



# エージェント時代の Build → Measure → Decide ループ

エージェント駆動チームのための「意思決定の頭脳」。  
作ることは安くなった。残す価値を知ることは、まだ高い。

稼働中 — セルフホスト型MVP

カスタマーゼロ — 自社経営をFlowで運営



# 作ることは、安くなった。 知ることは、まだ高い。

AIエージェントは、どんな人間のチームにも不可能な速度でコードを出荷できるようになりました。その結果は、雰囲気で作られたアウトプットの洪水 — 直感で出荷される機能、対照のない実験。ボトルネックは「作ること」から「判断すること」へ移りました。

## 雰囲気ドリブンの出荷

直感で出荷される機能。対照のないまま走る実験。  
「自分の環境では動く」が、スケールして起きている。

## 較正された判断の不在

エージェントの出力に溺れ、本物のシグナルとノイズを区別する手段を持たない創業者たち。

## エビデンスより思い込み

スタートアップの多くは、検証ではなく信念の上で反復して死ぬ — 仮説を試す代わりに、雰囲気を守り抜いてしまう。

**ビジョン:** 会社を「未採点 — 価値が主張されるだけ」の状態から、「採点済み — 現実に判定される」状態へ動かす。次の一手を自明にする。

# 本物のシグナルを、自動で収集する

エージェントは、出荷したものに自ら計測を組み込みます。テレメトリが流れ込み、システムは各シグナルの信頼性を推定します — 指標を並べるだけでなく、どこまで信じてよいかをスコアリングします。

App Store 評価

4.2 ▲0.4

127件の評価

D7 継続率

38% ▲6pt

先週は32%

DAU

2,840 ▲12%

前週比

シグナル信頼性

高

p = 0.82 ・ 収束中

機能利用率

オンボーディング

78%

ソーシャル共有

52%

ダークモード

31%

AIレコメンド

14%

顧客の声

★5 「次に何をやるべきか、ようやく決められるアプリ」

★4 「便利。ただしオンボーディングは少し長い」

★5 「数ヶ月を無駄にするとところだった機能を3つ止められた」

★2 「AIレコメンドがランダムに見える」

現在動いているもの

エージェント割り当て・ワークツリー分離・差分レビュー・マルチエージェントチェーン・cron自動化・エージェントREST API・日英バイリンガル

この画面はロードマップ — クレジットが実現します

自動テレメトリ・シグナル信頼性スコアリング・校正された意思決定レイヤー

Flow・Series T — Post AGI from Kyoto

PhaseShift, Inc.

# ダッシュボードではなく、意思決定を。

疎でノイズの多い初期シグナルを読み、信頼できる次の一手を返します — 実データが集まるほど狭まる、誠実な信頼区間つきで。

|                                                       |                                                   |              |      |
|-------------------------------------------------------|---------------------------------------------------|--------------|------|
| <div>オンボーディング v2</div> <div>シグナル</div> <div>本物の</div> | D7継続率 +12% (CI: 8–16%)・App Store評価 ▲0.4・機能利用率 78% | p = 0.82 収束中 | 強化 ↑ |
| <div>ソーシャル共有</div> <div>兆候あり</div>                    | 機能利用率 52%・バイラル係数 0.3・上昇傾向だがCIはまだ広い                | p = 0.61 上昇中 | 継続 → |
| <div>ダークモード</div> <div>データ不足</div>                    | 機能利用率 31%・継続率への影響は未観測・データはまだ5日分・CIが広すぎて判断不能       | p = 0.45 不確実 | 収集 + |
| <div>AIレコメンド</div> <div>ノイズ</div>                     | 機能利用率 14%・コンバージョン改善なし・★1「ランダムに見える」・9日間シグナル横ばい     | p = 0.09 横ばい | 中止 × |
| <div>プッシュ通知</div> <div>ノイズ</div>                      | エンゲージメント改善なし・オプトアウト率 41%・初期の山は物珍しさによるもの           | p = 0.11 横ばい | 中止 × |

ダッシュボードではなく、**較正された判断**: 継続 / 中止 / 転換 / 強化を、明示的な信頼区間つきで。信頼は、現実採点された実績から生まれます — それが私たちのRL研究が可能にするものであり、クレジットの使い道です。

# モートは生成ではない。校正された判断だ。

## コモディティ化するもの

- **生成** — 誰でも、何でも出荷できる
- **自動化** — どのチームも同じエージェントを持つ
- **配布** — 全員が等しく負けるコンテスト
- **ダッシュボード** — 信じられない指標は、増えても役に立たない

## 複利で効くもの

- **校正された意思決定** — 初期シグナルへの誠実な信頼区間
- **レビューデータ** — エージェントの仕事への構造化された人間の判断
- **実験横断のアウトカム** — 現実に採点される
- **実績** — モデルの自己申告ではなく、実現した結果からの信頼

## 技術的イノベーション

スタートアップの反復を、ダッシュボードの問題ではなく**校正の問題**として扱う。指標を増やすのではなく、各初期シグナルの信頼性を推定し — 本物の需要とノイズを見分け — 継続 / 中止 / 転換 / 強化の判断を、実データで狭まる明示的な信頼区間つきで出力する。現実の結果に対して継続的に採点され続ける、学習されたゲート。

**トークンが燃料である理由:** 反復のたびに推論が回る — エージェントの実行、シグナルの採点、並行実験それぞれの次の一手の生成。クレジットが多いほど、ループが増え、結果データが増え、判断が鋭くなる。**クレジットは特典ではなく、エンジンです。**

## Valliappa Chockalingam

創業者 & CEO・強化学習研究者

チューリング賞受賞者 Richard Sutton 教授（アルバータ大学）の下で修士号。杉山将教授（東京大学）の下で博士課程研究。研究テーマ: fixed-horizon TD、方策条件付き価値関数、successor features。

[github.com/geekstor](https://github.com/geekstor) ・ [jp.linkedin.com/in/geekstor](https://jp.linkedin.com/in/geekstor)

## 上田 勇希

共同創業者 & CPO・日本GTM

営業歴10年。将棋のオンラインスクールを設立・運営。プロダクトの方向性と市場戦略を担当。

PhaseShift, Inc. ・ 東京で法人設立・人間2人、エージェントで増幅し、エージェント駆動チームのための「意思決定の頭脳」を作る。

## POST-AGI ビジョン

認知が潤沢になれば、何かを作ること自体は些事になる。希少になるのは、**現実**に根ざした判断だ。

個人や小さなチームが、かつて会社を必要とした挑戦をできる未来 — 1年かけて思い込みを守り抜く代わりに、**数日で自分の間違いに気づける**未来。

より多くの生成ではなく、**より良い見極め**を: 自分の仕事についての真実を、行動できるうちに教えてくれるインフラ。