

Typed Decisions

判断だけを返すAI — TypeSafe AI「Jev」 実地検証レポート

文章を生成しない System One モデルは、業務システムの分類処理を置き換えられるか。Choice / Score / Noul の3つの型と、実データ1,600マスの判定実測から導いた設計原則

SUMMARY

本書の要旨

本書の数値は、当社が2026年9月20日から21日にかけて実データで計測したものです。モデルの仕様はベンダー公式ドキュメントに拠っています。出典は巻末に一覧で掲載しました。

TypeSafe AI が2026年9月に公開した System One モデル「Jev」は、文章を書かずに判断と確率だけを返します^[1]。当社はこれを不動産物件の自動分類に適用し実データで検証しました。分かったのは、結果を左右していたのがモデルの性能ではなく、渡す情報の作り方と答えの型の設計だったという点です。

- 01 資料の記載に反した判定は、採点できる344マスのうち**4件（1.2%）**でした。ただしこれを決めているのはモデルの構造ではなく、選択肢ごとに書いた規則です。規則を入れる前の版は記載違反0件でしたが、言語化済みの事業ルールには180件反していました。
- 02 判定の質を決めていたのは**入力の作り方**でした。規則を同じにしたまま、情報を解いてから渡すか生のまま渡すかを変えるだけで、記載に反する判定が0件と16件に分かれます。
- 03 判定1,600マスのうち**77.9%には、根拠になる記載がありません**。ここに「分からない」を表す型を用意しないと、判断のつかない物件が黙って検索結果から消えます。

4 件

資料の記載に反した判定
（採点可能な344マス中・1.2%）

事業ルールを入れた推奨構成。規則を入れる前は0件だが、事業ルールに180件反する

100 %

資料を書き換えたときに
判定が動いた割合（148/148）

記載の無いマスから物件ごとに1マスを抽出。可・不可の双方向で計測

0.06 %

同一の入力を2回流した
ときの判定のぶれ

推奨構成。1,600マス中1マス

CONTENTS

目次

01	判断だけを返すAI	System One モデルと3つの型	P.04
02	入力の設計	渡す情報の作り方が判定を決める	P.05
03	検証の設計	「精度」が出せないデータの測り方	P.06
04	出力の設計	「分からない」を型で表す	P.07
05	横展開	卸売業の受発注・商品マスタへの応用	P.08
	適用シナリオ	受注ファクスを一次仕分けする場合の実装	P.09
06	まとめ	3つの設計原則と、出典一覧	P.10

METHODOLOGY

調査概要

実施	2026年9月20日～21日
対象	事業用賃貸物件の自動分類（業種10グループ・こだわり条件）
使用モデル	TypeSafe AI「Jev」（jev-latest／実体は jev-1.13.0）。公式は英語が主たる学習言語でCJKは同等ではないと明記しており ^[4] 、本書はすべて日本語データでの実測です
前工程	図面・写真の読み取りは Google Gemini。Jev は画像を扱えないためテキストで渡します
データ	図面90件／流通データ987件（正解ラベル1,867個）／本番160件
判定数	7,896件（流通データ）／1,600マス×11パターン＋感度テスト444件（本番160件）
出典	ベンダー公式5件／官公庁2件／企業公表2件

本書の作成方針

- 数値は当社の実測値か、一次情報が確認できたものだけを掲載しました。
- 物件名・物件ID・流通サービス名は開示していません。
- 当社サービスの品質に関する個別の数値は、本書の範囲外です。

01 判断だけを返すAI — System One モデルと3つの型

SYSTEM ONE MODELS

TypeSafe AI は2026年9月15日、System One モデルと、その第1号「Jev」を公開しました^[1]。文章を生成せず、答えの型と確率だけを返します。学習方法は RLCD（Reinforcement Learning for Calibrated Decisions）です。

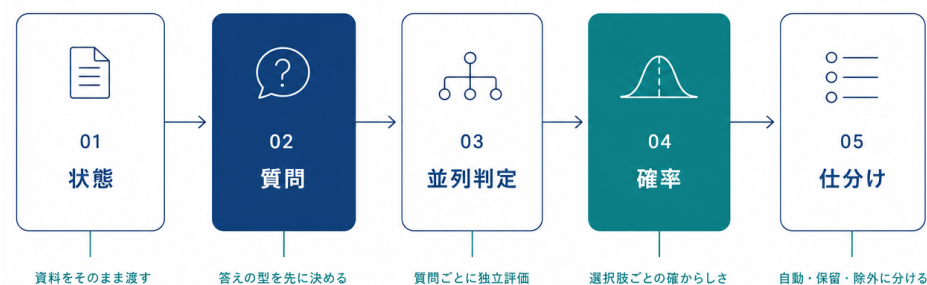


Fig. 01 — System One モデルの処理の流れ。答えの型を先に決めてから資料を渡す点が、文章を生成するモデルとの構造的な違いです。

FIG. 02 — JEV が備える3つの型（プリミティブ）と、公式が定める制約 ^{[2][3]}

型	公式の説明	返ってくるもの	制約と、今回の使い方
Choice	Choose an option from a list	選択肢・確率・確信度	選択肢は最大255個。10の業種グループと3つの条件の判定に使いました
Score	Score the state on a rubric	段階の値・確率・確信度	2～10段階。今回は未使用です
Noul	Is this statement true?	0～1の値	真偽を一言で聞くととき向けの型です

公式は System One モデルを a new class of frontier models built to make fast, structured decisions that software can use directly と定義しています^[1]。接続先は POST <https://api.typesafe.ai/v1/systemone>、上限は1リクエスト64kトークン（state と最長の質問で32k）、料金は入力100万トークンあたり \$0.042 で出力は無料です^{[2][4]}。質問は同じ state に対し並列かつ独立に一度で評価されるため^[3]、13問を1回にまとめています。

02 入力の設計 — 渡す情報の作り方が、判定をそのまま決める

DESIGNING THE STATE

Jev は推論の途中経過を作らず、一度で確率を返します。考えさせる余地が無いので、**こちらで解いてから渡すほど正確になり、生の文字列のまま渡すと読み違えます**。実データで3通りを比べました。

解いた情報を抜くと、総称の読み違いが戻る

業種ごとの可否を整理した情報を state から外し資料の記載だけを渡した版では、規則は同じまま、記載に反する判定が0件から16件に増えました。

「飲食店可」という総称から重飲食まで可と読む、要相談を可と読む、といった誤りです。

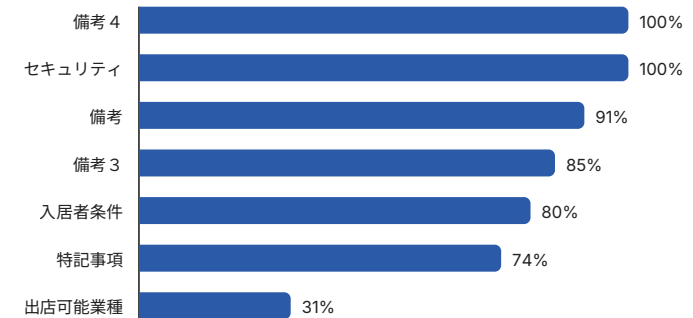
所在階は126通りの文字列があった

160件で所在階の文字列は126種類ありました。書式は14パターンで、残りは階数の組み合わせです。少数の規則で解けるのに、文字列のままでは一致しません。

本章の結論

解いた情報と、解いていない生の情報を**両方渡す**。前者は項目どうしの関係を保ち、後者は要約で落ちた事実を供給するので、役割が違い競合しません。

FIG. 03 — 判定工程まで届いていなかった情報の割合



対象160件。元データ側に値があるのに、要約工程を経た後には現れなかった項目の割合。

当社実測（対象160件）。読み取り工程が書き出した要約だけを判定に渡していたため元データにあった情報の多くが判定側に届いていませんでした。

要約に残らない事実がある

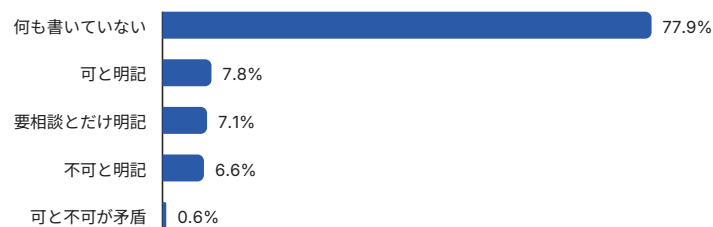
図面90件の検証では、整理済みの要約だけを渡した場合と、解釈しない粗い書き起こしを併記した場合を比べました。室内写真に写っていたシャッターは要約に残っておらず、併記して初めて判定が動いています（該当確率 0.00 → 0.94）。入力は6%増えただけで劣化した項目はありませんでした。

ただし公式は、質問に関係しない情報を state に入れると精度が落ちると明記しています^[5]。上は入力が6%増える範囲での実測であり、state が大きく伸びる場合は再計測が必要です。

03 検証の設計 — 「精度〇%」が出せないデータを、どう測るか

DESIGNING THE EVALUATION

FIG. 04 — 根拠になる記載が存在するか（1,600マスの内訳）



160物件 × 10業種グループ。上位4層の344マスだけが、記載に照らして採点できる。

当社実測。160物件 × 10業種グループ。正解が存在しない8割に対して一致率を計算しそれを精度と呼ぶことはできません。

そこで採点を3本立てにしました。①資料の記載に反していないか ②事業ルールに反していないか ③資料を書き換えたとき判定が正しい方向へ動くか。③は記載の無い8割を測る代わりになります。

FIG. 05 — 3本立ての採点結果（事業ルールを入れる前と後）

測ったこと	規則なし	規則あり
資料の記載に反した判定	0 / 344	4 / 344
事業ルールに反した判定	180	5
「可」「不可」を足すと判定が動く	—	各 148 / 148
同一入力を2回流したときのぶれ	5 / 1,600	1 / 1,600

感度テストは記載の無いマスから物件ごとに1マスを抽出した148件が対象。解析済みの可否欄に記載を足しているため、原文から読み取れるかまでは測っていません。ぶれの差は反復2回ずつのもので因果は言えません。

どちらか一方はゼロにできない

規則を入れる前は記載に一度も反しませんが、事業ルールには180件反します。規則を足すと逆になり、ルール違反5件・記載違反4件です。両方を同時にゼロにはできません。

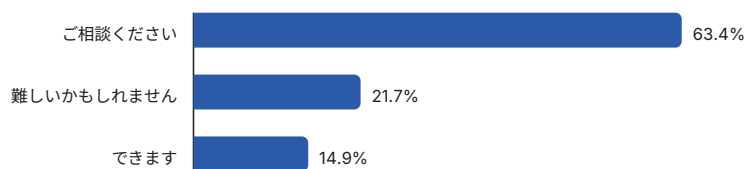
規則は、書く場所を間違えると効かない

「この選択肢を選ぶな」を別の選択肢の説明欄に書いたところ1件も動きませんでした。説明は「その選択肢を選ぶ条件」として読まれるためです^[5]。規則には「何に負けるか」まで書きます。

04 出力の設計 — 「分からない」を型で表す

DESIGNING THE ANSWER

FIG. 06 — 3値設計での判定分布（1,600マス）



2値設計では、この63.4%が「該当なし」に吸収され検索結果から消える。

当社実測。判定の6割強は、こちらでは決めきれないという帯に入ります。ここを表せる型が無いと、この帯は検索結果から消えます。

答えの型は当初4段階でしたが、3つに統合しました。「**不可**」という言葉は**使いません**。事業用物件では、断りの記載が無い限りたいていの業種が入れます。「絶対に無理」と読める言葉を置くと、問い合わせの芽まで潰しかねません。

なお、記載が1つも無い1,246マスのうち**345マス（27.7%）では断定が返ります**。用途区分や所在階から判断してよいと規則で決めた分です。記載が無ければ必ず保留になる、という作りにはなっていません。

FIG. 07 — 内部の値と、画面に出す言葉を分ける

内部の値	画面の表示	選ぶ条件
possible	できます	資料に可と明記／前業態が一致／必要な設備が明記
consult	ご相談ください	既定。落とす理由が見つからない
hard	難しいかもしれません	断る記載がある／設備・階・客導線と噛み合わない

語尾には確信度の差を持たせました。言い切る／相手に委ねる／濁す、の3段階です。表示の言葉だけを差し替えられるよう、判定ロジックは内部の値で持ちます。

2値で持つと何が起きるか

「該当する／しない」の2値では、判断のつかないものが「該当しない」に吸収されます。落ちたことに誰も気づけないまま、その業種を探しているお客様の目に触れません。**取りこぼしを検知できる形で持つことが、精度を上げるより先に効きます。**

05 横展開 — 卸売業の受発注・商品マスタへの応用

APPLICATION TO WHOLESALE

決まった書式ではないテキストが大量に届き、決まった型に仕分けたい業務なら構造は同じです。卸売業はその代表格で、事業所34万件超・年間商品販売額389兆円の規模があります^[9]。従業員400名規模の電材卸で月2万件の受注のうち6割が電話とファクスという調査もあります^[6]。

FIG. 08 — 同じ型が当てはまる卸売業の業務

業務	渡すもの	使う型	返るもの／受け取ったコードの仕事
商品マスタへの割り当て	商品名・規格・荷姿の原文。単位は正規化	Choice	自社カテゴリと確率。割れた行は新カテゴリ候補へ
受注テキストの仕分け	ファクスとメールの本文と取引先コード	Choice ×3	担当部署・緊急度・商品コード。3つ高ければ自動
与信・取引条件の見直し	取引履歴の定性メモと公開情報	Score	2～10段階の区分。またいだ取引先を審査へ
返品・クレームの一次分類	受付メモ。出荷ロットを引き当てて渡す	Choice + Noul	品質・物流・受発注の別と再発の恐れ

先行事例も「全部は自動化しない」

トラスコ中山は見積回答AI「即答名人」を2020年1月から運用し、2025年12月末時点でWeb見積依頼率49.4%・見積自動化率30.5%と公表しています^[8]。受注の9割がファクスで届く食品卸でも手書きの識字率は7～8割^[7]。

読めないものと決めきれないものは必ず残ります。

この型が効く業務の見分け方

- 決まった書式ではないテキストが、継続して届く
- 仕分け先が有限で、言葉で定義できる
- 間違えたときに、後工程の人が気づける

3つが揃うなら次ページの形が移せます。本章は適用案です。

05 適用シナリオ — 受注ファクスを、一次仕分けする

A WORKED EXAMPLE

前ページの「受注テキストの仕分け」を例に、実装の形まで具体化します。不動産で使った3原則が、そのまま順番どおりに効きます。

① 渡すもの — 解いた情報と生の原文を両方

解いた情報=取引先コード、マスタ突合の候補20件、正規化した数量・単位・納期。**生の情報**=OCRの粗い書き起こし。**入れないもの**=集計値と日付比較^[5]。

② 質問の並べ方 — 1回にまとめて聞く

質問は同じ state に並列かつ独立に評価されます^[3]。

FIG. 09 — 1リクエストにまとめる質問の例

型	聞くこと	返るもの
Choice	担当部署はどこか	部署と確率
Choice	商品マスタのどれか	商品コードと確率
Choice	緊急度は	通常／至急／要確認
Noul ×2	特別な取引条件／手書きの追記	それぞれ0～1の値

③ 確率を、コードでどう振り分けるか

モデルに「自動化してよいか」は聞きません。**帯を決めるのはコードです。**

FIG. 10 — 確率から帯を決める（閾値は自社データで定める）

帯	条件	処理
自動	3つのChoiceが高確信・両Noulが低い	基幹システムへ流す
保留	確信度が中位／候補が割れている	担当者の確認待ちへ
除外	特別条件か手書きのNoulが高い	必ず人が見る

Noul の閾値を Choice に持ち込まないこと。確率の意味が違います^[5]。

④ どこから始めるか

01 **粒度を先に決める。**200分類は割れる。20分類に束ねる。

02 **正解が無い層を数える。**不動産では77.9%だった。

03 **「分からない」の帯を先に作る。**取りこぼしは後から気づけない。

06 まとめ — 3つの設計原則

CONCLUSION & REFERENCES

原則 01

入力を解いてから渡す。途中経過を作らないモデルは、渡した情報の作り方がそのまま判定に出ます。解いた情報と生の情報を両方渡し、計算はコードに残します。

原則 02

正解が無い前提で測る。一致率を精度と呼ばず、記載準拠・ルール準拠・感度の3本立てで健全さを確かめます。

原則 03

「分からない」を型に入れる。判断のつかないものを表せないと、取りこぼしたことに誰も気づけません。

モデルの選定より先に決めるべきなのは、この3つです。今回の検証で判定が動いた場面はどれも、入力か答えの型を直したときでした。モデルを替えても、この3つを決めなければ結果は変わりません。

出典一覧

- [1] TypeSafe AI「Introducing System One Models & Jev」 2026年9月15日
<https://typesafe.ai/blog/introducing-system-one-models-and-jev>（最終閲覧：2026年9月21日）
- [2] TypeSafe AI「API reference」
<https://docs.typesafe.ai/api.md>（最終閲覧：2026年9月21日）
- [3] TypeSafe AI「How to use TypeSafe's System One API」
<https://docs.typesafe.ai/lms.txt>（最終閲覧：2026年9月21日）
- [4] TypeSafe AI「Models」
<https://docs.typesafe.ai/models.md>（最終閲覧：2026年9月21日）
- [5] TypeSafe AI「Jev 1.13 jaggedness」 2026年9月17日時点
<https://docs.typesafe.ai/model-jaggedness/jev-1.13.md>（最終閲覧：2026年9月21日）
- [6] 中小企業庁「令和3年度 中小企業受発注のデジタル化推進方策に関する調査」
https://www.chusho.meti.go.jp/keiei/gijut/digitalization/download/report_r3.pdf（最終閲覧：2026年9月21日）
- [7] 全国農協食品株式会社 導入事例（掲載：株式会社インフォーマート）
<https://www.infomart.co.jp/ai-ocr/case/07.asp>（最終閲覧：2026年9月21日）
- [8] トラスコ中山株式会社「AIへの取り組み」 2025年12月末時点
https://www.trusco.co.jp/trusco_is/ai.html（最終閲覧：2026年9月21日）
- [9] 総務省・経済産業省「令和3年経済センサス-活動調査 産業別集計（卸売業、小売業に関する集計）」
https://www.stat.go.jp/data/e-census/2021/kekka/pdf/oroshikouri_outline.pdf（最終閲覧：2026年9月21日）



AI を核に、経営課題を解決する。

株式会社PERVA ITカンパニー

本社：〒103-0004 東京都中央区東日本橋2丁目24-9 LIT 4F

TEL：03-6428-7357 MAIL：info@perva.co.jp

お問い合わせ：<https://perva.co.jp/contact>

本書に記載された内容は発行時点の情報に基づくものであり、予告なく変更される場合があります。

記載された会社名・製品名は各社の商標または登録商標です。

© PERVA Inc. All rights reserved.