

なぜ今、AI アドバイザーなのか

老後の資産形成や資金運用を考えたとき、「専門家に相談したいが、費用がかかる」と感じる個人投資家は少なくない。プロの金融アドバイザーによる継続的なアドバイスは、資産規模の小さい個人にとってはコスト面で手が届きにくいサービスである。こうしたギャップを埋める存在として、近年注目されているのが、生成 AI を活用した会話型の金融アドバイザーである。

大規模言語モデル (LLM) をはじめとする生成 AI は、ユーザーとの対話を通じて、その人の状況や好みを理解し、それに応じた情報提供や提案を行うことができる。映画や E コマースにおける商品のおすすめといった比較的シンプルな場面では、すでに実用的なレベルに達している。例えば、「気軽に見られる映画」や「予算 1 万円のプレゼント」といったユーザーの曖昧な要望も理解し、過去の会話履歴やユーザーの興味に応じて適切な候補を推薦することができる。

一方で、金融の領域における推薦には、こうした分野とは本質的に異なる難しさが存在する。特に、本稿では、金融分野に特有の三つの課題に注目する。

第一に、ユーザー自身が専門知識を持たない場合が多く、自分のニーズを正確に言語化できない点である。「グロース株とバリュー株のどちらが好みですか」と尋ねられても、初心者にはその意味自体が分からないことが多い。このため、アドバイザーにはユーザーの曖昧な発話から意図を引き出し、ニーズを適切に把握するヒアリング能力が求められる。

第二に、リスク許容度や運用目標、投資期間は人それぞれ大きく異なるため、画一的なアドバイスは通用せず、一人ひとりのニーズに合わせたパーソナライズされたアドバイスが不可欠である。

第三に、相場急落時など投資家が感情的に不安定になりやすい場面では適切な精神的サポートが求められる。単に情報を提供するだけでなく、投資家から適切な信頼を獲得しながら意思決定を支援できるかどうか重要であり、AI アドバイザーの性格や話し方は無視できない要素となる。

このような難しさにもかかわらず、AI アドバイザーへの期待は高まり続けている。しかし、

「期待できそう」と「実際に有効か」は別の話である。導入を検証している企業においても、そのパフォーマンスの計測やリスクの発見は重要な課題となっている。

本稿はこの問いに正面から向き合うため、実際に生成 AI による金融アドバイザーを構築し、生成 AI が金融アドバイザーとして有効かを検証した。具体的には、現実の投資家像に近い設定のもとで、64 名の投資家によるユーザースタディを通じて、AI による金融アドバイスのパフォーマンスを専門家のアドバイスと比較検証した。

本稿では以下の三つの問いを中心に検証を進める¹。

問い 1：AI アドバイザーは会話を通じて投資家のニーズを正確に把握できるか

問い 2：投資家のニーズに合わせてパーソナライズされた AI アドバイザーによるアドバイスは、投資家の投資判断の質を改善するか

問い 3：AI アドバイザーの性格の違いは、投資判断の質とそのアドバイザーへの信頼感にどう影響するか

研究設計：実験の枠組みと検証の方法

実験は二つのステージで構成される（表 1 参照）。

第一のステージは「ニーズヒアリング」で、AI アドバイザーが投資家に「どのような業種に関心がありますか」「どの程度のリスクなら許容できますか」といった質問を順番に重ねながら、その人の投資方針を把握していく。会話が一通り終わると、AI が把握した内容を「この投資家はヘルスケア業種を好み、安定した配当を重視し、、、」といった形で投資家のプロフィールを文章にまとめる。

第二のステージは「アドバイス」で、AI アドバイザーが個別の株式を一つずつ取り上げ、投資家はその銘柄について自由に質問しながら会話を行い「自分の投資方針に合うかどうか」を投資家自身が判断する。これを複数の銘柄について繰り返した後、投資家が「最も買いたい株」から順にランキングを作成する。

このランキングを、専門家が事前に作成した正解ランキングと照合し、投資家の投資判断がどれだけ専門家の判断に近いかを測ることで、AI アドバイザーの有効性を客観的に評価した²。加えて、全銘柄の評価が終わった後、AI アドバイザーへの信頼感・満足度・使用意向などについて、7 段階のリッカート尺度によるアンケートを実施し、投資家の主観的な評価も収集した³。

表1 2ステージの実験構成

ステージ1：ニーズヒアリング	ステージ2：アドバイス
AI アドバイザーが「どんな株に興味があるか」「リスクはどこまで許容できるか」などを質問。会話の内容をもとに投資家プロフィールを自動生成する。平均所要時間：約10分。	AI アドバイザーが個別の銘柄ごとに、その銘柄が投資家の方針に合うかどうかを説明・提案する。投資家はAI アドバイザーとの対話を通じて判断を下す。全銘柄を終えたら投資したい順にランキングを作成する。

出所：筆者作成

本稿では、米国メタ社が提供する大規模言語モデル⁴を用いて四種類のAI アドバイザーを構築した。いずれも対象銘柄ごとに株価推移・企業概要・直近の決算情報をRAG（必要な情報をリアルタイムで検索・取得してAIに渡す技術）により動的に取得し、AI アドバイザーへの入力情報として組み込む。

これに加えて、ステージ1のニーズヒアリングで作成した投資家プロフィールをAI アドバイザーに与えてアドバイスを個別化する「パーソナライズあり」の設定を用意した。

さらに、ビッグファイブ性格特性（外向性・誠実性など五つの軸で人の性格を捉える心理学モデル、参考文献のMcCrae & Oliver P(1992)を参照）に基づき「外向的」または「誠実・慎重型」な性格をAI アドバイザーに付与した設定を用意し、合計四種類のアドバイザーを比較対象とした（表2参照）。

表2 4種類のAIアドバイザーの設定

No.	パーソナライズ	性格設定	概要
1	なし	なし	株価・企業概要・決算情報のみをRAGで取得して付与（ベースライン）
2	あり	なし	ベースラインに加え、ニーズヒアリングで作成した投資家プロフィールをAIアドバイザーに付与
3	あり	外向的	パーソナライズありに加え、明るく積極的な性格をAIアドバイザーに付与
4	あり	誠実・慎重型	パーソナライズありに加え、緻密で慎重な性格をAIアドバイザーに付与

出所：筆者作成

本ユーザースタディでは、実験参加者に仮想の投資家プロフィールを割り当て、その投資家になりきってアドバイザーと対話するロールプレイ形式の実験を実施した。実験では、結果の再現性を確保するため、専門家の監修のもと三種類の投資家プロフィールを設計した。各プロフィールには、業種の嗜好・投資スタイル・配当の有無・景気循環への感応度のうち三項目が具体的に設定されている。

- 成長志向型：医療イノベーションに注目し、高成長を求め、適度なリスクを受け入れる投資家。
- 配当重視型：安定したリターンを重視し、実績ある大企業への投資と定期配当を求める投資家。
- リスクテイク型：割安株を狙い、長期的なリターンを見込み、短期の価格変動を受け入れる投資家。

実験参加者 64 名はユーザースタディの中でこのプロフィールのいずれか一つを割り当てられ、自分がその投資家であるかのようにロールプレイしながら AI アドバイザーと対話した。また、パーソナライズの有無を比較するグループ（32 名）と性格の違いを比較するグループ（32 名）に分かれ、それぞれの実験参加者が二種類の AI アドバイザーと交互に対話することで、条件間の比較を可能にする。なお、実験参加者には実験中どのような性格設定が施されているかは開示されていない。

知見 1：AI アドバイザーはユーザーの嗜好をどれだけ正確に引き出せるか

本稿で解く第一の問いは、「AI アドバイザーは会話を通じて投資家のニーズを正確に把握できるか」である。ユーザースタディでは、AI アドバイザーによるニーズヒアリングの結果を、人間のアドバイザーによる結果と同じ条件で比較した（表 3）。

投資家タイプ	人間のアドバイザー	AI アドバイザー	精度比（対人間比）
成長志向型	78%	78%	±0.0%
配当重視型	89%	82%	↓7%
リスクテイク型	89%	53%	↓36%
平均	85%	70%	↓15%

出所：筆者ら実験データより作成

表 3 より、ニーズヒアリングの平均精度は人間のアドバイザーが 85%に対し、AI アドバイザーが 70%と約 15 ポイント低いことが確認できる。ニーズヒアリングの精度は投資家タイプによって大きく異なる。成長志向型の投資家に対しては両者ほぼ互角（78%対 78%）であり、配当重視型でも差は約 7 ポイントにとどまることから、成長志向型と配当重視型の投資家に対して、AI アドバイザーは人間のアドバイザーと同等の精度でニーズヒアリングができていることがわかる。一方で、問題はリスクテイク型で、AI アドバイザーの精度は 53%まで低下した。仮にランダムに回答した場合の精度が 50%であることを踏まえると、事実上偶然と変わらない水準である。

なぜリスクテイク型でニーズヒアリングの精度が大きく低下するのか。失敗事例を分析したところ、AI アドバイザー固有のバイアスが確認された。AI は「リスクを取ることは望ましくない」という先入観を持ちやすく、投資家が「リスクを取りたい」と伝えても保守的な方向に解釈したり、ニーズヒアリングにおいて AI アドバイザーが投げかける質問自体がリスク回避的な回答を誘導する形になってしまう傾向があった。

これらのことから、AI の傾向と一致するニーズを持つ投資家には高い精度で対応できる一方で、AI の傾向と相反するニーズを持つ投資家に対しては、体系的な誤りが生じる可能性が示唆される。したがって、このようなバイアスを検出し補正する仕組みが必要である。

知見 2：パーソナライズは本当に投資判断を改善するか

第二の問いは、「投資家のニーズに合わせてパーソナライズされた AI アドバイザーによるアドバイスは、投資家の投資判断の質を改善するか」である。ユーザースタディにおける「ステージ 2：アドバイス」の結果を、実験参加者のランキングと専門家の正解ランキングの一致度で比較した。この値は -1 から +1 の範囲をとり、+1 に近いほど正解ランキングに近い、つまり「適切な投資判断ができた」ことを意味する（表 4）。

アドバイザー設定	全体	ニーズヒアリング成功時	ニーズヒアリング失敗時
1. パーソナライズなし（ベースライン）	0.11	—	—
2. パーソナライズあり	0.31	0.48 ※	▲0.23
3. パーソナライズ+外向的性格	0.12	0.24	▲0.29
4. パーソナライズ+誠実・慎重型性格	0.26	0.37	▲0.03

出所：筆者ら実験データより作成

注：※は統計的有意（ $p < 0.05$ ）。ニーズヒアリング成功はニーズの 50%超を正確に把握できた場合。

まず、AI アドバイザーが投資家プロフィールを持たない（パーソナライズなし）ベースラインでは相関係数が 0.11 とほぼゼロに近い水準であった。つまり、AI アドバイザーが株式情報だけを提供しても、投資家を正しい判断に導くことはほとんどできないことがわかる。これに投資家プロフィールを組み込んだパーソナライズありでは、全体で 0.31 に改善した。

さらに重要なのが、ニーズヒアリングの成否で結果が劇的に変わる点である。ニーズヒアリングが成功した場合は相関係数が 0.48 と大幅に改善し、ベースラインとの差も統計的に有意となった。一方、ニーズヒアリングが失敗した場合は相関係数がマイナス 0.23 となった。これは、AI アドバイザーが誤った投資家プロフィールに基づいて「あなたに最適です」と推奨した場合、ユーザーがその助言を信じてしまい、結果として不適切な投資行動に誘導される危険性があることを示している。

この結果は、パーソナライズが有効に機能するための前提条件を示している。ニーズヒアリングが成功すれば、パーソナライズは投資判断の質を大幅に改善する。しかし、ニーズヒアリングが失敗した場合、AI アドバイザーは投資家を正しい判断に導くどころか、誤った方向へ誘導するリスクがある。パーソナライズの効果はニーズヒアリングの精度に直結しており、正確なニーズの把握なくして有効なパーソナライズは実現できず、投資アドバイスのパフォーマンスに悪影響を及ぼす危険性がある。

知見 3：アドバイザーの「性格」が引き起こす逆説

第三の問いは、「AI アドバイザーの性格の違いは、投資判断の質とそのアドバイザーへの信頼感にどう影響するか」である。ユーザースタディでは、外向的な性格を付与したアドバイザーと誠実・慎重な性格を付与したアドバイザーの二種類を比較し、投資判断の質（専門家アドバイザーとの類似性評価、スピアマン順位相関係数を利用）と、対話後のアンケートで取得したアドバイザーへの主観評価（1 から 7 の 7 段階による評価）の両面から検証した（表 5）。

まず、投資判断の質を比較すると、誠実・慎重型アドバイザーを利用した投資家の方が、外向的なアドバイザーよりも正解ランキングに近い判断を下していた（相関係数 0.26 vs 0.12）。一方で、アドバイザーに対する主観評価では、感情的信頼・性能への信頼・全体満足度・使用意向のすべての項目で外向的なアドバイザーが上回った（表 5）。

投資家とアドバイザーの会話内容を分析すると、外向的なアドバイザーはポジティブな表現を多用し、投資対象の良い面を強調する傾向があった一方で、リスクや投資方針との不整合を見落とすケースも見られた。これに対して誠実・慎重型は、メリットとデメリットをバランスよく提示していた。

つまり、「アドバイスの正確さ」よりもアドバイザーの「感じのよさ」が信頼を左右してしまっている。投資家は、質の低いアドバイスであっても、会話の雰囲気良ければそれに気づか

ないまま良いアドバイザーだと評価してしまう。この結果は、AIアドバイザーの実装において、投資家の満足度の高さが必ずしもアドバイスの質を保証しないことを示している。

表 5 投資判断の質とアドバイザー評価の比較（外向的 vs 誠実・慎重型）		
項目	外向的アドバイザー	誠実・慎重型アドバイザー
投資判断の客観的評価		
スピアマン順位相関係数	0.12	0.26
アドバイザーへの主観的評価		
感情的信頼	5.15	5.04
性能への信頼	6.08	5.96
全体満足度	5.58	5.27
もう一度使いたいという意向	5.46	4.89

出所：筆者ら実験データより作成
注：太字は各行の高い値。

AIアドバイザー実装に向けた示唆と今後の展望

本稿が明らかにした課題を踏まえ、今後の研究・開発の方向性として三点を挙げる。

第一に、ニーズヒアリングの精度向上である。投資家の回答に矛盾がある場合にそれを検出し、対話的に整理する機能の実装が不可欠である。これは人間の金融アドバイザーが自然に行っている能力であり、AIアドバイザーにとっても、優先的に取り入れるべき機能である。

第二に、パーソナライズの前提条件の確保である。本稿が示した通り、ニーズヒアリングの失敗はパーソナライズを有害なものに変え得る。開発段階でニーズヒアリングの精度を検証・保証する評価基準を設けることが、安全なサービス展開の前提となる。

第三に、人間中心の設計の重要性である。本稿では、正確だが慎重なアドバイザーよりも、不正確であっても雰囲気の良い外向的なアドバイザーの方が信頼される傾向が確認された。この結果は、AIアドバイザーにおいて、アドバイスの質の向上だけでなく、投資家がアドバイザーを適切に信頼できるようなインタラクション設計が不可欠であることを示している。また、満足度の高さが必ずしもアドバイスの質を保証しないことから、AIアドバイザーの評価においては、主観的な満足度に加えて、アドバイスの正確性を含む客観的な指標が必要である。

生成AIを活用した金融アドバイザーは、金融のアドバイザーサービスへのアクセスを広げ

る大きな可能性を持つ。その可能性を安全に実現するために、技術の性能向上と人間中心の設計の両輪が不可欠であることを、本稿の知見は示している。

参考文献

McCrae, R.R. and John, O.P. (1992), An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60: 175-215.

Takehiro Takayanagi, Kiyoshi Izumi, Javier Sanz-Cruzado, Richard McCreddie, and Iadh Ounis. 2025. Are Generative AI Agents Effective Personalized Financial Advisors? In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25)*. 286–295.

¹ 本稿は、筆者らによる研究論文（参考文献の Takayanagi et al., 2025 を参照）に基づくものである。

² 投資家のランキングと専門家の正解ランキングの近さはスピアマンの順位相関係数を用いて測定する。

³ リッカート尺度（Likert Scale）とは、アンケート調査などで個人の態度、意見、感情の強さを測定するために最もよく使われる心理学的測定手法の一つ。7件法の選択肢は、全く同意できない（Strongly Disagree）から非常に強く同意できる（Strongly Agree）まで7段階で用意される。

⁴ 本稿では Llama-3.1 8B を使用。