

人工知能裁判官が持つ思いやりが与える信頼度への影響

Effect on reliability given by consideration of artificial intelligence judge

神先 凌太郎
Ryotaro KANZAKI中島 遼
Ryo NAKAJIMA井上 綾
Aya INOUE田中 瞭
Ryo TANAKA小島未央奈
Miona KOJIMA所属:同志社大学商学部田口研究室
University of Doshisha Taguchi seminar

Abstract: ①This paper aim to clarify the indicators that influence on the reliability in the two situations where it is a judge and AI judge. We used consideration in addition to the honesty and ability which were indicated previous study (Nakayachi, 2018) as indicators. The necessary of consideration is mentioned in the field of trial (Supreme Court, 2013), and it is closely related to the ability, honesty and the use of compassion as an indicator influencing the trust of Mayer (1995).

②As a result, there is a significant difference that improving perception of reliability both the judge and AI judge in condition of consideration. However, the honesty and ability also improve in condition of consideration. Therefore, the possibility that the consideration doesn't influence directly on improving perception of reliability.

1. 研究背景

近年、人工知能が目覚ましい発展をしている。その発端となっているのは 2012 年の画像認識のコンペティションで優勝した、トロント大学が開発、実装したディープラーニングである (松尾, 2015)。2016 年には囲碁の世界チャンピオンを Google Deepmind 社の Alpha Go が倒し、話題になった (日本経済新聞社, 2018)。こういった話題の中心となっているディープラーニングの技術は大量のデータを材料にして自らの中にパターン認識をし、新たなアルゴリズムを生成し、答えを出すことを得意とする。米国ミシガン州のノースポイント社 (現在社名はエクイバント) では COMPAS と呼ばれる裁判 AI が発明されており米ウィスコンシン州などで保釈をするか否かなど、裁判官の判断をする補助的役割として活用されている (2012, Northpointe Inc.)。裁判の分野は、ルールである法律が文書としてあらかじめ決められていることに加えて、過去の判例という大量のデータに照らし合わせ判断を下すことが可能である。そのためディープラーニングをはじめとする AI のテクノロジーはリーガルの分野と相性が良い。これからの日本においても、本格的に裁判に AI が導入されることも想定できる。

そのような状況において信頼は重要である。信頼とは相手の行為が自分にとって否定的な帰結をもたらさうな不確実性がある状況で、それでも、そのようなことは起こらないだろうと期待し、相手の判断や意思決定に任せておこうとする心理的な状態 (中谷内, 2008) である。

提案の受容に信頼が関係することは、AI の場合でも示されている (Tay, Jung, & Park, 2014)。つまり、どれだけ優れた AI が裁判に対して判決を下したとしても、その AI に対し被疑者、被害者の信頼度が低いと結果の正しさ、合理性に関わらずおお不満な結果と感ずる場合が考

えられる。また、信頼度が下がると提案や判決への受容度が下がり、再審率が上昇することも考えられる。ひいては司法という国の三権分立の一角を担う存在も揺るぎかねない。

では信頼度とは何によって構成されているのか。信頼の構成に関しては多くの研究で言及されている。Peters, Covello, & McCallum (1997) の知識の豊富さ・透明性・正直さ・気遣い、Johnson (1999) の危険か安全かを区別するスキル・責任感・思いやりなどがある。この様な多様な信頼度を構成する要素について中谷内 (2018) では誠実さと能力に大きく分けられるとしている。この主要二要因は山岸 (1998) の能力に対する期待と意図に対する期待、Metlay (1999) の認知的要素と感情的要素とはおおよそ一致する。一方 Mayer (1995) の中では信頼に影響するものとして能力、正直さ (誠実さ) に加え慈悲という三つの項目をおいている。

また、国民が求める裁判官像についての審議会において、弁護士・元名古屋高等検察庁検事長である水原 (2010) は「人間味あふれる心の温かさのわかる裁判官でないと信頼される裁判はできない ~中略~ 公正さに加えて、思いやりのある裁判官に裁いてもらうことが望ましい」と答弁している。弁護士・元広島高等裁判所長官である原田 (2010) は裁判官に一番必要な素質として能力をさし置き「人間的な温かさ」をあげている。ダニエル (2000) もまた理想的な裁判官の要素として法の尊重、公正さ、良心に従う勇氣に加えて、人間に対する思いやりを述べている。これらに共通して言えることは裁判官において能力や公正さも大事であるが、思いやりを持って相手の気持ちを汲むことができることも必要であるということである。これは Mayer (1995) が述べている三つの信頼を構成する項目の中の「自己中心的な利益動機とは別に、相手にとって善になることを願う「慈悲の度合い」が信頼度に

影響する”という考察と関連すると捉えられる。

そこで今回、裁判における裁判官への信頼度は主要二要因に加え、「思いやり」が影響を及ぼすものとして仮説を立てる。この思いやりとは相手の心情を推し量る能力であると定義する。

また、人間と AI では同じ反応を示すという Nass & Moon (2000) という研究がある。そのため人間の裁判官において思いやり条件の信頼度が上昇することと同じ様に、AI でも思いやり条件の信頼度が上昇すると考えられる。以上の研究背景、先行研究を経て仮説を下記の様に設定する。

仮説 1: 思いやりは裁判において裁判官への信頼度を高める。

仮説 2: 思いやりは裁判において AI 裁判官への信頼度を高める。

2. 調査方法

2.1 実験

場面想定法を用いて AI が裁判官を行うモデル、人間が裁判官を行うモデルに思いやりあり条件となし条件を想定し、計 4 つのシナリオを作成し、サーベイ実験を行った。Google フォームのアンケートを用い、期間は 2019/1/16～1/23 で行った。

2.2 分析方法

2 × 2 の被験者間二要因分散分析を行う。独立変数に、属性の違い (AI か人間か) と思いやりの有無を置く。従属変数には、信頼度、能力、誠実さを置く。

3. 結果

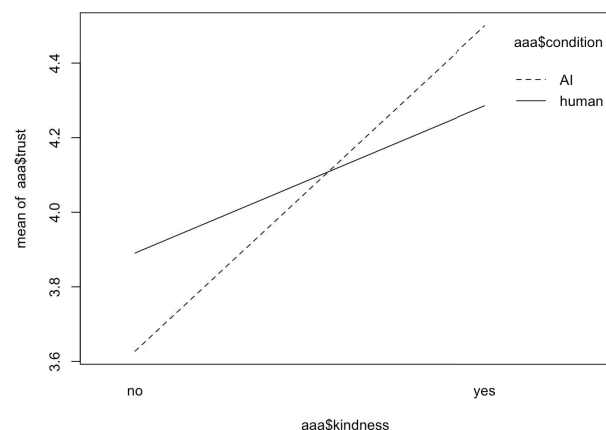


Figure 1 独立変数を属性である AI, 人間と思いやりあり, なしとし, 従属変数を信頼度として出力した信頼の平均値

	度数	信頼	能力	誠実さ	思いやり
AI・思いやりなし	110	3.63	4.33	3.61	2.13
人・思いやりなし	64	3.89	3.83	3.41	3.06
AI・思いやりあり	62	4.50	4.92	3.92	3.03
人・思いやりあり	56	4.28	4.11	4.39	4.60

Table1 各条件における記述統計量

実験の結果、思いやりによる主効果が見られ、仮説が支持された。つまり、思いやり条件は AI 裁判官においても人の裁判官においても信頼に影響することが統計的に有意であると分かった。また、属性の違いおよび交互作用は、統計的に有意な差を示さなかった。

ただし補足分析において、思いやり条件による能力と誠実さの上昇を調べた結果、有意であることが分かった。中谷内 (2008) では、信頼は能力、誠実さに影響を与えるとされる。そのため思いやり条件がこの能力と誠実さを上昇させていることで、能力もしくは誠実さ、または両方を経由して、信頼度の上昇に繋がらなかったとは言いきれない。

4. 考察

思いやり条件において誠実さと能力も上がった理由を、能力、誠実さそれぞれに分けて考える。能力が上がった理由としてはシナリオ中の裁判官及び AI 裁判官の相手を思いやる発言が影響したと考えられる。この思いやる発言により裁判官及び AI 裁判官に相手を推量する力、コミュニケーション能力、伝達能力が備わっていることを見出したと考えられる。誠実さが上がった理由としても、上記同様にシナリオ中の裁判官及び AI 裁判官の相手を思いやる発言が影響し、裁判官に対し一生懸命さや、真面目に取り組んでいる姿勢を感じ、誠実さが上がったと考えられる。

5. 研究の限界と今後と展望

限界としては、思いやり指標を測ったが、思いやりありなしの条件に関わらず AI にだけ思いやりを全く感じていない割合が高かった。このため、AI に思いやりを感じない層が一定数いると考えられる。また、これは Wolter Pieters (2010) が述べるように、AI では説明は信頼を与えるために使用されるが、情報を過度に与えず不信任感を強めることが考えられる。これにより、AI に対して思いやりを感じさせるシナリオという説明が不信感を抱かせたことも考えられる。加えて、思いやり条件において思いやりのみならず、能力と誠実さも同様に向上させてしまう結果となったため思いやりが直接信頼度に影響しているとは言いきれない結果となった。

また従属変数を信頼度、説明変数を能力と誠実さにおき、全 4 条件において分析した結果、人間においては能力、誠実さのどちらも信頼度への影響が有意であるという結果が得られたが、AI においては、能力のみの信頼度への影響が有意であるという興味深い結果となった。つまり AI においては誠実さが信頼度に影響しているとは言えない結果となった。この理由については今回の論文では明らか

にすることはできなかった。

また今回、有意差は出ていないが思いやりなし条件において、人の能力と誠実さを AI が上回ったにも関わらず、AI は人よりも信頼度が低い結果となった。この能力と誠実さは主要二要因であり信頼度を構成する中谷内(2018)と先行研究で明らかにされているにも関わらず、このような逆転する結果となった理由については今回の論文中では解明できなかった。

今後の展望としては、思いやりが直接的に信頼度に影響するのかを確かめることが出来なかったことから、思いやり条件において、能力や誠実さが信頼に影響しないものを採用し、間接的に影響するのかが、それとも直接的に影響するのかを明らかにしたい。また、AI に思いやりを感じない人が一定数いると考えられることから、否定的固定観念を元から抱いている人がどれくらいおり、それをどういった要素が取り除くのかを研究していく必要性がありそうである。加えて、AI に対して誠実さが信頼度に影響しなかったことを踏まえ、AI に対して新たな信頼の要因があること、もしくは能力のみが影響するのかなどを明らかにしていきたい。

参考文献

Ang, W. & Benbasat, I. (2005). Trust in and adoption of online recommendation agents. *“Journal of the Association for Information Systems”* 6, 72-101.

Communicating Autonomous Agents」, “Group Decision and Negotiation 2 (3)”, pp-237-258, Kluwer Academic Publishers

Northpointe Inc. (2012)
“COMPAS Risk & Need Assessment System”
Selected Questions Posed by Inquiring Agencies

johnson, B.B. (1999). Exploring dimensionality in the origins of hazard related trust. *“Journal of Risk Research”*

Nass, C. & Moon, Y. (2000). “Machines and mindlessness: Social responses to computers. The Journal of Social Issues” 56, 81-103.

Peters, R.G., Corelli, V.T., & McCallum, D.B (1997) “The determinations of trust and credibility in environmental risk communication: An empirical study” *Risk Analysis*, 17, 43-54

PIOTR J. GMYTRASIEWICZ • EDMUND H. DURFEE (1993) “Toward a Theory of Honesty and Trust Among”

Renn, O. & Levine, D. (1991). Credibility and trust in risk communication. In R. E. Kasper-son & P.J.M. Stallen (Eds.), “communicating risk to the public” Dordrecht; Kluwer Academic Publisher. pp. 175-218

Tay, B., Jung, Y., & Park, T. (2014). “When

stereo- types meet robots: The double-edge sword of robot gender and personality in human-robot interaction” *Computers in Human Behavior*, 38, 75-84.

Wolter Pieters (2010) “Explanation and trust: what to tell the user in security and AI?”

Wolter “Pieters Ethics Inf Technol” (2011) 13pp. 53-64

最高裁配布資料 2003 年 10 月 21 日 (火)
[裁判官の人事評価についての仕組みの整備]
裁判官に求められる資質・能力について記載された文献等

ダニエル・H・フット(ワシントン大学教授)
(2000.1.1-15) “ジュリスト 1170 号・司法に何を望むか”

中谷内一也・横井良典(2018) 「治療方針の共有が人工知能への信頼に及ぼす影響」, “社会心理学研究 34 (1)”, pp16-25, 日本社会心理学会

中谷内一也(2008) 「リスク管理機関への信頼: SVS モデルと伝統的信頼モデルの統合」, “社会心理学研究第 23 巻第 3 号” pp259-268, 日本社会心理学会

日本経済新聞社, 金子豊(2018). 『AI2045』
日本経済新聞出版社

藤田耕三(弁護士・元広島高等裁判所長官)
2013 年 8 月 8 日(火)
“司法制度改革審議会集中審議第 2 日目議事録”

松尾豊(2015). 『人工知能は人間を超えるのか: ディープラーニングの先にあるもの』 KADOKAWA

水原敏博(弁護士・元名古屋高等検察庁検事長)
2013 年 8 月 8 日(火)
“司法制度改革審議会集中審議第 2 日目議事録”

山岸俊男(1998). 『信頼の構造: こころと社会の進化ゲーム』 東京大学出版会

本研究で用いたシナリオ

シナリオ前半
(AI 条件)

あなたは裁判の傍聴をしています。この裁判所では、人工知能が裁判官を務めています。人工知能裁判官は法律と過去の判例をもとに、最終的な判決を下します。

(人 条件)

あなたは裁判の傍聴をしています。裁判官は法律と過去の判例をもとに、最終的な判決を下します。

シナリオ後半

(思いやりあり条件)

<判決のシーン>

人工知能 裁判官：「被告人は今回犯した罪により，3年の刑と処す．3年という期間は，人生の中においても少なくない時間ではありますが，しっかりと罪を受け入れて今後改心して，新たな第二の人生を歩むことを願っています．以上で閉廷します．」

(思いやりなし条件)

<判決のシーン>

人工知能 裁判官：「被告人は今回犯した罪により，3年の刑と処す．以上で閉廷します．