 他部門・他クラウドから利用しづらい
- 移行・運用コスト（お金・時間）が発生する

○ オープンテーブルフォーマット

- AI やツールだけを切り替えられる
- 技術の進化に追従しやすい
- クラウドを超えてデータを共有しやすい
- コストを抑えてデータ活用が可能



「Cross-cloud Lakehouse (Preview)」
AWS S3 上の Apache Iceberg データを
BigQuery で参照可能

「高精度×低コスト」を両立する AI-Ready データレイクハウス

◎Apache Iceberg を中心とした、データを「動かさない」アーキテクチャ

Cloud Storage 上のデータを動かさず、BigQuery から直接参照し、SQL と Gemini の連携だけで自動的に『AI-Ready』なデータに変換してレイクハウスに蓄積する



精度を上げるだけでなく、コストも抑えることが可能

『AI-Ready』なデータ基盤で押さえるべき2つのポイント

Point

1

AI の精度を分けるのは「データの形」である

人間にとっての見やすさは AI にとってはノイズ。AI の理解力を最大化するオープンフォーマット「Markdown / Mermaid」への変換が精度向上の鍵

Point

2

データは「オープン」に保ち、未来の AI 進化へのロックインを防ぐ

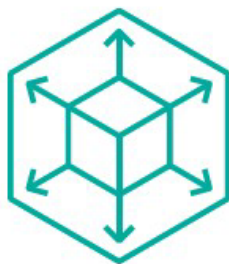
Apache Iceberg 等のオープンフォーマットでデータを蓄積することで、特定のベンダーに依存せず、いつでも最新 AI にデータを直接接続できる

AI 活用の成否は、モデルの性能だけでは決まらない
AI が理解できるデータと、将来も使い続けられる
データ基盤が重要である

AI-Ready 設計・サービス

年間350件を超えるデータ分析基盤の構築実績で培ったノウハウにより、各フェーズにおいて人とAIが共存する設計・サービスを提供し、AI-Readyな分析基盤構築を支援しています。

【データレイヤー】



データ設計の拡張

AIが理解できる
データ設計へ

人とAIが“同じデータ”を利用

【アーキテクチャ】



基盤設計の再定義

人とAIが共存する
基盤設計へ

AIを後付けにしない構造

【利用レイヤー】



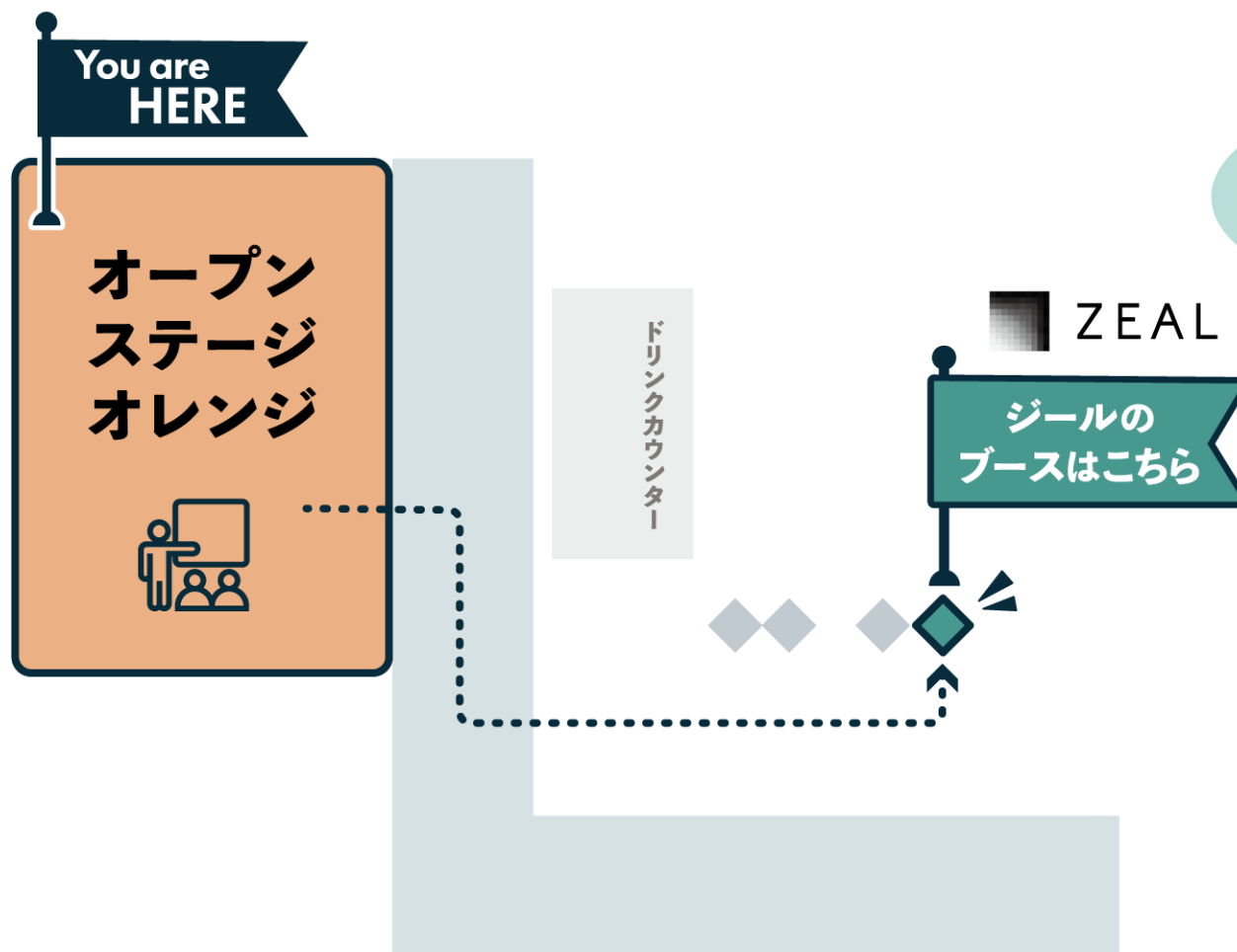
業務設計の進化

AIを業務判断に
組み込む設計へ

“可視化”から“自律化”へ

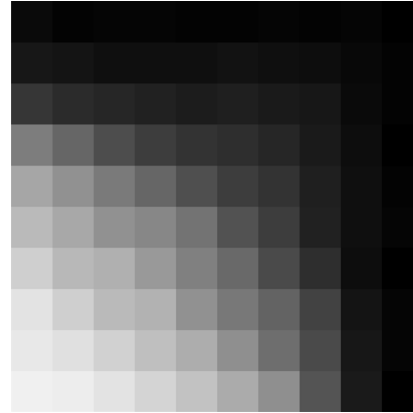
ジールのブースはこちら

◎ジールのブースはオープンステージを出てドリンクカウンター近く



お気軽に
お立ち寄りください！





ZEAL

SUSTAINABILITY IS VALUE