

大規模言語モデルによる心理尺度開発支援フレームワーク

分寺杏介(神戸大学)・丹亮人(エム・アール・アイリサーチアソシエイツ株式会社)・
福島健太郎(大阪大学)・島田大祐(東京都立大学)

✉ bunji@bear.kobe-u.ac.jp

※本研究は、LRM 株式会社から提供された共同研究資金およびJSPS 科研費24K16811 を受けて行われました。



▲ ポスター

- 心理尺度の作成は職人技が要求され、時間とお金がかかってしまう
 - 近年では生成AIによる尺度フレームワークも整備されてきた(e.g., Russell-Lasalandra et al., 2026)
 - ▶ 構成概念の定義や既存尺度情報を「事前知識として与える」必要がある
 - 制度・技術・運用慣行が頻繁に変わる領域(例:情報セキュリティ意識 [ISA])では、定期的に、また対象属性(業種)に合わせて尺度項目を更新していく必要が生じる
- タスクを分解したプロンプトチェイニングによって、事前知識の検索だけでなく標準的なテスト作成フローの全体を自動化することは可能か?**

- 心理尺度の性質は主に**信頼性(reliability)**と**妥当性(validity)**の観点から評価される
 - ▲ そのための多様な手続きに関する知見は相当に蓄積されてきた(e.g., AERA et al., 2014)
 - 近年ではコンピュータ内で完結させる評価法も見られる(e.g., Russell-Lasalandra et al., 2026)が、特に回答者視点に関する評価を十分に行うことは原理的に難しいと考えられる
 - ゆえに、自動生成尺度の「人間の回答に基づく」評価に関する知見の蓄積が必要である
- 生成AIによって作成された尺度項目は、実用に耐えうるものになっているか?**

目的 生成AIを利用して、特定の属性を対象にした心理尺度項目の作成を試みる

その心理統計学的性質を実データに基づいて標準的な手続きによって検証する

提案フレームワーク

- 2026年1月にGemini 3 Pro で実施
- 各ステップのアウトプットはJSON形式
 - ▶ 次ステップのプロンプトに追加した
- 対象業種はサイバー攻撃のリスクが高いと考えられている「医療」「金融」「IT」



◀ プロンプトおよび生成された項目

1 ドメインに関する最新の知識の同定
代表的な情報セキュリティガイドラインを抽出(業種共通/業種固有および国内/国際)
【出力例】ISO/IEC 27001(共通・国際) FISC安全対策基準(金融・国内)

2 測定対象の概念の整理
各ガイドラインの要点を抽出 ▶ ガイドライン間で共通した要素をまとめる
【出力例】「ガバナンス・組織体制」「アクセス制御」「ID管理」「インシデント対応・復旧」

3 測定領域の抽出
ガイドライン間に共通するテーマのうち、一般労働者に関係するものを選出
【出力例】「情報の持ち出し」「ID・パスワード管理」「インシデント報告」など

4 シナリオと誤概念の抽出
項目作成時に利用するため、具体的な状況と正しい対応、広まりやすい誤概念を整理
【出力例】「ファイル共有プロトコル(通称PPAP)」「パスワードの定期的な更新」など

5 テスト仕様に基づく項目生成
今回はISAにおける代表的な枠組み(KABモデル; e.g., Parsons et al., 2014)に基づき作成
知識(Knowledge) → 態度(Attitude) → 行動(Behavior)
項目形式や項目数だけでなく、難易度配分、テーマのバランス、および回答バイアスを誘発しやすい「避けるべき表現」などを指示して項目を生成

6 項目の自動選定・表現の洗練
別のペルソナを割り当て、各項目の表現の最終チェックおよび項目数の厳選(形式的要件)逆転項目の比率・困難度比率のチェック(文言修正例)K項目:曖昧な主観表現(例:たまたま)、誘導(例:すべき)などの排除

生成された項目例(全体では各領域×側面に6項目ずつ生成)

- 知識(共通)** 使用しているブラウザから「更新が必要です」というポップアップ通知が出たため、ポップアップをクリックせず、ブラウザのメニューから正規のアップデート機能を開いて更新を実行した。【適切】
- 知識(医療)** ナースステーションのPCで記録入力中にナースコールが鳴ったため、すぐに処置に入れるよう、認証カード(HPKIカード等)をリーダーに挿入したまま病室へ向かった。【不適切】
- 態度(金融)** 有名人や知人の口座情報を検索しても、誰にも言わなければ、銀行(会社)に損害を与えるわけではないので問題ないと思う。
- 行動(IT)** 私は、.gitignoreファイルを適切に設定し、ローカルの設定ファイルや一時ファイルがリポジトリにコミットされないようにした。

実データによる評価

対象者

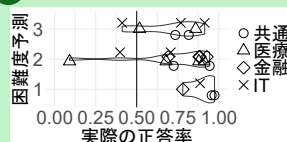
- 主な勤務先が「金融業・保険業」「病院・一般診療所」「ソフトウェア業」に属する
 - 業務でPCまたはタブレットを1日に30分以上利用する
- ウェブ調査(Freeasy)により得た1,825名分の回答を使用(除外率52.4%)

手続き

- (抜粋)
- 生成されたISA尺度項目(KAB各12項目)に回答する
 - 尺度の全体的な印象(例:回答しやすいと感じた)を回答
 - 関連尺度としてRSeBISまたは社会的望ましき尺度に回答
 - 行動(B)項目の想定場面(例:.gitignoreファイルの操作)の経験有無を回答

分析結果

I K項目(正誤問題)の正答率 II B項目の場面経験率



場面	共通	医療	金融	IT
1	5	2	1	3
2	4	2	6	1
3	6	3	5	4
4	6	5	1	4
5	3	2	3	2

III 「質問の意味が分からない」選択率

知識:0.00 - 6.52% 態度:0.00 - 1.12% 行動:0.11 - **22.16%**
【最大値の項目】「私は、.gitignoreファイルを…(左下例)」(**22.16%**)

IV 尺度の全体的な印象(5件法)の平均値

	質問文	医療	金融	IT
答えにくさ	質問の意味を理解するのが難しかった	2.20	2.00	2.21
	意味がわからない単語(専門用語など)が使われていると感じた	2.22	1.90	2.48
	回答しやすいと感じた	3.36	3.52	3.39
答えやすさ	同じ業種の大抵の人は正しく理解できるだろうと思った	3.74	4.03	3.91

V 1因子モデルのフィット

VI 関連尺度との相関係数

尺度	業種	EFA	CFI	TLI	SRMR	RMSEA	RSeBIS	自己欺瞞	印象操作
知識	共通	0.829	0.989	0.981	0.092	0.068	0.261	0.062	-0.054
	医療	0.716	0.990	0.983	0.056	0.035	0.236	-0.054	-0.127
	金融	0.750	0.853	0.756	0.074	0.170	0.113	-0.103	0.105
	IT	0.635	0.862	0.770	0.086	0.132	0.151	-0.050	-0.046
態度	共通	0.826	0.934	0.889	0.044	0.120	0.474	0.013	-0.030
	医療	0.737	0.951	0.918	0.049	0.082	0.414	-0.006	-0.104
	金融	0.845	0.960	0.933	0.039	0.100	0.402	0.074	-0.040
	IT	0.856	0.925	0.875	0.053	0.142	0.362	0.004	-0.009
行動	共通	0.653	0.600	0.333	0.124	0.222	0.666	0.084	-0.082
	医療	0.828	0.848	0.857	0.132	0.211	0.460	0.108	-0.030
	金融	0.654	0.914	0.397	0.053	0.088	0.522	0.116	-0.060
	IT	0.712	0.638	0.747	0.132	0.231	0.409	-0.058	-0.048

考察

項目文自体はそれなりに良いものが生成できたように見えても、「心理尺度」としての性質はイマイチ

▲ 特に一次元性が満たされたとはいえない尺度もあるのでブラッシュアップが必要

対象属性および背景知識を明確にすると、項目も個別具体的にしやすい可能性

▲ ある意味では生態学的妥当性につながる一方で「帯域幅と忠実性のジレンマ」も

▲ 具体性が高すぎる項目は認知負荷が高くなるため脱落も多くなる可能性がある

尺度・項目の評価の段階では、やはり実際に回答してもらい、心理測定学的手続きを踏むことが大事

参考文献

- AERA, APA, & NCME (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Parsons, K. et al. (2014). Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q). *Computers & Security*, 42, 165-176.
- Russell-Lasalandra, L. L. et al. (2026). Generative psychometrics via AI-GENIE. *Behavior Research Methods*, 58, Article 217.