

チューリングテストによる AI と人の特徴分析の予備的研究

赤堀 侃司

NPO 教育テスト研究センター

一般社団法人 ICT CONNECT 21・一般社団法人日本教育情報化振興会

抄録

人工知能（以下、AI と略す）と人の特性についてチューリングテストを用いて抽出した結果を基に、学習に適用することを目的として本研究を実施した。チューリングテストを用いて、問いかけに対して AI が答える回答と特定の人が答える回答を、60 名の実験協力者に提示し、どちらの回答が AI かを判定してもらいその正答率を求めた。同時に AI の回答内容を吟味して、もし人間だと仮定したらどの年齢レベルかを推定してもらった。その結果、AI だと正答した率は 17 問全体の平均値が 0.81 と高い値であった。また人間だと仮定した時の推定年齢は、平均的にはおよそ中学生レベルと推定された。ただし、この結果は提示した問いの内容に強く依存することが分かった。さらに年齢推定において、その理由を自由記述で書いてもらい分析した結果、AI と人間のいくつかの特徴が見いだせた。

キーワード：AI、AI 時代、チューリングテスト、ビッグデータ、学習方法

1. はじめに

これからの時代は Society5.0 に代表されるように、大きな社会変革と、未来の学びの姿が不連続的に進化すると言われている（Society 5.0 に向けた人材育成に係る大臣懇談会、2018）。その社会において、AI やビッグデータが大きな役割を果たすことは言うまでもない。その時代を AI 時代と呼べば、その未来を生きる子供たちには、どのような資質・能力が必要とされ、どのような学習が求められるだろうか（赤堀侃司、2019a）。プログラミング教育（赤堀侃司、2018）や STEM・STEAM 教育も、その学習内容・方法として有効であるかもしれないが、まだ模索中と言ってよいだろう。

そこで AI 時代における学習の在り方を探求するためには、AI と人間はどこが違うのかを、始めに明らかにする必要がある。これまでも、いくつかの優れた研究報告がある。例えば、新井紀子（2018）や奈良潤（2017）、ゲルトギーゲレンツァー（2010）などが参考になる。そこで論じられてきたことは、AI の強みと弱みや読解力などの小中学生が身に付けたい学力や、人間のもつ暗黙知や優れた直感力とこれらの力を生かす方法など、これからの学習への示唆を含んでいる。本研究では、これまでの先行文献の知見を参考にしながら、チューリングテストによる AI と人の比較や特徴分析を行い、その結果を元に AI 時代の学習について考察をして示唆を得ることを目的としている。

2. チューリングテストによる実験方法

2.1. チューリングテスト

チューリングテストは、アラン・チューリングの1950年の論文「Computing Machinery and Intelligence」の中で提案された人間と機械を判別する思考実験として知られているが、その概念図を図1に示す。

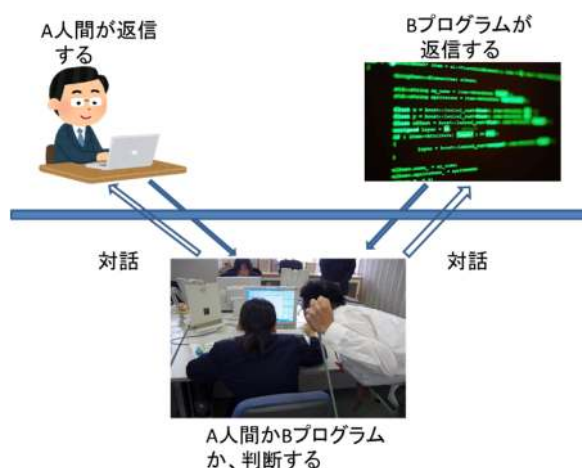


図1 チューリングテスト

図1のように、壁で仕切られた部屋を想定する。壁の手前の人間が、A人間とBプログラムの両方に問いを発信する。問いを受けたAとBはどちらかが返信するが、その問いと回答を繰り返す。問いを発信した手前の人間は、図1のように返信を受けるが、その返信はAの人間からなのかBのプログラムからなのかを判定する思考実験である。1回の試行では、Aの人間かBのプログラムのどちらかが返信するが、この試行を何回も繰り返して、その判定結果である正答率が同じ程度であれば、プログラムは人間と同じ知能を持つと言ってもよいのではないか、という考え方である。

2.2. 実験方法

<p>指示</p> <ul style="list-style-type: none"> ・ 問題で、AかBか、コンピュータか人間か ・ コンピュータのほうを答える ・ コンピュータのレベルを選ぶ ・ その理由を書く 	<p>回答</p> <p>1 (1) あいさつ *</p> <p><input type="radio"/> AIは、Aです。</p> <p><input type="radio"/> AIは、Bです。</p> <p>1 (2) あいさつ *</p> <p><input type="radio"/> 人間に例えると、幼児レベルです。</p> <p><input type="radio"/> 人間に例えると、小学生レベルです。</p> <p><input type="radio"/> 人間に例えると、中学・高校生レベルです。</p> <p><input type="radio"/> 人間に例えると、大学生かそれ以上のレベルです。</p> <p>1 (3) あいさつ *</p> <p>その理由は例でしようか、自由に書いてください。</p> <p>回答を入力</p>
---	--

図2 回答方法の例

本研究では、図1のようなりアルタイムの判定をする実験ではなく、実験協力者に提示する教材を予め準備して、その教材に対して判定してもらう方法を用いた。その教材は、筆者が問いを発信して、AIの回答はGoogle Homeの応答とし、人間の回答は筆者の応答として、テキストにして提示する教材である。

実験協力者には、AIはAかBのどちらですかという質問形式で判定をしてもらった。実験協力者には、AIと判定した場合、そのAIは人間だとしたらどのレベルかを判定してもらった。さらにその理由や根拠も自由記述で回答してもらった。その回答方法を図2に示す。

なお、そのレベルを本研究では推定年齢と呼び、以下の4段階に分けた。

1：幼児レベル、2：小学生レベル、3：中学・高校生レベル、4：大学生かそれ以上

実験協力者は、都内の大学生で男女それぞれ30名からなる合計60名であり、2018年10月20日に都内の大学で実施した。合計17の教材を用意した。

3. 分析と結果

図3に、正答率と推定年齢の相関分析のグラフを示す。相関係数は0.51で、この値自身にあまり意味はないが、図3の回帰直線の上部に知識カテゴリー（■で示す）が、下部に思考カテゴリー（●で示す）が、全体に認識カテゴリー（▲で示す）が分布していることが興味深い。サブカテゴリーの内容に依存することはあるが、カテゴリー毎の特徴が読み取れる。

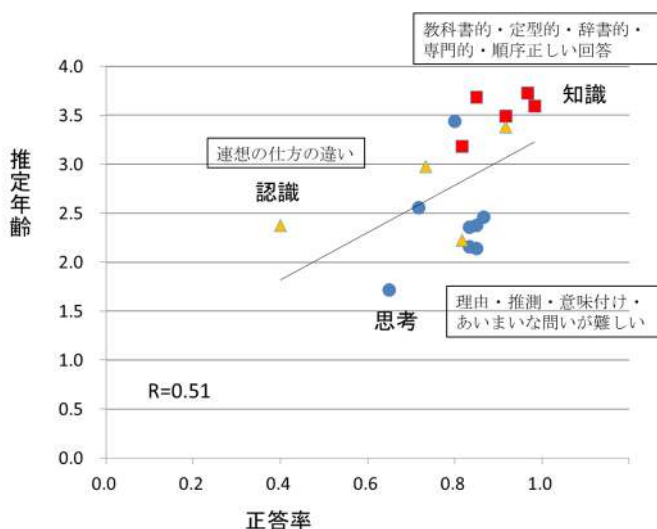


図3 正答率と推定年齢の相関グラフ

実験協力者の何故AIだと判定したのかという問いに対する自由記述の回答を分析した結果の特徴を、図3に書き加えた。

知識カテゴリーは、「教科書的・定型的・辞書的・専門的・順序正しい回答」という特徴があった。人間はあいまいで、その時その場で対応する柔軟性を持っているが、AIの回答した専門的な内容を読んで、その推定年齢は高いと判断したと思われる。思考カテゴリーは、逆に推定年齢は低く、その理由として「理由・推測・意味付け・あいまいな問いが難しい」が挙げられた。AIは、まだ思考レベルでは人間に

劣っていると大学生たちは判断した。特に、理由を述べる、文章を理解する、意味付けをするなどでは、人間の能力が優れていると感じたと言える。さらに認識カテゴリーでは、写真やイラストを見てどのような連想をするのか、どのように認識したのかという問いで、連想の違いによってばらつきが大きい結果になっている。

4. 結論と考察

以上の結果と分析を、結論としてまとめる。

- (1) 大学生60名を対象にしたチューリングテストでは、AIと人間の判定は0.81と高い正答率を示した。これは、現代の若者が日常的にスマホなどに触れていて、AIを身近に感じているからと考えられる。
- (2) AIの正答率や推定年齢は、問いの内容である知識・思考・認識のカテゴリーに大きく依存する。全体的には、知識カテゴリーの推定年齢が高く、思考カテゴリーが低く推定された。全体の平均は、ほぼ中学生レベルと推定された。

なお本研究は、学会論文誌に投稿中であることを、お断りしておきたい（赤堀、2019b）。

最後に本論文は、(NPO) 教育テスト研究センターの支援と、科学研究費助成金・基盤研究C（代表、赤堀侃司、課題番号15K01034）の支援を受けたことを明記して、厚くお礼申しあげる。

参考文献

赤堀侃司（著）（2018）プログラミング教育の考え方とすぐに使える教材集，ジャムハウス
 赤堀侃司（著）（2019a）AI時代を生きる子どもたちの資質・能力，ジャムハウス
 赤堀侃司（2019b）チューリングテストによるAIと人の特徴分析の予備的研究，AI時代の教育学会論文誌，269号掲載予定
 新井紀子（著）（2018）AI vs. 教科書が読めない子どもたち，東洋経済新報社
 ゲルトギーゲレンツァー（著），小松淳子（翻訳）（2010）なぜ直感のほうが上手いくのか？，インターシフト
 奈良潤（著）（2017）人工知能を超える人間の強みとは，技術評論社

