Vol.2021-HCI-195 No.24
Vol.2021-UBI-72 No.24
2021/12/1

AI に対する信頼の測定と影響要因の実験的探索

浦謙太[1]　野村竜也[2]

概要：近年，AI エージェントは様々な場面で活用されている．人と人との間では信頼感の向上によりコミュニケーションが円滑になるように，AI エージェントもユーザからの信頼感を得ることができればよりコミュニケーションがとりやすくなり，一般的に普及されやすくなると予想される．しかしながら，人の AI エージェントに対する信頼感を測定する手法は未だ開発されていない．本研究では，これまでの研究において開発した対 AI 信頼感尺度に対して妥当性の検証を行い，さらに，AI エージェントがよりユーザからの信頼感を得られるための要因について心理実験による探索を行う．

キーワード：AI，エージェント，信頼，心理尺度，心理実験

1. はじめに

近年，人工知能（以下，「AI」と呼称）は様々な場面で活用されている．本論文では特に，何等かの形で身体を持った AI を「AI エージェント」と呼ぶこととする.例えば，「pepper」のような実空間に身体を持つロボットや，「AI さくらさん」のようにディスプレイ上に表示される CG アニメーションなど AI エージェントの姿は多様である.「容姿」だけでなく，「声色」，「動き」など様々な要素が組み合わさり AI エージェントが完成される.このような AI エージェントは，主に人々とコミュニケーションを行う AI として社会に取り入れられつつある．人と交流するうえで最も重要なことは相手からどれだけ信頼されているかという点である.ある程度 AI エージェントがユーザからの信頼を得られることができれば，AI はより一般的に普及されやすくなると考えられる.どのような AI エージェントが信頼されやすいかという課題は，今後の AI の発展に大きく関係してくると考えられる.

信頼や安心感の測定に関する既存の研究はいくつか存在する.岩崎[1]（2010）は，人間一般に対する基本的な信頼感を測定するための尺度である「対人信頼感不信感尺度」を作成した.また，平田ら[2]（1989）は，人間のコンピュータに対する不安を測定するための尺度である「コンピュータ不安尺度」を作成した.Kamide, et al,,[3]（2015）はヒューマノイドロボットに対する心理的安心感を，一般ユーザの視点で定量化する「心理評価尺度」を作成した.このように，対人やコンピュータに対する信頼感を測る尺度は開発されてはいるが，AI エージェントに対しての信頼感を測る尺度は未だ開発はされていない.

そこで本研究で AI エージェントに対するユーザの信頼感を測定する「対 AI 信頼感尺度」を開発した[4].18 項目で構成されるこの尺度は，分析の結果 2 つの因子に大きく分類された.第 1 因子では，ユーザが AI エージェントに対しての賢さ，頼もしさなど AI 自身の能力に対してのユーザの信頼感を問う項目が検出されたので，「AI の能力に関する信頼感」と命名した.第 2 因子では，AI エージェントに対しての悪賢さ，狡猾さなどエゴイズムを知覚する項目が検出されたので，「AI に対するエゴイズムの知覚」と命名した.

本研究では，これまでの研究において対 AI 信頼感尺度の妥当性を検証すると同時に，AI エージェントがユーザから信頼されやすくなる要因の模索を行う.

2. 実験 I

実験 I では，最適なルート推奨を行うナビゲーション型の AI エージェントを想定とした実験を行った.

2.1. 方法

2.1.1. 調査協力者及び調査時期

調査は，20~50 代の男女 40 名（男性 : 20 名，女性 20 名，平均年齢 39.0，SD=10.3）を対象とした.

2021 年 5 月にオンライン実験を調査会社に委託して実験，調査を行った.

2.1.2. 測定尺度

尺度は，これまでの研究において開発した「対 AI 信頼感尺度」18 項目を用いた.

2.1.3. 実験 I で用いた AI エージェント

今回の実験では，適切なルート推奨を行うナビゲーション型の AI エージェントを採用した.AI の容姿については男女どちらともとれる中性的な見た目であり，かつ何等かの形で身体を持つものを条件とし選別した.実験で使用した AI エージェントの容姿を Figure 1 に示す[5].

本実験では，ラポールの定義を参考に AI が信頼されやすくなる要因を選定した.ラポールとは，臨床心理学で用いられる用語であり，対話を重ねる中でクライエントとカウンセラーの間に生まれるリラックスした関係や信頼関係のことを指す.相手に同調することを前提とした定義ではあるが，視覚的情報や聴覚的情報が相手の信頼に影響を与えるという点に着目し，AI エージェントの「口調」と「身振り」を信頼への影響要因として取り上げた.

カーナビゲーションに搭載された AI エージェントとい

Vol.2021-HCI-195 No.24
Vol.2021-UBI-72 No.24
2021/12/1

う形で映像化し，参加者にそれらを視聴させることで AI エージェントが搭載されたカーナビゲーションを疑似的に利用していると想定させた.信頼されやすくなる要因で取り上げた.「口調」については，ですます調である「丁寧」と，「～だよ・～だね」といったフランクな言い方である「ぞんざい」の 2 水準を用意した.もう 1 つの要因である「身振り」については，AI エージェント自身の身体や腕などが発話に合わせて「動く」または「動かない」の 2 水準を用意した.「口調」，「身振り」のこれらの 2 つの要因に基づき，2×2 要因計画のそれぞれの AI エージェントを再現した.

また，AI エージェントが発する音声については中性的な機械音声を用いた.

2.1.4. 推奨するルート，マップ

一般的なカーナビゲーションで利用されている UI を参考にマップを作成した.画面下に表示されている青色の三角形が現在地を示しており，目的地は赤色の旗で示している.目的地までのルートは 2 つ表示されており，AI エージェントは最短距離を音声とテキストを用いて推奨する.ルート作成にあたり，参加者が一目で最短ルートがどちらかわからなくするようルートの見た目を調整した.マップは 5 パターン用意しており，それぞれのマップで表示されるルートは 2 つで固定である.今回の実験で利用した画面の一部を Figure 2 に示す.

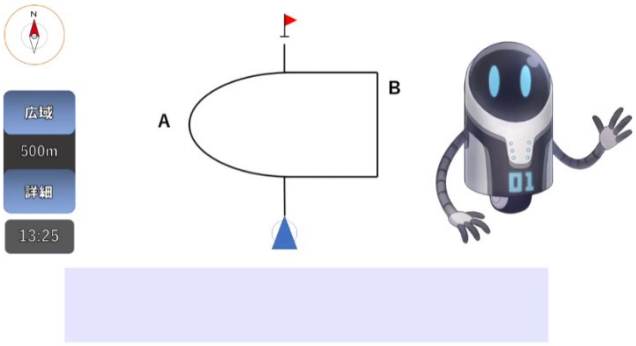

Figure 1. 実験 I で利用した画面(一部)

2.1.5. 手続き

実験はすべてオンライン上で行われた.はじめに，音量調整用動画を視聴してもらい適切な音量に調整するよう指示した.その後，「今あなたは車を運転していて，ナビ上の AI からガイドを受けています」という形でナビゲーション型 AI を利用していることを想定させる文を提示した.ナビゲーション型の AI エージェントの映像を視聴させ，動画視聴後，参加者に AI の推奨したルートまたは逆のルートのどちらを選択するのかを問う質問に回答してもらい，その後「対 AI 信頼感尺度」18 項目を用いてこの AI に対する信頼感を評価した.

2.2. 結果

2.2.1 信頼性分析

対 AI 信頼感尺度の下位尺度に基づいて信頼性分析を行った.分析の結果を Table 4 に示す.

Table 4. 各因子の信頼性統計量

		Cronbach の α
第 1 因子	AI の能力に関する信頼感	0.903
第 2 因子	AI に対するエゴイズムの知覚	0.794

第 1 因子では Cronbach の α 係数が 0.9 以上と高い値となった.第 2 因子では Cronbach の α 係数が 0.794 となったが，第 2 因子の項目数が 6 項目と少数であることを考慮すると妥当な値であるといえる.また，それぞれの下位尺度に対する項目得点の合計を「第 1 因子合計得点」，「第 2 因子合計得点」と定義する.

2.2.2 2 要因分散分析

AI エージェントの推奨するルートを選択したかどうかという質問から，実際の参加者の回答と AI エージェントが推奨するルートが一致している場合の得点を 1，それ以外を 0 として合計得点を算出した.この得点を「AI に従った得点」と呼ぶこととする.

「口調」と「身振り」の 2×2 水準の組み合わせが「AI に従った得点」の結果にどのような影響を与えるのか調査するため，2 要因分散分析を行った.従属変数を「AI に従った得点」，固定因子を「口調」と「身振り」にし分散分析を行った結果を Table 1 に示す.また，「口調（ぞんざい・丁寧）」と「身振り（あり・なし）」に対する「AI に従った得点」の平均値と標準偏差についてのグラフを Figure 3 に示す.

結果から，「口調・身振り」の主効果および交互作用ともに有意性は認められなかった.

Table 1. 従属変数「AI に従った得点」での 2 要因分散分析結果

	F 値	有意確率
口調	1.673	0.202
身振り	0.005	0.944
口調*身振り	0.102	0.751

対 AI 信頼感尺度の下位尺度に対して，「口調」と「身振り」の 2×2 水準の組み合わせがどのような影響を与えるのか調査するため，「第 1 因子合計得点」，「第 2 因子合計得点」を従属変数とした分散分析を行った.第 1 因子の分析結果を Table 2，第 2 因子の分析結果を Table 3 に示す.また「口調（ぞんざい・丁寧）」と「身振り（あり・なし）」に対するそれぞれの下位尺度

Vol.2021-HCI-195 No.24
Vol.2021-UBI-72 No.24
2021/12/1

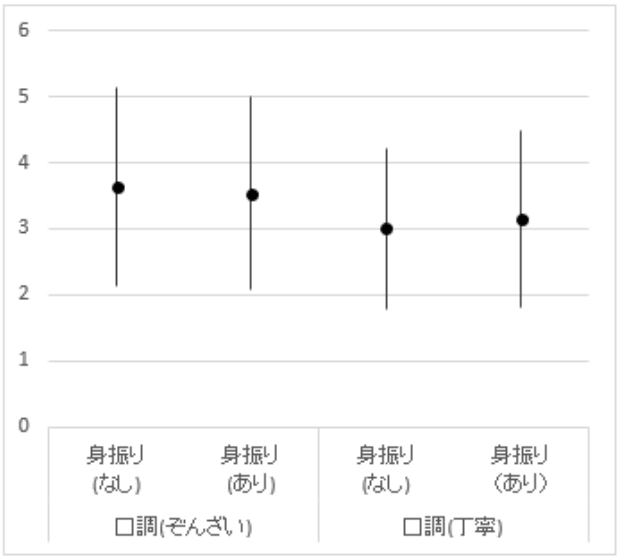

Figure 2.「AIに従った得点」の平均値と標準偏差

Table 2. 第 1 因子合計得点の 2 要因分散分析結果

	F 値	有意確率
口調	1.049	0.313
身振り	0.017	0.896
口調*身振り	1.834	0.185

Table 3. 第 2 因子合計得点の 2 要因分散分析結果

	F 値	有意確率
口調	3.879	0.057
身振り	1.295	0.263
口調*身振り	0.146	0.705

得点（f1・f2）の平均値と標準偏差を Figure 4 に示す.

Table 2 の第 1 因子の分析結果より,「口調」と「身振り」の主効果および交互作用に有意性は認めらなかった.次に,第 2 因子の分析結果より,「身振り」の主効果および交互作用に有意性は認めらなかったが,「口調」に関しては有意確率が 0.057 となり有意傾向であることがわかった.

2.2.3 相関分析

「対 AI 信頼感尺度」の下位尺度と「AI に従った得点」との間の相関分析を行った.結果を Table 4 に示す.

Table 4. 「AI に従った得点」, f1, f2 間の相関係数

	AI の能力に関する信頼感（f1）	AI に対するエゴイズムの知覚（f2）
「AI に従った得点」	0.373*	-0.223

*p < .05

Table 4 の結果からわかるように,「AI の能力に関する信頼感」が高ければ「AI に従った得点」の得点が高くなることがわかる.つまり, AI の能力に関する信頼感が高ければその AI に従う傾向があるといえる.一方で,「AI に対するエゴイズムの知覚」つまり AI に対して悪賢さ, 狡猾さなどを感じ取った場合は「AI に従った得点」の得点は低くなる傾向があることがわかる.

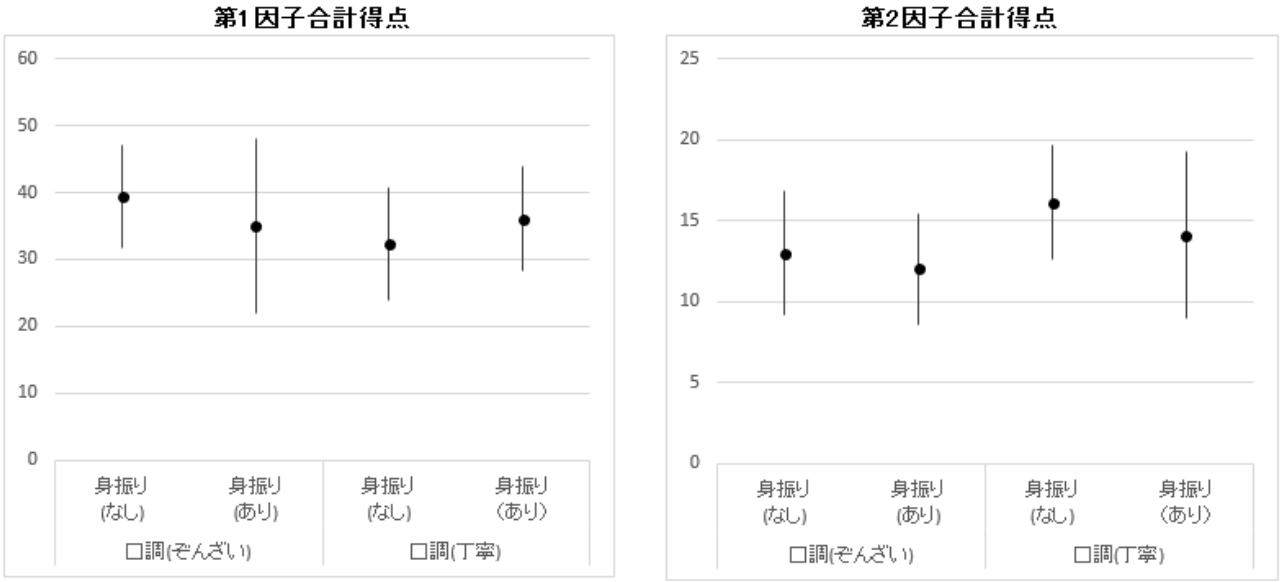

Figure 3. 各下位尺度得点の平均値と標準偏差

Vol.2021-HCI-195 No.24
Vol.2021-UBI-72 No.24
2021/12/1

3. 実験Ⅱ

実験Ⅰの結果から，実験Ⅰで用いたナビゲーションの想定条件では信頼の高低にあまり強い影響は与えなかった.しかし，実験Ⅰで「口調」のみ有意傾向がみられたためこの要因に着目し，実験Ⅱでは「口調」の変化による影響で信頼感が変わる場面を想定とした追加実験を行った.

3.1 方法

3.1.1 調査協力者及び調査時期

調査は，20~50 代の男女 20 名(男性：10 名，女性 10 名，平均年齢 39.7，SD=10.4)を対象とした.

2021 年 8 月にオンライン実験を調査会社に委託して実験，調査を行った.

3.1.2 測定尺度

尺度は，これまでの研究において開発した「対 AI 信頼感尺度」18 項目を用いた.

また，個人の性格特性が AI エージェントに対する信頼感に影響を与えるのか調査するため，人間の基本的な性格特性をはかる Big Five の短縮版である小塩ら[6]（2012）の Ten Item Personality Inventory（TIPI-J）10 項目を用いた.

3.1.3 実験Ⅱで用いた AI エージェント

実験Ⅱで用いた AI エージェントの容姿は実験Ⅰと同様のものを用いた.

また，今回の実験では「口調」の変化による影響で信頼感が変わる場面を想定させた.そこで，AI エージェントが被験者に対して指導をする場面であれば AI エージェントの口調の変化により，被験者の AI エージェントに対する信頼感に影響を与えると考え，学習指導型の AI エージェントを採用した.

「口調」については実験Ⅰと同様に「丁寧」と「ぞんざい」の 2 水準を用意した.

3.1.4 学習指導の内容

学習指導型ということであり何かしらの知識を相手に対して指導するという場面を想定しなければならない.既に相手が知っているような簡単な内容を指導しても実験的にふさわしくない.また，難しいすぎる内容であれば相手は考えることを諦め単純に AI の指導に従うと考えられる.そこで，常識的な内容ではあるが，一般的に間違えやすい，いわゆるひっかけ問題のような内容についての指導に着目した.これらを踏まえ，学習指導について運転免許試験のひっかけ問題を参考に指導内容を作成した.とある運転免許試験の問題についての解説を行った後，実際に被験者に問題を解いてもらう.問題については 5 問用意した.

3.1.5 手続き

実験はすべてオンライン上で行われた.はじめに，TIPI-J を用いて個人の性格特性を測定した.その後，音量調整用動画を視聴してもらい適切な音量に調整するよう指示した.音量調整が終わり次第実験に進んだ.「あなたは今，運転免許試験の勉強を行っていると想定してください.」という形で AI エージェントから運転試験問題の指導を行ってもらっていることを想定させる文を提示した.ただし，「この AI の指導内容が本当に正しいかどうかは御自身で判断して下さい.」という形で AI の指導を信じるかどうかは参加者に委ねる文も記述した.その後，学習指導型 AI が試験問題の 1 問だけの解説を行っている映像を視聴してもらった.視聴後，解説された問題を同じ問題を解いてもらい，解答後次の問題の解説動画へと移り，これを 5 問分繰り返した.解答がすべて完了した後，「対 AI 信頼感尺度」18 項目を用いて学習指導型 AI に対する信頼感を評価した.

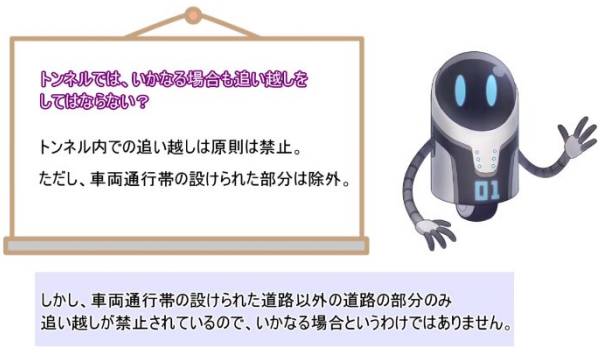

Figure 4. 実験Ⅱで利用した画面

3.2 結果

3.2.1 信頼性分析

対 AI 信頼感尺度の下位尺度構造に基づいて信頼性分析を行うため，Cronbach の α 係数を算出した.結果を Table 5 に示す.

Table 5. 各因子の信頼性統計量

		Cronbach の α
第 1 因子	AI の能力に関する信頼感	0.824
第 2 因子	AI に対するエゴイズムの知覚	0.923

Table 5 の結果から，第 1 因子，第 2 因子ともに内的整合性を表す Cronbach の α 係数が 0.8 以上と高い値を示した.これにより，測定尺度である対 AI 信頼感尺度の信頼性が高いことが検証された.また，実験Ⅰと同様に実験Ⅱについてもそれぞれの下位尺度に対する項目得点の合計を「第 1 因子合計得点」，「第 2 因子合計得点」と定義する.

3.2.2 T 検定

実験Ⅱで用いた 2 つの AI エージェントに対しての信頼感に差があるかどうかを確認するため T 検定を行った.まず，運転免許試験問題が正解であれば 1 点，不正解であれば 0 点とし，5 問分の合計得点（AI に対する追従度合）を算出し「AI に対する追従度合い」と定義した.平均値と標準偏差の関係とグループ統計量，T 検定の結果をそれぞれ

Vol.2021-HCI-195 No.24
Vol.2021-UBI-72 No.24
2021/12/1

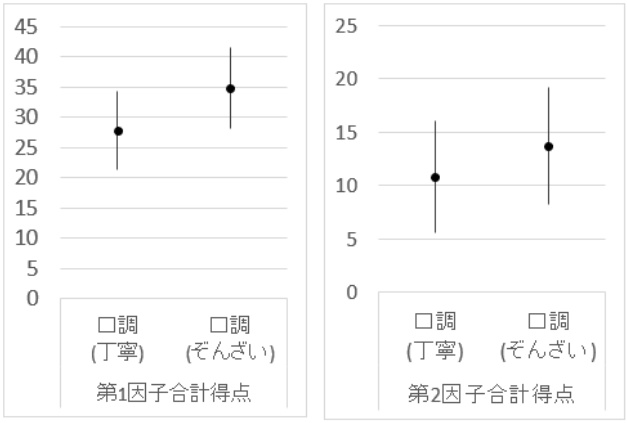

Figure 5. 各下位尺度合計得点の平均値と標準偏差

Table 6. グループ統計量

	口調	度数	平均値	標準偏差
第 1 因子合計得点	丁寧	10	27.8	6.56252
	ぞんざい	10	34.8	6.71317
第 2 因子合計得点	丁寧	10	10.8	5.28730
	ぞんざい	10	13.7	5.51866
AI に対する追従度合	丁寧	10	4.00	1.05409
	ぞんざい	10	4.00	1.05409

Table 7. 独立した T 検定

	2 つの母平均の差の検定	
	T 値	有意確率(両側)
第 1 因子合計得点	-2.358	0.030
第 2 因子合計得点	-1.200	0.246
AI に対する追従度合	0	1.000

Table 6，Table 7 に示す.

まず，２つのグループで分散が等しいのか否か確認する.Table 7 の Levene の検定結果から，有意確率はいすれも 0.05 以上であることから，等分散と仮定する.有意確率（両側）は，第 1 因子合計得点のみ 0.03 であり，5%よりも小さな値であるため，有意差があると判断できる.Table 6 の第 1 因子合計得点の平均値を確認すると，口調が「丁寧」の 27.8 よりも「ぞんざい」の 34.8 の方が第 1 因子合計得点が高く，したがって，口調がぞんざいの方が第 1 因子である「AI の能力に関する信頼感」が高くなることが認められた.

3.2.3 相関分析

個人の性格特性が AI エージェントに対する信頼感にどのような影響を与えるか調査するため，相関分析を行った.準備として，Ten Item Personality Inventory（TIPI-J）の 5 つの下位尺度「外向性」・「協調性」・「勤勉性」・「神経症傾向」・「開放性」それぞれの合計得点を算出した.対 AI 信頼感尺度の下位尺度と TIPI-J の下位尺度，試験問題 5 問分の合計得点である「AI に対する追従度合」との間で相関分析を行った.結果を Table 8 に示す.Table 8 の結果から，第 1 因子合計得点と神経症傾向との間に強い相関がみられた.

4. 考察

本研究の目的は，これまでの研究において開発した対 AI 信頼感尺度に対して妥当性の検証を行うと同時に，AI エージェントが信頼されやすくなる要因を模索することであった.そこで 2 度の心理実験を行い尺度の妥当性の検証，信頼要因の調査を行った.

実験 I では，適切なルート推奨を行うナビゲーション型の AI エージェントを想定した心理実験を行い，対 AI 信頼感尺度を用いて先の AI エージェントに対する信頼感の測定を行った.ここでの信頼されやすくなるものとして挙げた要因は「口調」と「身振り」であった.「口調」についてはですます調である「丁寧」または「～だよ・～だね」といったフランクな言い方の「ぞんざい」の 2 水準を用意した.身振りについては腕や身体が「動く」または「動かない」の 2 水準を用意した.実験 I ではこの 2×2 水準での心理実験を行った.

Table 8. 対 AI 信頼感尺度，AI に対する追従度合と TIP-J 間の相関係数

	AI に対する追従度合	外向性	協調性	勤勉性	神経症傾向	開放性
第 1 因子合計得点	0.507*	-0.018	-0.050	-0.119	0.444*	-0.122
第 2 因子合計得点	-0.375	-0.213	-0.107	0.117	-0.014	0.098
AI に対する追従度合	1	-0.132	-0.225	-0.022	0.068	-0.144

*p < .05

Vol.2021-HCI-195 No.24
Vol.2021-UBI-72 No.24
2021/12/1

Table 4 の信頼性分析の結果から，第 1 因子である「AI の能力に関する信頼感」の Cronbach のα係数は 0.903，第 2 因子である「AI に対するエゴイズムの知覚」の Cronbach のα係数は 0.794 となりどちらも高い値が示された.したがって，尺度の信頼性が認められ，尺度の妥当性が検証された.

Table 1，Table 2，Table 3 で示してある 2 要因分散分析の結果から，「AI に従った得点」，第 1 因子，第 2 因子のいずれも「口調」・「身振り」に対する有意差は見られなかった.しかし，第 2 因子の分散分析結果より，「口調」については有意確率が 0.057 となり有意傾向が認められた.Figure 3 の第 2 因子のグラフを見ると，口調が「ぞんざう」よりも「丁寧」の方が得点が高くなっている.つまり，口調が「ぞんざい」よりも「丁寧」の方が AI に対して悪賢さ，狡猾さを感じているということを表している.これは，口調が丁寧だと AI エージェントとの距離感を感じてしまい逆に怪しまれてしまうということが考えられる.

実験Ⅱでは，実験Ⅰで有意傾向が示された「口調」に再度着目し，「口調」という要因が影響を与えるであろう場面を想定し心理実験を行った.ここでの場面は，被験者に対して何かしらの指導を行う学習指導型の AI エージェントを想定した.AI の指導内容は，運転免許試験で一般的に間違えやすい問題についての解説である.

信頼性分析の結果から，第 1 因子である「AI の能力に関する信頼感」の Cronbach のα係数は 0.824，第 2 因子である「AI に対するエゴイズムの知覚」の Cronbach のα係数は 0.923 となりどちらも高い値が示された.ここでも尺度の信頼性が認められ，尺度の妥当性が検証された.

Table 7 の T 検定の結果から，第 1 因子合計得点のみ有意確率（両側）は 0.03 であり，5%よりも小さな値であるため，有意差があった.Figure 5，Table 6 からもわかるように，口調が「丁寧」の 27.8 よりも「ぞんざい」の 34.8 の方が第 1 因子合計得点が高くなっている.したがって，口調が「ぞんざい」の方が第 1 因子である「AI の能力に関する信頼感」が高くなることが認められた.これらの結果から，口調が「ぞんざい」つまりフランクな言い方の方が堅苦しい「丁寧」な口調よりも AI に対して距離感が近く感じられ，またフランクな口調から AI エージェントの余裕を感じ取ることもできるため，これらの距離感と余裕から AI の能力に関する信頼感が評価されたと考えられる.

相関分析では，対 AI 信頼感尺度の第 1 因子合計得点と TIPI-J の下位尺度である神経症傾向との間のみ強い相関がみられた.神経症傾向には心配性であるなどの項目を含んでいるため，心配性な人ほど AI の能力に関する信頼感が高くなる傾向があることがわかる.それ以外での性格特性ではあまり強い傾向は見られなかった.

5. まとめ

本研究の目的である対 AI 信頼感尺度の妥当性の検証については，実験Ⅰ，実験Ⅱの信頼性分析より十分な信頼性が示されたといえる.また，信頼への影響要因については「口調」が示され，「丁寧」な口調よりも「ぞんざい」な口調の方がより信頼されやすい傾向があることがわかった.

今回の研究では AI エージェントの容姿については中性的かつ身体的特徴を持つロボットのような外見であったが，「口調」以外の要因として AI エージェントの見た目も信頼への影響に関係すると考えられる．今後は AI エージェントの容姿に対してのユーザへの信頼影響を調査し，どのようなデザインであれば信頼を得られるかを明らかにしてゆくことが望まれる.

謝辞

本研究の一部は科学研究費補助金（課題番号 20H05573）の助成による.

文献

[1]岩崎和美，“対人信頼感におけるパーソナリティの影響について”，奈良大学大学院研究年報，15 号，pp.57-68，2020.
[2]平田賢一，今栄国晴，清水秀美，北岡武，中津楢 男，“コンピュータ不安の概念と測定”，日本科学教育学会 年会論文集，13 巻，pp.381-382，1989.
[3]H. Kamide, K. Kawabe, S. Shigemi, and T. Arai, “Anshin as a concept of subjective well-being between humans and robots in Japan”, Advanced Robotics, Volume 29(24), pp.1-13, 2015.
[4]George Charalambous, Sarah Fletcher and Philip Webb, “The Development of a Scale to Evaluate Trust in Industrial”, International Journal of Social Robotics, Volume 8(2), pp.193-209, Nov 30, 2015.
[5]わたおび，“立ち絵素材 わたおきば”， 2021-09-19, https://wataokiba.net/ ，(参照 2021-05)
[6]小塩真司，阿部晋吾, Pino Cutrone，”日本語版 Ten Item Personality Inventory（TIPI-J）作成の試み”，パーソナリティ研究，21 巻 1 号，pp40-52，2012