GLG

AIからROIへ:
AI投資を価値実現へ導く

5名の有識者が語る、構想から運用までの実践知



目次

はじめに 3

第1章　手段の目的化 ― なぜ多くのプロジェクトは変革に至らないのか 5

第2章　エンタープライズAIは、価値を成果につなげられているか 9

第3章　進捗は、何を基準に測られているのか 14

第4章　AIの経済構造はSaaSの常識を覆す 18

第5章　最も価値の高いAIほど、実現までに時間を要する理由 23

おわりに 27

はじめに

2026年、世界のAI支出は前年比**47%増の2兆5,900億ドル**に達すると予測されています。その圧倒的な金額の大きさもさることながら、特筆すべきは、AI投資の大半が他の分野の投資であれば通常課される厳格な承認プロセスを踏むことなく進んでいるという現実です。

「巨額のAI投資が、通常の承認プロセスを経ずに進んでいます。」

多くの企業はすでにAI投資の手応えを実感しており、AI施策による具体的な成果を挙げています。例えば、マッキンゼーが2025年に発表した「State of AI」では、**回答者の64%が「AIによるイノベーション効果をすでに実感している」と回答**し、ソフトウェアエンジニアリング、製造、マーケティング、営業などの個別ユースケースにおけるコスト削減や収益向上効果が報告されています。

しかし、投資が加速し技術革新が進む中で、部門単位でのAIの実質的効果と、経営陣が自信を持って示せる企業レベルの価値や収益効果のギャップは広がり続けています。同調査では、**AI投資を企業全体の業績向上に結びつけられている企業はわずか39%にとどまり**、その大半においても、影響はEBIT（利払前・税引前利益）全体の5%未満にとどまっています。米国マサチューセッツ工科大学（MIT）のProject NANDAはこの傾向をより鮮明に示しており、**企業向け生成AIパイロットの約95%は損益に対して測定可能な効果をもたらしていないと報告しています。**個々の施策レベルでは効果を測定できるケースもあるものの、企業全体でのROIの実証には、依然として大きな課題があるのです。

このギャップを埋めるためには何が必要なのか――。GLGは、AI導入の最前線で活躍する5名の有識者にインタビューを行いました。それぞれの有識者が、この課題に対して異なる視点を持っています。

- **Matt Baker**
  Dell Technologies 元コーポレート戦略・AI戦略担当シニアヴァイスプレジデント（SVP）
- **Divya Ashok**
  Altera Ventures 創業者、Salesforce AI Research 元最高執行責任者（COO）
- **Yiannis Antoniou**
  Lydatum 最高技術責任者（CTO）兼 AI・データ統括責任者
- **Mallika Iyer**
  AIインフラストラクチャ専門家、Google Cloud AI および CoreWeave 元エンジニアリングリーダー、Arya AI 創業者
- **Dr. Ryan Mackey**
  LifePoint Health 最高医療情報責任者（CMIO）兼 AI／RCM／CDI 担当医師役員

5名の経歴は多岐にわたります。Fortune 500企業でのAI戦略立案、金融機関のAI基盤となるデータインフラの構築、ベンダーおよび顧客の双方の立場でのエンタープライズAIの売買、病院システムの臨床・収益サイクル業務へのAI展開、そして大手IT企業のエンジニアリング組織でのAIプロダクト開発――いずれも、現場で培われた知見です。

GLGは、独自のAIモデレーター（**GLG AIインタビュー**）を活用して、5名の有識者に同じ問いを投げかけました。その問いとは、「AIの導入において、技術が機能している段階から、経済的価値を企業レベルで証明できる段階に至るまでの間で、どこに行き詰まりが生じるのか」というものです。5名の有識者は全員、このギャップが重要であるという点では一致しました。一方で、その原因と示唆については見解が分かれました。

5名の見解を重ねると、AI導入の構想から成果の測定までの一連の流れが見えてきます。それぞれが、経済的価値の証明に行き詰まりが生じやすい局面として、異なる点を挙げています。プロジェクトの範囲設定や契約締結の前にあたる上流の局面もあれば、すでに成果の測定が始まっている下流の局面もあります。

自社のAI投資の成果を示せない原因は、どこにあるのか。本レポートは、その見極めと、測定の枠組みの構築に役立つ実践知を提供します。



第1章
手段の目的化 ―
なぜ多くのプロジェクトは
変革に至らないのか

By Matt Baker

Dell Technologies 元コーポレート戦略・
AI戦略担当シニアヴァイスプレジデント（SVP）

GLG

現在のAIをめぐる動向を見ていると、かつてビッグデータやHadoopが注目を集めた時代を思い出します。当時は、クラウドの導入そのものが目的とされるケースが目立ち、価値を示せないままのパイロットが多くの企業で相次ぎました。

いまなお多くの人がAIに同じような姿勢で向き合っており、その価値をめぐる議論も当時と同様に未成熟なままです。現在のAIプロジェクトの大半は、AIの導入そのものが目的となっています。リーダーは単発のイノベーション施策にAIを結びつけ、技術の力に目を奪われたまま事業の一部に投入しています。個別の業務では一定の効果が見られることもありますが、プロセス全体は実質的に変わっていません。

測定の課題は、実はここから始まっています。問題は数式や指標ではなく、プロジェクトの意図が「事業重視」か「技術重視」かにあります。この違いは、AIプロジェクトを成功させるうえでも、その効果を示すうえでも極めて重要です。

要するに、成否を分けるのは、意図と焦点、そしてマインドセットです。問われるのは、技術を導入することではなく事業を変革するというマインドセットを持てるかどうかです。事業変革が具体的であれば、事前に定義し測定できる目標と成果は、自ずと見えてきます。

数百のAIプロジェクトが教えてくれたこと

Dell Technologiesでは、生成AIへの期待が高まった黎明期に、800件以上のプロジェクトを棚卸しました。いずれも、事実上「AIが何か興味深いことをしている」例として扱われていたものです。しかし振り返ってみると、800件すべてが、価値を生み出しうる領域のごく限られた一面しか捉えていませんでした。一方で、カスタマーサポートのあり方を根本から変える、あるいはソフトウェア開発全体を刷新するといった大規模な変革は、より大きな成功を収めました。その効果ははるかに具体的で包括的であり、測定が可能でした。

全体を見渡すマインドセットを持つと、変化が何をもたらすかを、より的確に予測しやすくなります。個々のAIプロジェクトの導入を単独で測るのではなく、事業の抜本的な変革を見据えるからです。そのような全面的な変革では、人、ツール、プロセスを含む、事業のあらゆる層に変化が及ぶことが見込まれます。期待するのは事業上の成果であり、最終的に測定するのもその成果です。

「全体を見渡すマインドセットが、成果の予測精度を高めます。」

私が見てきた中で最も優れた成果のいくつかは、病院内の放射線科の業務全体とワークフローを管理する、特定のプロセスに焦点を当てた取り組みから生まれました。そこでは、導入した自動化によって、放射線科医の時間が生み出されました。ただし、この取り組みは、その分野向けのAIツールを導入するだけにとどまりませんでした。チームは、患者体験の全体にわたって、プロセス全体を再設計することを目指したのです。その結果、患者がX線検査の結果を受け取り、放射線科医の診察を受けるまでの時間が大幅に短縮されました。病院は多大な

コストを削減し、患者ははるかに早く結果を得られました。おそらく、より健康にもなったことでしょう。

このように意味のある測定可能な効果は、胸部X線の平均読影時間を変えるだけの施策とは対照的です。この施策が測定するのは、プロセス全体から切り離された一つの工程にすぎません。ワークフロー全体の中の一要素への影響だけを測定しても、効果は曖昧なままです。

賛同の度合いとプレッシャーの関係

変革には時間がかかります。報告サイクルが短く、成果を急ぐ圧力が強い組織は、見せかけの成果を生みやすくなります。AI導入のコストが帳簿に計上されてから財務上の効果が表れるまでには、通常1四半期では測れないほどの時間差があります。

成果を急ぐ圧力は、多くの企業で、AIの可能性とそこに至る課題への理解不足だけでなく、当初からの足並みの乱れや、関係者の賛同の弱さをも示しています。取り組みが組織全体の変革として正しく位置づけられていれば、取締役会や予算更新時の圧力は小さくなります。全員が賛同し、何が起きているのかを理解しているからです。難しくなるのは、リーダーが、AIを活用した限定的な機能に一定額を投じると約束し、ステークホルダーが成果を求め始めたときです。圧力が高まるのは、意図が明確でなく、成果がどのような形で、いつ表れるべきかについて、誰も認識を揃えていないからです。取締役会やCEOが「あらゆるものをAI化せよ」と言うような状態は、組織にとって危険な状態です。

「成果への圧力が強く、結果を急ぐ組織では、見せかけの成果を生み出しがちです。」

リーダーは、適切にスコープを定めたプロセス変革には、少なくとも1年かかると見込むべきです。まず、現状分析から始まります。例えば、対象のプロセスは何か、工程はいくつあるのか、何人が関わっているのか、といったことを確認します。変革に着手する前に、既存の業務プロセスと向き合い、深く理解します。そのうえで、範囲を限定した導入で、仮説の検証に進みます。

大規模な変革は、複数年かけて進みます。私がDellで関わった取り組みは、当初から複数年にわたるものと位置づけられていました。異なる時間軸で多くのプロセスを対象とし、それらが互いにどう積み上がっていくのかも、明確に理解されていました。

全員が変革のマインドセットに賛同しているとき、それは、混乱とコストを伴うプロセスに全員で向き合うと理解しているということです。組織の全員が、取り組みの意義と必要な時間を十分に把握していれば、進捗報告を求める圧力はそれほど高まりません。定期的な報告は、取り組み全体に当然の構成要素として組み込まれます。

「変革への賛同とは、全員が、混乱を伴う変化と費用について理解している状態です。」

ツールではなく、プロセスから始める

AIから確かな成果を得たいリーダーへの私の助言は、明快です。広範な事業変革に取り組んでいるか、確認してください。そうでなければ、AIは、すでに制御不能になっているプロセスをさらに加速させるだけです。プロセス全体の再構築は、AIの有無にかかわらず大きな成果をもたらすはずです。AIは、その成果を加速し、何倍にも高める役割を担います。価値を生むのは、新たに使えるようになったツールを前提に、事業のやり方を見直し、変えることです。AIツールを既存の業務プロセスに押し込み、うまくいくことを期待するだけでは、価値は生まれません。

クラウドの導入時を振り返ると、価値を示すまでに長い時間がかかりました。しかもその価値は、必ずしも財務面ではっきり表れたわけではありません。クラウドコンピューティングの導入が、常に安く済むとは限らなかったのです。同じことは、AIの導入時にも起こり得ます。期待と現実のギャップを埋める鍵は、ここでも変革のマインドセットです。勝者となるのは、技術そのものの力ではなく、事業変革と、この新しい種類の技術があらゆる変化の成果を劇的に加速させる可能性に、焦点を当てる組織です。



Matt Baker
Dell Technologies 元コーポレート戦略・
AI戦略担当シニアヴァイスプレジデント（SVP）

詳しくはこちら

第2章
エンタープライズAIは、価値を成果につなげられているか

By Divya Ashok
Altera Ventures 創業者、Salesforce AI Research 元最高執行責任者（COO）



GLG

エンタープライズAIの課題は、測定の問題というより、価値の獲得にあります。AIが意図どおりに機能していても、事業価値が生まれないことはあります。モデルは機能し、利用も進んでいるのに、それを事業の変化に結びつけられずにいるのです。

「どのような経済的価値を生み出したいのか。それをどう獲得するのか。」

企業のリーダーは、部門横断の重要な問いに答えなければなりません。私たちはどのような経済的価値を生み出したいのか、そして組織はそれをどのように獲得するのか、という問いです。

その価値は、ユースケースによって異なります。目的は、余力の創出、顧客離反やリスクの低減、追加収益の創出、採用の抑制、あるいは、従業員がより価値の高い業務に使える時間を取り戻すことかもしれません。いずれも正当な成果になりえますが、成功の定義も測定の仕方も、それぞれ異なります。

だからこそ、大規模なエンタープライズ導入にはすべて、「AI価値契約（AI value contract）」が必要になると私は考えています。これは、事業、技術、財務、ベンダーの間で、次の4点について認識を揃えるための合意です。追求する経済的価値、測定の基準となる現状の数値、その価値を実現するために事業で変えるべきこと、そして実現に責任を負う経営幹部です。

ベンダーと顧客の認識のずれ

価値契約が崩れやすい場面の一つが、ベンダーと顧客の間です。ベンダーは契約の更新、顧客は取締役会への報告を前に、成功について異なる考えを持ち込みがちです。ベンダーは利用の浸透を示して更新を正当化したく、顧客は事業への影響を示す必要があります。二つの説明は関連しているものの、通常は異なって聞こえます。そして契約を結ぶ前に、関係者がそれらをすり合わせることは、ほとんどありません。

ベンダーの説明は、利用率、エージェントによる処理件数、顧客満足度（CSAT）スコアに焦点を当てがちです。一方、顧客の取締役会が聞きたいのは、P&Lのどこが動いたかを軸にした、少し異なる説明です。更新をめぐる摩擦の多くは、この二つの捉え方の間にあります。例えばベンダーは、時給100ドルの業務を5万時間節約したとして、その価値を500万ドルの生産性向上とモデル化するかもしれません。余力の見積もりとしては妥当です。しかしそこには、プラットフォームの費用、必要な連携、人による例外対応、そして多くの顧客が当初は過小評価するAIスタックのその他の層のコスト

「契約を結ぶ前に、成功を測る指標がベンダーと企業の間ですり合わせられることは、ほとんどありません。」

が、含まれていません。より難しい問いは、その余力がどうなったのか、そして事業が最終的にどれだけの価値を獲得したのかです。

AIは余力を生み出すことができます。その余力を何に変えるのかは、経営陣が決める必要があります。例えば、採用の抑制、コンバージョンの向上、対応できるサービス問い合わせ件数の増加、従業員が考える時間を取り戻すことなどです。

組織内部で生じる同様の課題

顧客組織の内部でも、AIを日々使うチームと、それを評価する経営幹部との間で、同じようなずれが起きます。技術部門や業務部門のチームは、日々の生産性に注目し、それを測定しがちです。例えば、1時間かかっていたレポート作成が15分で済むようになった、あるいは、AIが数分でアカウントリサーチを仕上げたおかげで、BDR（営業開発担当者）に1時間の余裕が生まれた、といった例です。こうした改善は確かに意味のあるものですが、売上、利益率、コスト低減といった事業成果に、必ずしも結びつくとは限りません。

リーダーは、これを生産性向上そのものへの否定ではなく、成果が段階を踏んで表れていく過程と捉えるべきです。AIの構築には、それだけで時間がかかることがあります。そのうえで、従業員がツールを大規模に使うようになって初めて、組織は経済的な価値を目にします。さらに、ツールの使い方を学び、意味のあるワークフローを築く時間も必要です。創出された余力が財務諸表に表れるのは、その後のことです。このタイムラグは織り込んでおくべきもので、必ずしも導入の失敗と受け取る必要はありません。

「生産性の向上をすぐに疑う必要はありません。成果は、段階を踏んで現れます。」

このギャップは、私たちがベースライン（導入前の基準値）を整えないままAI時代に入ったことにも起因します。以前のプロセスを測定していなければ、導入後の変化も測定できません。事前に合意した指標を設定すること。それも、プロンプト数や処理トークン数のような代理指標ではなく、確かな成果を示す指標を設定することが、価値を実証できる導入かどうかを分ける決め手になります。

成果を起点に、ベースラインを確立する

組織が利用状況ではなく成果を軸に据えると、価値の説明が明確になります。多くの組織は、ベースラインを整えないままAI時代に入りました。導入前に測定していなければ、何が変わったのかを判断するのは難しくなります。指標は、単なる活動ではなく、成果を反映する必要があります。プロンプト数、処理トークン数、エージェントによる処理件数、利用の浸透度が示すのは、利用状況であって、経済的価値ではありません。

「利用状況ではなく成果を軸に据えると、価値の説明が明確になります。」

SalesforceによるAgentforceの導入は、分かりやすい例です。AIは数百万件の顧客対応を処理し、サポート担当者がより複雑な案件に集中できるようになりました。解決率は時間とともに改善し、人へのエスカレーションは減りました。経営陣は、AIの活動から財務上の影響までを、かなり明確にたどることができました。

SalesforceでのAIの社内導入からも、導入効果の測定が一筋縄ではいかないことを示す教訓が得られました。当初は、従業員がツールの使い方に戸惑い、効果も限定的でした。利用が定着すると、人員を企業内の別の部門に再配置できることが分かりました。これは、人員の削減としても、支出報告書の項目としても表れませんでした。実質的な効果は、既存の従業員をより価値の高い業務に再配置することで、外部採用を避けられた点にありました。また、再配置された人材はすでに事業と顧客を理解していたため、コストの削減と利益率の改善だけでなく、顧客体験の向上にもつながりました。

AIの価値を捉える方法は、ほかにも数多くあります。節約できた時間よりも、さらに財務成果に近い指標を使うものです。例えば、開発や営業の余力の拡大による追加収益、手作業による対応の削減、スループットの向上、顧客の待ち時間の短縮、顧客離反の抑制などです。重要なのは、導入前に成果を定義し、ベースラインを測定しておくことです。

導入前に「AI価値契約」を作成する

このギャップを埋める鍵は、導入を始めてからではなく、始める前に価値契約を定義することです。そのプロセスには、常に3つのステップを含める必要があります。

1. **価値を定義する** — どの指標を重視するかについて、明示的に合意します。P&Lに直結する数値でも、余力の指標でも、顧客体験や従業員に戻された時間でも構いません。大切なのは、追求する成果を明確にすることです。

2. **現状をベースラインとして確立する** — 対象とするワークフローについて、現状の数値をできるだけ明確に把握します。これがあって初めて、AIが機能しているかどうかだけでなく、どれだけ改善し、そのためにいくらかかったのかを測定できるようになります。完璧でなくても構いませんが、信頼できる数値でなければなりません。

3. **成果の責任者を明確にする** — 成果に責任を負う経営幹部を指名し、全社的な実行権限を与えます。多くの組織ではこれが最も難しい部分のようですが、極めて重要です。責任が曖昧なまま委員会やタスクフォースに分散すると、結局、導入は利用の指標で評価されることになります。より確かな指標を測ると決める、十分な権限を持つ人が不在のためです。

業界は確かに、正しい方向に向かっていると見ています。契約更新や投資をめぐる議論では、利用実績よりも本質的な指標に焦点が移りつつあります。例えば、採用の必要人数の変化、営業チームの成約率の向上、顧客離反の低減、処理1件あたりのコストの低下です。ベンダーも顧客も、どのモデルを使い、どのツールを避けるかの見極めが洗練されてきました。成果だけでなく、処理そのものの効率にも目を向け始めています。

価値のある成果のすべてが、損益に表れるわけではありません。結局のところ、価値の大部分は、多くの組織で見えていない領域に残り続けるでしょう。その大きな一つが、従業員が定型業務から相当な時間を取り戻すことの、本当の意味です。AIにはまだ代替できない、真の判断力や創造性、関係構築を要する仕事に人が力を注げるようになること。その価値は、スプレッドシートには表れません。価値が生まれるのは、組織が望む成果を決め、その成果を生み出す形で事業を変え、価値が実現するまでやり遂げたときです。

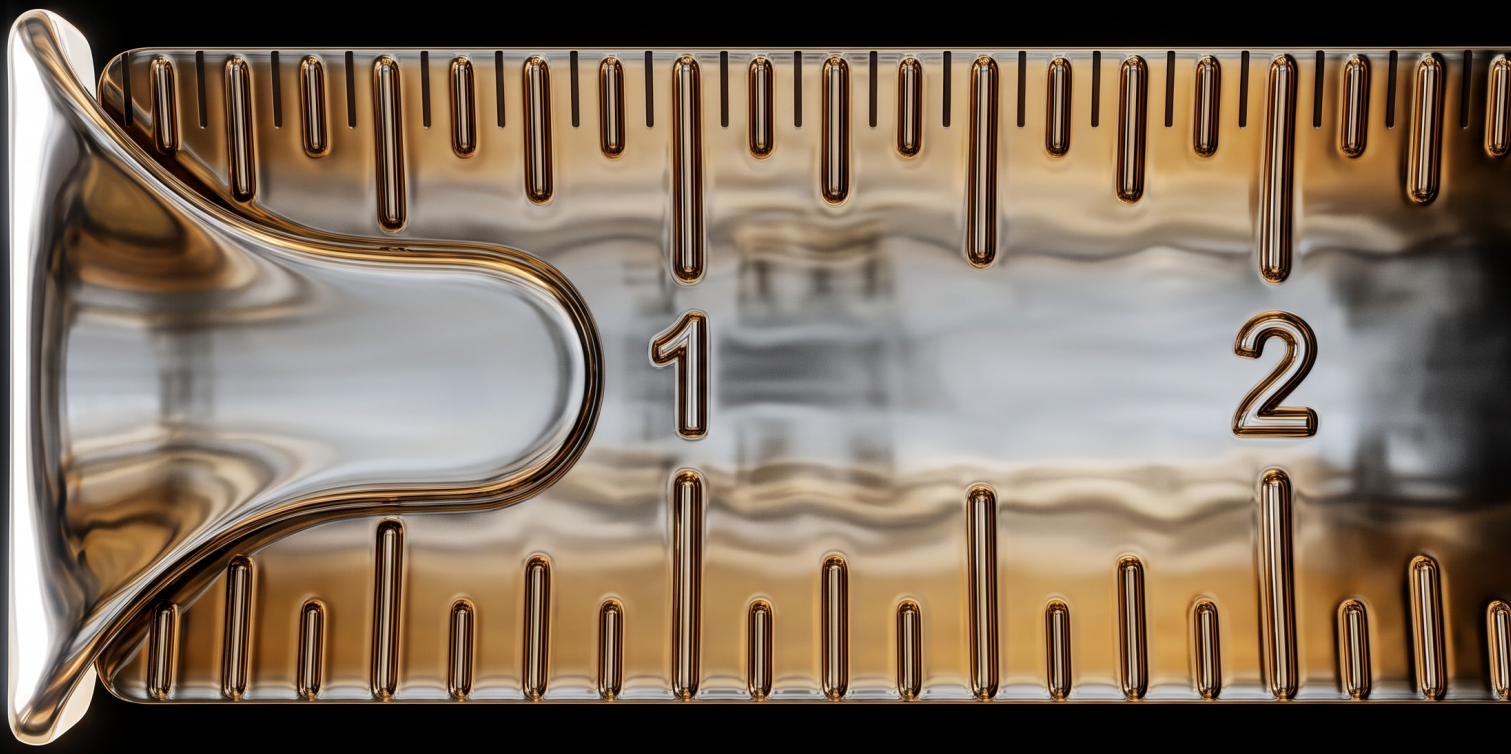


Divya Ashok
Altera Ventures 創業者、Salesforce AI Research 元最高執行責任者（COO）

詳しくはこちら

第3章
進捗は、何を基準に測られているのか

By Yiannis Antoniou
Lydatum 最高技術責任者（CTO）兼 AI・データ統括責任者



GLG

AI導入に取り組む組織で、私が最もよく目にするつまずきは、技術が「機能する」ことを示せるかどうかとは、ほとんど関係がありません。その点は、むしろ容易な部分です。チームが技術を構築すれば、遅かれ早かれ何らかの変化がもたらされ、何らかの目的に向けて「機能する」ようになります。多くの企業が価値の証明に苦労するのは、もう一つの基本的な問い、「進捗を、何と比べるのか」に答えるのが難しいからです。

私は、導入前の状態を記録していなかったために失速したり失敗したりしたプロジェクトを、数多く見てきました。そうしたプロジェクトでは、導入の動機はたいてい、技術の目新しさや性能への漠然とした関心と、「近代化」しなければならないという切迫感に行き着きます。そのような環境で、リーダーが「自分はどの事業成果を改善しようとしているのか」と真剣に自問することは、めったにありません。構築に入る前にその問いに答えられなければ、どれほど先進的な製品を作り上げても、価値の実証にはつながりません。

「失敗に至る場合の多くは導入の動機が新技術への漠然とした関心や、近代化への切迫感にとどまっています。」

投資リサーチを例に考えてみましょう。ある企業に、特定のセクターを担当するアナリストチームがあり、調査できるのはその分野の企業の5%にとどまっているとします。このチームにとって、「AIリサーチツールを導入した」ということは、事業成果にはなりません。「同じ人数のアナリストが、同じ時間で、10倍の数の企業を調査した」と言えて初めて、事業成果になります。これは、導入前後の変化が明確かつ測定可能で、事前に合意し、追跡し、投資効果を問われたときに示すことができます。しかし、ほとんどの企業は、いまだにそのような観点で考えていません。だからこそ今、このギャップを取り上げているのです。

なぜベースラインが存在しないのか

私が関わるクライアントの多くでは、技術部門と事業部門が緊密に連携していない場合に、ベースラインの欠如が生じます。独自のスケジュールと経緯を持つエンジニアリング主導のプロジェクトでは、事業成果とのつながりがまったくないことも少なくありません。さらに、技術部門と事業部門の足並みがそろっている、運営の行き届いた組織でも、事業側がベースラインの測定をまだ考えていないことがよくあります。ベースラインの設定は、「この作業にX時間かかっており、それをYだけ減らしたい」という程度の、単純なもので構いません。

「多くの企業では、ベースラインに必要なデータが存在しないか標準化されていません。」

しかし、問題の根本は組織内の足並みの乱れよりも深いところにあることが分かりました。私が訪れるほぼすべてのクライアント環境で、ベースラインの確立に必要な基礎データがそもそも存在し

ないか、信頼できるほど標準化されていません。企業は、自社のデータ資産の品質やガバナンスを過大評価しがちです。「顧客」や「製品」が何を指すのかについて共通の定義がなければ、ベースラインは確立できません。データの発生源や保存方法によって複数の定義が混在している場合も同様です。

AIは、この弱点をとりわけ浮き彫りにしました。データ資産が分断され、統制が効かず、所有者もまちまちで、共通の定義にも沿っていない組織では、AIへの投資が成果に結びつかないからです。裏を返せば、AIは、たとえ半ば強いられる形であっても、データ資産を刷新し、統制する重要な機会になります。データ資産をAIやエージェントの文脈情報として活用し、本当のベースラインを測定できる基盤を築く機会でもあります。この問題に気づくベンダーは増えており、AIがデータを技術面だけでなく、実際の事業の文脈でも理解できるよう、セマンティック（意味・文脈）レイヤーへの投資を進めています。

組織内の足並みの乱れは、データの問題に、さらに別の課題を加えます。技術部門と事業部門は、常に見解が一致するとは限らず、同じインセンティブで動いているわけでもありません。エンジニアリング主導のプロジェクトは、技術的に優れていながら、組織の中では存在が見えていないことがあります。

さらに事態を悪化させるのは、何かが機能していることを証明するよう迫られた組織が、事業上の本当のベースラインがないまま、何らかの進捗指標を無理に当てはめようとすることです。そうした組織は、空白を埋めるために利用の指標に頼りがちです。例えば「対象ユーザーの80%が先月このツールを利用した」といった数字が、財務上の効果の証拠と取り違えられてしまいます。利用の浸透もエンジニアリングの卓越性も確かに重要ですが、どちらも価値を直接示すものではありません。根本的には、AIであれ、ほかの取り組みであれ、その達成を助けるものとして、組織に共有された事業目標が必要です。その目標は、明確な構成要素に分解し、構築を始める前に指標を設定して合意しておく必要があります。社内の利用率やエンジニアリングのマイルストーンだけでは、十分ではありません。

「利用の浸透度も技術力も重要ですが、それだけでは価値を示せません。」

また、利用の浸透そのものが、主に組織の問題を表しているという見方にも、異論を唱えたいと思います。人々が実際に価値を得られるなら、利用は自然に広がるはずです。「AIを使わなければならないから、全員がAIを使うことにする」とトップダウンで強制されるのであれば、そもそも事業上の根拠がなかったことの表れです。

早い段階で現れる危険信号

私の経験では、技術の性能にかかわらず、構築のごく早い段階、早ければ最初の1週間で、価値の根拠を後から示すのが難しくなる兆候が現れます。

第一の兆候は、関係者と簡単なデモや要件確認のセッションを行った後でも、何をもって成功とするのかを関係者自身が説明できない場合です。スケジュールや技術的なアプローチに不確実性があるのは

普通のことですが、事業の観点で「うまくいった」とは何を指すのかを説明できないことは、深刻な問題です。

第二の兆候は、より気づきにくいものです。それは、改善が、プロジェクトに関与していないチームの行動変容に依存している場合です。指標を達成するには、特定のチームがAIの機能を使うためにプロセスを変更する必要があるのに、そのチームは議論に加わっておらず、当事者意識も持っていません。これでは、周囲から切り離された状態で構築を進め、自分たちでは制御できない依存関係を抱え込むことになります。技術的には「完璧」なシステムを作り上げたあとで、一度も関与しておらず、優先する理由もない別のチームに、利用やプロセスの変更を頼っていると気づくこともあります。このパターンが原因で、ほかの点では順調に進んでいた多くの導入が頓挫するのを、私は見てきました。

承認を出す前に問うべきこと

経営陣は、社内外のどちらのリソースを使う場合でも、大規模なAI導入に1ドルでも投じる前に、2つの問いに答えられなければなりません。

第一の問いは、「事業成果として何を達成しようとしているのか、測定可能な成功基準は何か」です。事業の観点から端的に言えば、他のあらゆる使い道と比べて、なぜこの資金を投じるのか。最終的に、指標は実際にどのような数値になるのか。これらに明確に答えられれば、進める根拠が得られます。あるいは、その根拠がないと気づき、リソースを無駄にする前に立ち止まって、仕切り直すこともできます。

第二の問いは、「適切な人々が議論に加わっているか」です。組織としての足並みを揃え、実質的な後ろ盾を得るには、行動やプロセスを変える必要が生じるチームからの真の賛同が欠かせません。着手前に賛同を得ておかなければ、後になって賛同を求めて苦労することになります。その時点で問題を解決するには、はるかに大きなコストがかかります。

この2つの問いは、一見シンプルです。しかし、構築を始める前にこれらに明確に答えられる組織こそが、価値を手にし、それを実証できるようになります。



Yiannis Antoniou
Lydatum 最高技術責任者（CTO）兼 AI・データ統括責任者

詳しくはこちら

第4章
AIの経済構造はSaaSの常識を覆す

By Mallika Iyer

AIインフラストラクチャ専門家、Google Cloud AI および
CoreWeave 元エンジニアリングリーダー、
Arya AI 創業者



GLG

多くの経営者は、今なお、SaaSを導入するのと同じ感覚でAIの導入に臨んでいます。ソフトウェアであれば、ライセンスの取得と構築にかかる費用を支払った後は、利用を重ねても追加コストはほとんど発生せず、価値は時間とともに着実に積み上がります。一方、AIの経済構造はそれほど単純ではありません。推論リクエストのたびに新たなコンテキストが参照され、GPUの処理時間が消費されてコストが発生します。つまり、AIは「うまく機能していながら、コストも膨らむ」という状態になり得るのです。この点を見落とすと、コストの増加を失敗のサインと受け止めてしまいがちですが、実際には利用が拡大していることの表れです。この取り違えは、その後の重要な評価・測定の判断にも影響を及ぼしかねません。

ここに、エンタープライズAIにおけるROIギャップの、最も見過ごされている要因があります。多くの組織が、AIには当てはまらない前提に立った基準で、成果を測定しているのです。過去10年以上、SaaSでの経験がテクノロジー投資の成否を判断する拠り所となってきた以上、多くの経営者が今もその発想から離れられないのは無理もありません。しかし、AI時代においては、価値を見出し、それを実証するための新たなモデルが求められています。

「SaaSの評価基準のまま、AIの成果を測っていませんか。」

KPIとの紐づけが成り立たなくなる理由

従来のソフトウェア導入では、因果関係が見えやすく、成果の帰属を比較的明確にできました。例えば、手作業の営業プロセスをCRMに置き換えた場合、インバウンドの問い合わせ件数、対応時間、ARR予測の精度といった指標を追跡し、その改善を特定の機能に結びつけて示すことができます。

一方、AIの出力は非決定論的です。同じ入力でも、毎回同じ出力が得られるとは限りません。そのため、KPIの改善をAI機能に直接結びつけることはできません。これは測定上の課題である一方、AIの価値の源泉でもあります。モデルは、固定された単純な指示を実行するのではなく、コンテキスト情報を活用して新たな出力を生成するからです。それゆえ、KPIが直線的に帰属することを前提とした費用対効果分析は、実態を歪めて映します。AIの効果が明確な帰属関係で説明できるという前提でROIの根拠を組み立てても、詳しく検証すれば長くは通用しません。

「AIの出力は毎回変わるため、KPIの改善を機能に直接結びつけることはできません。」

さらに、企業はAIの成果を、従来型の技術導入のために設計されたOKRやKPIで測りがちです。指標そのものが誤っているわけではありません。ただ、AI導入に特有のタイムラグや変動性に合わせて調整されていないのです。測るべきは、AIに関連する成果が変化しているか、それはどの時間軸で起きているか、そして総コストはいくらか、という点です。そこから、総保有コストの問いが浮かび上がります。

総保有コストの問題

AI導入を計画する際、多くの組織は、初年度にかかるホスティング費用を算出し、2〜4年目は保守費用を見込むというコストモデルを用います。従来のクラウド型アプリケーションであれば、この考え方でも通用するかもしれません。しかし、本番環境でAIを運用するコストについては、このモデルでは過小評価されてしまいます。

適切な計画では、例えばモデルドリフトを考慮する必要があります。モデルドリフトとは、運用を続けるうちに実際に入力されるデータの傾向が学習時とずれていき、モデルの性能が徐々に低下していく現象です。自社のユースケースで競争力と有効性を保つには、モデルの評価、ファインチューニング、そして定期的なアップグレードを継続的に行わなければなりません。そのため、人による介入にかかるコストは変動し、モデルが成熟するにつれてむしろ増える場合もあります。さらに、モデルがうまく処理できない例外的なケース（エッジケース）に対応する評価・テストの仕組みを構築する作業にも、継続的な工数とコストが発生します。リーダーは、総保有コストを検討する際にこうした要因を見落としがちです。AIがすべてを自動化するのだから初期導入が済めばSaaSのようなコスト構造になる。そう考えてしまうのです。

期待する成果の捉え方でも、同様の誤りが見られます。組織は複数の成果を同時に実現できると考えがちで、コスト削減、収益拡大、余力（キャパシティ）の創出、リスク低減といった期待リターンを、単一のROI指標にまとめてしまう傾向があります。しかし実際には、ある成果を追求すれば、別の成果とのトレードオフが生じることが少なくありません。例えば、収益拡大を狙う施策ではキャパシティを追加で消費し、人による監督もより多く必要になります。また、創出された余力は、活用方法を明確に決めない限り、自動的にコスト削減には結びつきません。だからこそ、AIが4つの成果をすべて同時にもたらすと期待するのではなく、どの成果の最大化を目指すのかをあらかじめ具体的に定めておくことが重要です。

生産性向上が真に意味すること

「削減できた時間は、そのままコスト削減や新たな収益につながる」。この思い込みこそが、企業のAI活用で最も多い測定ミスと言えるでしょう。とりわけ、取締役会で「その投資は損益計算書のどこに表れているのか」と問われたときに、この問題が顕在化します。

コード生成を例に考えてみましょう。AIコーディングエージェントでソフトウェア開発を支援すると、ほとんどの場合、導入初期は人員が増えます。モデルのエッジケースがまだ把握できていない段階では、生成されたコードを評価する「人間による確認」の担い手が、より多く必要になるためです。AIがコード生成で節約した時間は、自分たちが書いていないコードをレビューする時間で相殺されます。その後、モデルの弱点が見えてきて、それを補う評価ツールが整備されるにつれ、実装のスピー

「節約した時間は、そのままコスト削減や新たな収益になる。これは、AIの効果測定でよくある思い込みです。」

ドは上がっていきます。ただし、その過程では、削減効果が現れる前に、まずコストが増加するというカーブを描きます。初期段階だけを測定した組織は、誤った結論に至るおそれがあります。

カスタマーサポートでも同様の動きが見られますが、その現れ方は異なります。十分なコンテキストと明確なガードレールを備えたエージェント型AIを適切に導入できれば、人員とコストを削減しつつ、顧客の離反も減らすことができます。ただし、この成果は、導入初期にガードレールを適切に設計できていることが前提です。初期設計で成功する可能性はありますが、それを当然と考えるべきではありません。多くの場合、システムの調整が進み、エッジケースが表面化するなかで、コストはいったん増加し、その後に減少へ転じます。

利益率の改善という観点では、リーダーは利用量の多さを、成功の証拠ではなく前提条件として捉えるべきです。例えば、ユーザーがAIとやり取りを重ねても、問題解決につながらず、ユーザーにも企業にも経済的な便益が生まれないのであれば、利用が増えるほどホスティングコストが増えるだけで、利益率は向上しません。利益率が改善するのは、次の2つが同時に成り立つときです。1つは、AIモデルの生産性が高いこと。もう1つは、その生産性が経済的な価値につながっていることです。つまり、本来は人手が必要だった問題や、これまで解決されてこなかった問題を、AIモデルが解決できている場合を指します。

測定方法と、導入前に問うべきこと

測定の出発点となるのはベースラインですが、多くの組織が設定するものよりも、具体的なものでなければなりません。「導入前」の状態が明確であってこそ、「導入後」の変化を正しく測定できるのです。例えば、人間のカスタマーサポート担当者をエージェント型AIに置き換えるケースを考えてみましょう。何件の問い合わせが、どの期間に、どのコストで、どの解約率のもとで解決されたのか。解決時間は変わったのか。解約は増えたのか、減ったのか。AIが対応できずエスカレーションされた案件は、解決までに平均でどれだけの時間を要したのか。こうした指標は、AIが関与する工程だけでなく、ワークフロー全体を対象に追跡するものでなければなりません。

「導入判断を
検証するために、
あえて異論を
唱えてみましょう。」

私は、ほとんどの組織が実践していないベストプラクティスを推奨しています。それは、導入の意思決定そのものに、あえて異論を唱えてみることです。例えば、チームが経営陣に「特定の領域で優位を強める競合他社に対抗するためにAIを導入する必要がある」と主張してきたとします。このとき返すべき問いは、「30日遅れて導入すると、実際に何を失うのか」です。待つことで生じる現実的なコストと、ROIの根拠が十分に整理されないまま導入し、意味のある経済的便益が得られない領域に投資するコスト。この両者を比較します。導入案がこの基準を満たすのであれば、価値を実証できる可能性は高いといえます。

しかし、この問いに向き合おうとするリーダーは多くありません。何を失うのかを正確に答えられる人は、実のところ誰もいないことが多いためです。この不確実性がFOMO（取り残されることへの不安）を

招き、AI導入の多くは、リスクが低く安心できる反面、価値が小さすぎてP&Lには表れないユースケースにとどまっています。

こうした事態を避けている企業は、導入前に、極めて細かく精査する姿勢を持っています。特定のアプリケーションのAI機能が、生産性、収益、運用コストにどう影響するのか。AIの誤りを修復するための人的コストも含め、そのインパクトが実際にいくらになるのか。これらを具体的に把握しようとします。

そこまで詳細なデータを求めるのは気が引けることもあり、実際に用意するのも常に難しいものです。しかし、そこまで把握して初めて、ROIについて答えを出すことができます。



Mallika Iyer
AIインフラストラクチャ専門家、Google Cloud AI および
CoreWeave 元エンジニアリングリーダー、Arya AI 創業者

詳しくはこちら



第5章
最も価値の高いAIほど、実現までに時間を要する理由

By Dr. Ryan Mackey

LifePoint Health 最高医療情報責任者（CMIO）兼
AI／RCM／CDI 担当医師役員

GLG

医療分野におけるAIのROIを現実的に議論するには、まず医療システムの3つの階層、すなわち臨床、財務、運営（管理）を理解する必要があります。AIの成功の形は階層ごとに少しずつ異なり、それぞれ異なる時間軸で、特定の指標を通じて現れます。

財務と運営の階層では、成功のシナリオは比較的明快です。収益サイクル管理（RCM）にAIを適用すれば、請求プロセスが効率化し、保険請求の拒否率が下がるはずです。運営面では、患者エンゲージメントの向上、コンプライアンスの強化、日常業務の効率化が期待できます。こうした成果を裏付ける指標は早期に得られ、ダッシュボード上でも明確に示せます。医療システムが現在これらの領域でAIを積極的に導入しているのは、ROIをある程度の確信をもって説明できるためです。

一方、臨床面では状況が異なります。成果はより主観的で、表れるまでに長い時間を要する場合があります。しかし、AIが最終的に最も大きな成果をもたらすのは、この領域だと私は考えています。なぜ臨床面の測定が難しいのかを理解することは、AI投資に現実的な期待を設定したい、この分野のリーダーにとって欠かせません。

臨床領域が異なる理由

臨床AIの成果は、患者の健康回復という形で具体的に現れ得ますが、その効果だけを切り分けて定量化することは困難です。医療の質、医療従事者のバーンアウト（燃え尽き症候群）、疾患の早期発見率は、診療報酬コードやスループット（処理量）の指標には直接反映されません。より多くの患者を診ることや、より報酬の高い診療報酬区分で請求できるようにすることだけが、臨床AIの真の有効性とインパクトを示す唯一のROIではありません。

臨床では、因果の連鎖が長く、利用者も多様で、ケア環境も絶えず変化します。臨床AIが関わるのは、外来、救急、入院、そして代替ケア環境であり、それぞれ進むペースが異なります。複雑性が低い環境ほど、成果の裏付けは早く得られる傾向があります。外来での単純な記録作成ツールは、救急部門で診断推論を支援するAIシステムよりも、早く価値を示せる可能性が高いでしょう。適切な導入計画には、こうした臨床業務に固有のばらつきを最初から織り込む必要があります。

測定は難しいものの、医療に前向きな変革をもたらす最大の機会は、臨床AIにあります。エビデンスの方向性はすでに有望で、早期発見、ケア手法の改善、より意義のある患者エンゲージメントといった、可能性の高い重要分野で進展が見られます。ただし、AIにはまだ成熟の余地が大きく残っています。機械学習システムは、フィードバックの蓄積によって精度を高めますが、成果に意味のある変化をもたらすだけのシグナルが集まるには、時間がかかります。臨床現場でのAI活用について、より広い合意と認識が形成されるまでは、分野によって進展の度合いに大きな差が生じます。影響が最も大きい用途、すなわち急性期医療の複雑性や患者への直接的な影響を伴う用途ほど、実用化に時間を要するでしょう。

人員配置の現実

医療分野のAIには、根強い誤解があります。AIが効率化の余地を生み、同じ業務をより少ない人数でこなせるようになる、という見方です。しかし、医療機関の人員体制を実際に見ると、この前提には無理が

あることが分かります。

「AIの重要な役割は、管理・運営の負担を軽減し、スタッフを専門性を活かす業務に集中させることです。」

医療現場に、AIが代替できるような余剰人員はいません。一般的な医療機関では、かつて1〜2の役割を担っていた人が、いまや3〜4の役割を兼ねています。医療分野の労働力不足が近いうちに解消すると考える根拠も、ほとんどありません。AIに期待できる最も重要な役割は、スタッフがこれまでやむを得ず抱え込んできた管理・運営上の負担を軽減し、本来の専門性を活かせる業務に集中させることです。

AIが、スタッフが本来担うはずのなかった業務を引き受けることで、人員削減ではなく、再配置の機会が生まれます。臨床医は、専門性が求められる、患者に向き合い判断を要する業務に充てる時間を取り戻せます。管理スタッフは、定型処理ではなく、例外対応やエスカレーションに集中できます。AIは、業務上のギャップを埋める形で導入すべきです。さらに重要なのは、AIモデルが進化し続けるよう、スタッフが監督と制御を行い、フィードバックを反映できる体制を整えることです。

AI導入は実際にどう進むのか

AIツールの導入から、経営層が重視する指標に変化が現れるまでの期間は、前述の階層（臨床・財務・運営）によって大きく異なります。臨床面では、迅速に導入・活用できるツールであれば、3〜4週間で現場への効果を明確に確認できます。一方、収益サイクル管理（RCM）の導入には90日近くかかります。AIが有効性を証明するには過去データの取り込みが必要で、さらに既存プロセスへの統合を経て改善効果が積み上がるからです。しかし、いったん稼働すれば、RCMの領域は安定して運用される傾向があります。関わる利用者が少なく、反映すべき主観的なフィードバックも少ないため、改善サイクルを回しやすくなります。

パイロットが成功に至るまでの期間は短くなっており、あらゆる階層で、導入効果を説明しやすくなっていることがうかがえます。従来、医療分野のテクノロジーのパイロットは、統合リスクへの慎重さや成果検証の不確実性から、評価フェーズが長引いて停滞することがよくありました。一方、AIでは、導入の判断に確信を持ちやすくなっています。その一因は、ROIで何を達成するのかが明確であり、解決する課題も、医療システムが不可欠と認識しているものだからです。ある地域（マーケット）で成功したパイロットが、速やかに他の5〜6地域へ展開されることも少なくありません。

「パイロット成功までの期間は短くなり、価値を説明しやすくなっています。」

この加速を可能にするのは、地域の推進者の存在、明確なマイルストーン、ベンダーとの連携の組み合わせです。各地域には、期待値を調整し、現場の声を吸い上げ、組織内で導入を後押しできる、主体的な臨床または運営のリーダーが欠かせません。実データに基づくマイルストーンは、取り組みの進捗を実態に即して示し、経営層に、曖昧な進捗報告ではなく、意味のある確認の機会をもたらします。一方、ベンダーには、製品を売るだけでなく、拡張性を支えるパートナーとしての関わりが求められます。ツールの成長に合わせて運用を支え、データが示す結果に応じて機能を進化させ、初期導入の後も伴走し続けることです。

現実的な期待値の設定

大規模なAI導入を承認しようとする経営陣が目を向けるべきは、技術の選定よりも、組織の準備度です。必要なのは、全社の経営層と各地域の推進者（臨床、運営、またはその両方）が一体となって、主体的に関わるチームです。各地域の推進者は、意思決定の後に助言を求められる存在ではなく、最初から取り組みの一員として加わっていることが前提です。意思決定に各地域が加わっていないと、導入は当事者の納得を得られないまま始まり、成果に結びつかずに終わりがちです。

明確なマイルストーンに加え、各階層（財務、運営、臨床）で何を測定するのかを整理しておくことも欠かせません。成果の裏付けが得られるまでの時間軸はそれぞれ異なり、それを混同すれば、全員が失望する結果になります。さらに、臨床AIが成熟するまでの忍耐も必要です。そのためには、AIはスタッフの業務を奪うものではなく、それぞれが最も得意なことをより効果的に行えるようにするものだと、伝えることが重要です。

医療におけるAIが真の変革をもたらすのは、臨床ケアの現場です。財務と運営の成果はすでに現れ始めていますが、今後12〜18か月で見込まれる変化の土台にすぎません。利用者層もケア環境も非常に多様であるため、臨床AIには時間を要します。しかし、フィードバックが積み上がり、AIモデルが成熟するにつれて、あらゆるケア環境と職種の臨床医、さらに患者体験と成果の各側面にまで影響を及ぼす可能性は、現実味を増していきます。

AI活用で最大の成果を得るのは、早い段階で期待値を現実的な水準に調整し、導入を場当たり的ではなく計画的に進め、価値が実を結ぶまで粘り強く取り組み続けた医療システムです。



Dr. Ryan Mackey
LifePoint Health 最高医療情報責任者（CMIO）兼 AI／RCM／CDI 担当医師役員

詳しくはこちら

おわりに

本レポートの各論点は、AI導入が価値を実証できなくなる要因をそれぞれ一つずつ取り上げています。これらを併せて見ると、企業全体で価値を実証することが、一つの障壁を越えれば済む話ではなく、複合的な課題になっている理由が見えてきます。成功の定義、成果を測る基準、適用するコストモデル、実証に許される時間軸のそれぞれが、独立したつまずきの原因になりうるからです。技術的に健全な導入であっても、そのいずれかが欠ければ価値を示せません。一つを乗り越えても、残りも乗り越えられるとは限りません。企業全体のROIが、技術そのものが機能しているかどうかよりも答えにくい問いになるのは、このためです。

エンタープライズAIの導入に、特効薬はありません。その変革をもたらす可能性を引き出すには、時間とリソース、そして適切なマインドセットが欠かせません。必要なのは、意欲と謙虚さ、創造性をあわせ持つ姿勢です。これは経営層から現場の担当者まで、組織全体に求められます。投資の成果を証明する過程には、この取り組みの複雑さと規模がそのまま表れます。組織もチームもリーダーも、直面する障壁はそれぞれ固有であり、成功の条件も一様ではありません。

本レポートを支える5名の有識者は、世界中から厳選されたメンバーで構成される「GLG AIリーダーシップ・カウンシル」の一員です。同カウンシルは、AIがクライアントにもたらす課題と機会に特化した、当社初のエキスパート・リーダーシップ・カウンシルです。メンバーはいずれも、AIインフラとコンピューティング、基盤モデルとエージェント型システム、企業におけるAIの導入と変革、金融サービスとヘルスケアにおける業界特化型の活用、そしてAIのリスク、政策、投資戦略といった幅広いテーマにおいて、実務を通じた知見を有しています。

AIのROIギャップを埋める道筋は、組織によって異なります。GLGは、世界をリードする信頼性の高い「人の知見」のプラットフォームとして、こうした複雑な課題に向き合うための知見を、あらゆる段階でお届けしています。エキスパートインタビューや専門家を対象としたBtoBの定量調査に加え、プロジェクトへの有識者の参画（伴走支援）、M&Aや投資の局面でのディール・アドバイザーとしての参画など、ご要望に応じた形でご支援いたします。また、AIリーダーシップ・カウンシルのメンバーによる、貴社の課題に合わせたワークショップや研修、社内外のイベントでのパネル登壇や基調講演のご依頼も承っています。

AIリーダーシップ・カウンシルについて詳しくはこちら

お問い合わせ

株式会社Gerson Lehrman Group
〒105-6226 東京都港区愛宕2丁目5番1　愛宕グリーンヒルズMORIタワー26F
TEL: 03-4589-1500 Email: info.japan@glgroup.cpm
公式Website: https://glginsights.com/ja/