

消費者委員会 人工知能（AI）技術の利用
と消費者問題に関する専門調査会（第9回）
議事録

内閣府消費者委員会事務局

消費者委員会 人工知能（AI）技術の利用と消費者問題に関する
専門調査会（第9回）
議事次第

1. 日時 令和8年9月9日（水）15:00～16:41

2. 場所 消費者委員会会議室及びテレビ会議

3. 出席者

（委員）

【会議室】

小塚座長、加藤委員、坂下委員、田中委員

【テレビ会議】

丸山座長代理、大塚委員、岡崎委員、唐沢委員、野村委員、馬籠委員

（オブザーバー）

【テレビ会議】

大澤委員、柿沼委員、善如委員、

（説明者）

【会議室】

矢谷 浩司 東京大学大学院情報理工学系研究科電子情報学専攻教授

（事務局）

小林事務局長、小堀審議官、三谷参事官

4. 議事

（1）開会

（2）① 人工知能（AI）技術の利用と消費者問題に関する専門調査会の検討状況

（事務局担当者）

② AI技術と消費者問題について

有識者ヒアリング（矢谷 浩司 東京大学大学院情報理工学系研究科電子情報学専攻教授）

（3）閉会

《1. 開会》

○小塚座長 それでは、定刻になりましたので、「消費者委員会人工知能（AI）技術の利用と消費者問題に関する専門調査会」の第9回会合を進めさせていただきます。

皆様、お忙しいところお集まりいただきまして、ありがとうございます。

本日の出欠ですが、専門調査会の委員のほうは、加藤委員、坂下委員、それから田中委員と私、小塚が消費者委員会の会議室におります。

それから、オンライン、テレビ会議システムのほうに、丸山座長代理をはじめとして、大塚委員、岡崎委員、野村委員、馬籠委員に御出席いただいております。岡崎委員は少し遅れるという御連絡を頂戴しているようです。

それから、消費者委員会のほうの委員ですが、大澤委員、柿沼委員、善如委員、山本委員の4委員がテレビ会議システムにて御出席ということで、専門調査会との関係ではオブザーバー扱いということになります。

それでは、いつものように事務局から御説明をいただきます。よろしくお願いします。

○三谷参事官 事務局でございます。

本日はテレビ会議システムを活用して進行いたします。一般傍聴者にはオンラインにて御視聴いただき、報道関係者のみ会議室にて傍聴をいただいております。

議事録につきまして、後日、消費者委員会のホームページに掲載いたします。その議事録が掲載されるまでは、YouTubeでの見逃し動画配信を行います。

それから、配付資料につきましては、お手元の議事次第に記載してございます。もし不足の資料などがありましたら、事務局までお申し出くださいますようお願いいたします。

事務局からは以上です。

○小塚座長 ありがとうございます。

ただいま御説明いただきました事務局は、三谷参事官でございます。前任の友行参事官には、我々は大変お世話になりました。本日会議室にはいらっしゃらないですけれども、改めてお礼申し上げたいと思います。そして三谷参事官、今後どうぞよろしくお願い申し上げます。

《2. （1）人工知能（AI）技術の利用と消費者問題に関する専門調査会の検討状況（事務局担当者）》

○小塚座長 それでは、議事次第の議事に従って進めていきたいと思います。

最初は議事（1）です。「人工知能（AI）技術の利用と消費者問題に関する専門調査会の検討状況」ということですが、これまで当専門調査会もいろいろとヒアリングを行

い、また議論をしてまいりました。

そこで、我々が検討すべき論点があったわけですが、事務局でそれについての検討状況をまとめていただいて、検討状況（案）という資料を作っていただきました。20分程度で御説明をいただけるということですので、よろしくお願いいたします。

○福島補佐 事務局でございます。

まず、検討状況の御報告の前に、前回、第8回専門調査会終了後に行われました質疑応答・意見について御報告させていただければと思います。

野村委員から、御発表いただいた丸山座長代理に対しまして、異なる事業間でのAIエージェントの連携について、あるAIエージェントで問題が可視化された場合、その影響範囲をどこまで法的に問えるのかという御質問をいただきました。

これに対し丸山座長代理から、重要な御指摘である点、また、透明性の確保といった場合に、何をどのように事業者の開示させるのかの判断と、当該開示が消費者にとってどのような意味があるのかという点が重要となるという趣旨のコメントをいただいたところでございます。

では、検討状況の御報告に移らせていただければと思います。次のページをお願いいたします。

本専門調査会では、上の四角囲いにごございます「本専門調査会の主な問題意識」に記載されております問題意識から議論をスタートいただきました。すなわち、AI技術は、消費者の意思決定の材料の提供を行っていた段階から、技術的には意思決定を代替し得る段階になってきたといえるのではないかと。また、生成AIにより生成物を人間が作成したものであるかのように消費者に認識させ、消費者の意思決定のプロセスに一定の影響を与えると考えられるのではないかと。さらに、消費者の意思決定のプロセスに問題が生じているのであれば、対策が必要なのではないかという点でございます。

そして、8月に開催されました第8回専門調査会までに検討していただいた主な論点といたしましては、まず1点目、AI技術の進展に伴い、消費者のAI技術の利用主体としての重要性はどのような観点から高まっているといえるか。2点目、AI技術の事業者による利用が、消費者の意思決定のプロセスに及ぼし得る影響及びリスクについて適切に把握・評価するとともに、その対応として、必要な環境整備を推進することが必要ではないかという点になります。

次のページをお願いいたします。

まずは論点1というところで、消費者のAI技術の利用主体としての重要性の高まりというところです。AI技術の利用の量的な拡大という点から議論をいただきました。

まず、四角囲いの上でございます。令和4（2022）年後半、生成AIが社会実装され、国民のAI利用は急速に拡大を続けております。こちらは総務省の「令和8年版情報通信白書」を参照させていただいておりますけれども、業務利用と業務外利用の双方を含む個人の利用経験者は、国民の半数を超え、令和5（2023）年度から令和7（2025）年度にかけて6倍超と

急増しております。

上記利用経験者の利用実態といたしましては、日常の行動・判断に関する相談を定期的に行っている方の割合が15%、日常的な会話を目的として利用している方の割合が11%程度に達しているという報告がありました。

近年では、産業分野のみならず消費者の日常生活においても、利用者との対話を通じてタスクを実行し、意思決定や行動を代替し得るエージェント型AIの実用化が進展しております。

そこで、AIの利用主体として、組織内における業務利用者に加え、業務外の日常生活においてAIを利用する個人利用者、すなわち消費者の存在感が高まりつつあるといえるのではないかという議論をいただいております。

次のページをお願いします。

次からは質的な深化について議論をいただきました。

まず、AI技術と消費者接点の変化①というところですが、キーワード検索から繰り返しの対話へというところがございます。

本専門調査会での主な議論といたしましては、消費者が自らAI技術が最適化した情報を探しに行くのではなく、AIとの対話により相互理解を深めながら、情報をパーソナライズすることが可能となっているといった点、また、対話は消費者の状況・迷い・感情・価値観を含んでおり、事業者側のデータと結合されることで、提案の精度と影響力が同時に高まっている点などについて議論をいただきました。下には各専門調査会で御発表いただきました資料を掲載させていただいております。

こちらの議論から、柔軟かつ配慮的な対話型インターフェースを備えた、従来の手法とは質的に異なる「AI版セールスパーソン」が出現しつつあるといえるのではないかという議論をいただいているところがございます。

では、次のページをお願いします。

次のページは、AI技術と消費者接点の変化②というところございまして、選好・問題認識・ニーズ等の形成段階への関わり深化という点が挙げられました。

本専門調査会での主な議論といたしましては、消費者の曖昧で潜在的な選好・問題認識・ニーズ等、以下、選考等と呼ばさせていただきますけれども、こちらを対話のログの中から推論し、一致する商品を提案することが可能となっているといったこと、また、消費者は必ずしも自らの選好等を当初から明確化できているとは限らず、自身の曖昧な「抽象欲求」といったものについて、AIとの対話を通じて整理・具体化される可能性がある。さらに、AIの質問の設定方法、対話の口調等により、選好等形成にも一定の影響を及ぼされ得るといった点が議論されました。

下に各回で議論いただいた御発表を掲載しておりますけれども、そのほかにも、日本経済新聞電子版より、対話型AIに広告を導入する動きについての報道があったというところを御紹介させていただいております。

これらの議論から、一般的には商品の購買が消費者の最終的な行動といえるものの、それ以前・以後のAIとの相談自体が主要な消費行動の一つとなり、相談自体が消費の対象となっているという議論をいただきました。

次のページをお願いいたします。

AI技術と消費者接点の変化③というところです。対話自体の消費価値の増大という点が議論されました。

本専門調査会での主な議論といたしましては、AIを利用するという行動は極めて習慣化しやすく、自己表現、社会的証明等の社会的動機を満たす消費が、AIとの相互作用によって満たされるという点、また、日本では特にAIを人格を持つ存在のように受け止めてしまう傾向があり、AIの擬人化は対話の流れや消費者との相互作用の在り方に大きな影響を及ぼし得るといった点が議論されました。下に各回で御発表いただいた資料を掲載させていただきます。

これらの議論から、AIによる個別化された親身なコミュニケーション、即時的な反応により、AIと消費者の間に信頼関係が形成され、AIがより大きな影響力を持つ結果につながるのではないかということが議論されました。

では、次のページをお願いいたします。

AI技術と消費者設定の変化の現時点での最終局面でございますけれども、意思決定の代替可能性の出現というところが議論されました。

本専門調査会での主な議論といたしまして、AI技術が消費者に意思決定の材料を提供する段階から、意思決定自体を代替することが可能となりつつあること、また、今後、代替される意思決定とあえて支援を受けるにとどまる意思決定が分かれることが想定されるものの、いずれもAI技術の影響力の強化が予想されるといった点が議論されました。下に各回での御発表資料を掲載しております。

以上の論点1の議論から導かれる次の検討課題というところでございますけれども、重要な利用主体である消費者の意思決定に対する影響の実態を把握・評価し、必要に応じて対応を検討することが、「信頼できるAI」の実現のために必要ではないかといった点になります。

次のページをお願いいたします。

論点2でございます。意思決定のプロセスへの影響の適切な評価と対応の方向性についてというところです。

本専門調査会で報告された、対話型AIもしくはエージェント型AIが意思決定プロセスにもたらし得る影響ということで、こちらはニュートラルな変化というところでございます。例えば第7回唐沢委員、第8回丸山座長代理御発表資料に基づいて事務局が作成させていただいた資料ですけれども、左の四角囲いのところでございます。ニュートラルな影響としては以下のようなことが挙げられるのではないかという点が議論されました。

情報の要約から購入誘導のシームレス化、また、同調性、迎合性、忖度、親密性、人間

関係代替性、また、長期対話による心理特性の推定精緻化、例えば個別化の促進と深掘りといった点が挙げられること、また、AIが消費行動に与える新たな影響といたしましては、AI自体が個人の反応を観察し、何をどのタイミングでどのように与えれば、行動変容につながるかを学習できること、また、消費者と会話（相互作用）しながら消費行動に影響する働きかけを行うことといった点が挙げられました。

また、第8回勝又教授の御発表からは、購買プロセスへの影響という点を御指摘いただいたところです。例えば上位の問題認識ですとか、情報探索の部分といったようなところは、AIの影響が大きいのではないかとこのところを御議論いただきました。

以上から、AI技術による意思決定プロセスへの影響について、AI技術によるポジティブな影響を利活用しつつ、消費者の意思決定プロセスに与え得るリスクを評価し、適切に対応することが必要なのではないかとこの議論をいただいているところです。

次のページをお願いいたします。

次のページはリスクという点でございます。本専門調査会で報告された、消費者に対してリスクとなるAI技術利用の主な対応というところでございまして、こちらは本専門調査会において中心的に報告がされたところでございます。

まず、生成AI実装以前からの問題の増幅といたしまして、第2回田中委員の報告では、偽情報生成・欺瞞的説明の高度化・巧妙化・スケール化といったことを御指摘いただきました。

第2回岡崎委員の報告では、プライバシーの期待化の拡大といった点が指摘されました。

第4回坂下委員の報告では、行動の予測による操作という点が指摘され、第8回丸山座長代理の報告では、ダークパターン、パーソナライゼーションの中でも脆弱性を狙う等の対応、中毒性のあるUI・UX設計といった点を御指摘いただきました。

次に、(2) 生成AI実装後の双方向コミュニケーションの拡張に伴うリスクといたしましては、まず、第2回岡崎委員からは、消費者の意図を酌み取り過ぎるがゆえに安全性が低下すること。

第4回野村委員からは、消費者の対話型AI依存傾向に悪影響を及ぼす迎合性、擬人化等が報告されました。

また、第4回坂下委員からは、シコファンシー等の手法による消費者の心理誘導などを御指摘いただきました。

(3) エージェント型AIによる新たな展開というところでございますけれども、第5回の小栗様の報告では、対話の誤りから現実の不利益に至るといった点が報告をされました。

これらの議論に基づき、以下のようにまとめることができるかと思います。すなわち、リスクとなるAI技術の開発・設計・提供について、AI技術を消費者が用い、または用いられる利用場面・文脈によってリスクの現れ方が異なるため、さらなる検討の精緻化が必要と考えられるといった点が議論されました。

次のページ、最終ページをお願いいたします。

最後のページでは、本専門調査会で議論された環境整備の必要性に係る主要な論点につきまして御紹介させていただきます。

こちらのページでは、公開の場で委員の皆様から御発言いただいたもの以外に、事務局にお寄せいただいた御意見も含まれる点、御留意願えればと思います。

まず、生成AIを利用した偽・誤情報による、消費者の認知に対する攻撃コストと消費者個人の対応コストの非対称性、対話型AIとの対話の継続による、消費者への累積・反復的影響、例えばキャパシティを超えた購買ですとか、価値観を超えた意思決定、また、通常納得とはいえない選択等が議論されました。

また、対話型AIによる働きかけが継続的で、個人化され、見えにくいことによる、消費者個人での自衛の困難さですとか、消費者個人による事後的な相談・救済の困難さ、例えば消費者自身が自身の意思決定への影響を自覚できず、苦情として消費生活相談現場に上がらないケースが多数ある可能性が非常に高いという点、また、消費者が消費生活相談員への相談前にAIに相談し、相談員の回答がAIと異なった場合、相談員が間違っているとしてトラブルとなるケースが増加しているような点が指摘されました。

これに基づきまして、以下のようにまとめさせていただいております。個人のリテラシー向上と併せ、AI技術の消費者の意思決定プロセスへのリスクに対応するための環境整備について、具体的に検討する必要があるのではないかという議論となっているというところでございます。

事務局からの報告は以上となっております。

○小塚座長 ありがとうございます。

この調査会の会合での御議論、あるいは事務局から先生方にいろいろ御協力等をお願いしている面もあるかと思います。そういう中で得られた内容が反映されているということですが、御自身の御発言、御意見と違うとか、不足しているとか、あるいは今まで発言していないのだけれども、ここまで書かれるのであればこういうことも書くべきではないかとか、そのような御意見がありましたらどなたからでもこの機会におっしゃってください。いかがでございましょうか。

ここで決めていくというのではなくて、むしろ今ここで御意見をいろいろ集めていきたいということでございます。

オンラインのほうから丸山座長代理のお手が挙がっているようです。お願いできますか。

○丸山座長代理 画面オフのまま失礼いたします。

一番最後のページで、一番最後のところに書かれている文章で表現されているとは思いますが、私の発表の回のときに、AIという技術によって問題を解決できる側面もあるということが指摘されたと思います。ただ、それを活用しようとするときに、AIによってまたそれが妨げられるという可能性もあるといった議論がされましたので、環境整備という中に、技術による防御というのと、それを法制度的に環境整備するという、そういったニュアンスが伝わるように書ければ、そういう方向でも調査が進められるのではない

かと思いました。

意見でございますので、そういった印象を持ったということでございます。

○小塚座長　ありがとうございます。

論点2のほうがそもそもまだ分量が少ないのですが、この先増えていきますので、環境整備のところは、AIの使用から消費者の利益を守るという意味での環境整備と、AIを用いて消費者の利益を実現していくという環境整備と、この2つがあるという形で膨らませたいと思います。ありがとうございます。

そのほかいかがでしょうか。ぜひこういうところで御意見を出していただけますと事務局は非常に助かります。

坂下委員、どうぞお願いします。

○坂下委員　坂下です。

御説明ありがとうございました。よくまとまっていると思います。

私、個人的に関心があるのは、5ページ目にある相談自体が消費の対象になっているというくだりで、実際にそうだとすると、消費者契約とか、広告とか、表示とか、私がやっている個人情報とかいうものが、従来の消費者制度の中でどう守られてきて、それをどう変えなければいけないかというところが論点になってくると思うのです。ですから、ここは委員の先生方ともよく議論したいなと思いました。

感想ですが以上です。

○小塚座長　ありがとうございます。

相談自体が消費の対象となった結果、では何なのだということで、これまでの制度上の概念と、それが実際に消費者に対して持つ機能といいますか、効果が少しずれてくる、あるいは変わってくる可能性があるのと、それに対応していく必要があるのではないか、そういうことですね。ぜひそれは盛り込んでいただけますでしょうか。

そのほかいかがでしょうか。どうぞ。

○坂下委員　もう一つ、今日の議論に関係してしまうのであまり言わないほうがいいかなと思ったのですが、説得と操作の境界線をどう考えるかという話です。どこまでが介入でどこまでが誘導なのかがとても重要ではないかと思っております。そういうところがちょっと気になっています。

以上です。

○小塚座長　ありがとうございます。

事務局、その話はどこにありましたか。

○福島補佐　こちらの資料にはまだ掲載がございません。こちらには載せていない議論にはなります。

○小塚座長　でも、関連のある話はどこかに出ていましたよね。4ページか。この辺りで、また今日のこの後の御議論も踏まえてということですが、しかるべく膨らませていただけますでしょうか。

オンラインのほうから馬籠委員のお手が挙がっています。馬籠委員、お願いします。

○馬籠委員 ありがとうございます。

論点2の8ページの部分で、ポジティブな影響を利活用しつつ、消費者の意思決定プロセスに与えるリスクを評価しというところがあるかなと思いながら見ておりました。評価するところ、購買プロセスへの影響みたいなところを書いている部分があると思うのですけれども、例えばもう少し絞って、どこに影響が多くありそうだとするならば、もうちょっとその中の購買の部分でのもっと細かい部分を見ていくとかということがあると、もしかするといいのかなと思った次第でした。

以上でございます。

○小塚座長 ありがとうございます。

スライド8枚目に前回、勝又教授がお出しくださった資料、その後ちょっと修正をしていたようですが、それが入っていますけれども、このプロセスの中のワンステップごとにAIの影響があるわけです。これが大きい小さいというはあるわけですが、少なくともそれをきめ細かく見ていくということは必要だと思うのです。AIによって消費者が大きな影響を受けると言っているだけではいつまでたっても政策になりませんので、そこは購買プロセスへの影響を分析し、きめ細かく見た上で、評価した上でということなのだと思うのです。その上で、このリスクは適切に評価し、対応する必要があるということだと思います。

○馬籠委員 ありがとうございます。

○小塚座長 それでよろしいですか。

それぐらいであれば事務局、書けますね。

○福島補佐 また御相談の上、検討させていただきます。

○小塚座長 分かりました。

大塚委員もお手が挙がっているようです。よろしくお願いします。

○大塚委員 ありがとうございます。

私からは、この資料には必ずしも明確には表れていない点について発言がございました。

資料としては、全体を通してということになるのですけれども、どういうAIを念頭に置くかという点も明らかにされるとすごく分かりやすくなるのかなと思います。

生成AIの使われ方としては、恐らく主に念頭に置かれているのが、消費者が生成AIを利用して、そこに相談をしたり、あるいはこの商品を買ったらどうかということ、これも相談したりするということだと思いますが、資料の中には、事業者が生成AIを使って消費者の購買行動を促していくというパターンであるとか、あるいは、これも両方の間ということになるかもしれませんが、売手である事業者が生成AIを利用者に提供しているというパターン、ある種宣伝のような形で利用しているパターンがあるかと思います。また、個人が生成AIを利用するパターンも、契約関係に結びつく利用の仕方と、そうではない依存してしまって社会から切り離されてしまう、そういった危険性という点というのは、恐ら

く問題の出方は違うと思いますので、そういった観点からも整理があるとよいかと思います。

今回は恐らく論点との関係で簡潔に説明していただいたので出ていないという話なのかもしれませんが、そういった点も今後含めて整理していただけると大変分かりやすくなるのかなと思います。

以上です。

○小塚座長 ありがとうございます。

報告書を取りまとめる段階になりますと、恐らく冒頭でAI技術の動向が書かれなければいけないので、その中で今、大塚委員が言われたような利用形態、想定される利用形態ですね。社会実装が生成AIは大分進んできましたけれども、まだAIエージェントは社会実装がようやく始まったということです。想定される利用形態は書いていかなければいけないと思います。

それから、論点との関係では、今、大塚委員が言われたことが、前回、丸山座長代理のお話にありました利益相反の問題とかにつながってきますので、それはそれでまた論点のほうできちんと受け止める必要があるのではないかと思います。

以前にもこの場で御了解をいただいているはずですが、座長代理と大塚委員と私で若干その辺り、その辺りというのは、法制度的に見てこういう論点がどのように散らばっているかを少しマッピングしていきたいと思いますので、それを踏まえてまた事務局に資料をアップデートというカリバイスしていただいて、皆様のお目かけられればと思います。

唐沢委員、お願いできますでしょうか。

○唐沢委員 資料をおまとめいただきましてありがとうございます。

主に6、7、8ページの辺り、特に今出しておられる8ページの辺りなのですが、AIが入ってくることによって、AIの擬人化、それからAIがいろいろ消費者の情報を蓄積し、それをAIが一種記憶を持ってパーソナライズしていった働きかけていくという中で、消費者のそのときそのときの行動に影響するということについて、いろいろと御記載いただきましたし、社会的な背景も書いていただいていると思うのですが、どこまで踏み込んで書くべきかというのは迷うところかもしれませんが、そのようなAIの働きかけのもう少し本質的に怖い部分というのは、そのときに物を買うとか、消費というところの意思決定とか、こういうものもいいという選好形成とか、その辺り以上にそれを超えてむしろ自己価値みたいなものを、消費行動を通じてAIが形成していくという中で、消費者に非常に強い影響を与えるというような、長期的な影響みたいなことが重要かと思います。

飛び飛びにいろいろと書いていただいていると思うのですが、その辺りをどのように読み取ったらいいか私自身も難しかったので、例えばある時期迷ったことが将来の働きかけにも使われてしまうとか、過去の行動が将来の提案を拘束してしまうとか、何かしら時系列を通じて起こってしまうような、それが自己というものに影響を与えてしまうと

いう辺りを少し踏み込んで御記載いただけると厚みが出るかなと思いましたので発言させていただきました。

御検討のほどよろしく願いいたします。

○小塚座長 ありがとうございます。

なるほど。そこは私自身も認識が不足していましたけれども、確かに伝統的に消費者取引というのは、その場面だけを想定して議論してきたわけですが、そこにAIが使われることによって、将来のその人の選択あるいは選択できる幅に影響が出てしまうのではないかと思います。この御指摘は非常に重要な点だと思いますので、うまく盛り込んでいけるように、さらに事務局と御相談したいと思います。唐沢委員、どうもありがとうございます。

それでは、これは論点の検討状況を整理してみましたというペーパーですので、また少し事務局のほうでこれを充実させていただいて、皆様といろいろ御議論をさせていただきたいと思います。取りあえず今日は頭出しでした。ありがとうございました。

《2. (2) AI技術と消費者問題について

有識者ヒアリング（矢谷 浩司 東京大学大学院情報理工学系研究科電子情報学専攻教授）》

○小塚座長 それでは、次の議事に進ませていただきます。次の議事は、議事次第上では「消費者によるAI技術の適切な利活用に向けて」ということなのですが、本日は有識者として東京大学の矢谷教授にお越しいただいております。

最初にプレゼンテーションをいただきまして、その後、質疑応答・意見交換とさせていただきます。

矢谷教授ですが、東京大学大学院情報理工学系研究科電子情報学専攻の矢谷浩司教授です。ヒューマン・コンピュータ・インタラクション（HCI）という領域を御専門としておられ、人のウェルビーイングに資するAI・対話システムの設計と評価に関する研究をしておられるということです。本日御用意いただいた資料ですが、「消費者の主体性とAI Human-AI Interactionの視点から」という資料を御用意いただいておりますので、そのお話がいただけるということです。

お忙しい中お越しいただきまして、ありがとうございます。それでは、どうぞよろしくお願いいたします。

○矢谷教授 東京大学の矢谷でございます。本日はお招きいただきまして、ありがとうございます。

私のほうから、私が行っている研究を中心に少し話題提供させていただければと思いま

す。

私自身は、先ほど御紹介いただきましたけれども、矢谷浩司と申します。専門はヒューマン・コンピュータ・インタラクションというところで、人と情報技術が接するところを研究としています。最近、特に人のウェルビーイングに資するようなAIあるいは対話システムの設計をいろいろな分野の先生方と手を取りながらさせていただいているという人間でございます。

HCIという分野はかなり広うございまして、いろいろな研究スタイルがあるのですけれども、私は作ったシステムを実世界において、それによって人がどういうふうに振る舞うのかということを研究するというスタイルを取っています。AIであったりメディアあるいはインターフェースによる介入を使って、人がどのような振る舞いを見せるのか、あるいはそれによって人の振る舞いがどう変わるのかということを、データを集めて検証するというようなことをしています。

やることとしては、実際に我々がシステムを作って、それを一般の方に使っていただく。100人前後ぐらいに何週間ぐらいか使っていただいて、データを集めて分析するということをします。その中で、どういうふうに使ったかというような量的なデータと、あと実際にどういうふうに使ったのか、これは何に使ったのですかみたいなインタビュー、質的な調査というのを組み合わせて、人の振る舞いを解明するといったことをやっております。

ここでも話題に挙がっています生成AIですが、様々な方が多く使われていて、今ここでお見せしているのはアメリカのとある会社のレポートですけれども、2025年になって、セラピーだとか自分の生活の整理あるいは人生の目的探してみたいな、いわゆる自分自身に関係するような使い方がかなり上位に上がってきているという状況にあります。ここから見てとれることとしては、AIというのが単なる情報を集めてくるようなツールではなくて、自分の話し相手、相談相手みたいな形になってきていると、当然、消費者の購買行動や意思決定にAIが入ってきて、これは買ったほうがいいかなとか、こっちとこっちどっちがいいと思うみたいな会話が出てくるのは当然の流れだろうということなわけです。

そういった中で、人とAIの関わりから見たときに、AIとどう付き合うかということの難しさもいろいろ研究で知られています。今日はまずそれを少し御紹介したいと思いますが、まずはパーソナライゼーション、いろいろ個別化していくという話がありましたが、それが必ずしもいい方向に向かうかという、そうではないという話をさせていただきたいと思います。

これは私の研究ではないのですけれども、カガワ先生とババ先生が行った研究で、3つの意思決定・予測タスクを設定しています。客観的なタスク、客観的に正確・不正解が分かるタスクと主観的なタスク、自分の感情とか好みに依存するようなタスクと、あと高リスクなタスクという3つを用意して、ユーザーの判断基準に合わせる度合いをAIの中で調整するということをして、それによって人の振る舞い、意思決定がどういう変わるかということを検証したのになります。

ここで面白いことがあったのは、客観的、主観的なタスクでは、AIの説明がユーザーの判断の基準に近づく、ユーザーの考え方に近づくほど、AIに対しての同意率が上がって判断を変えるということがある。一方で高リスクなタスクの場合には、近しいとか遠いというのはあまり関係なく、判断にもそれほど影響しない。ところが、主観的な指標、満足したとか、よく信頼できたとかというのは上がってしまうというのがあります。つまり、必ずしも意思決定の質に貢献しないということなわけです。

なので、消費者にとって非常に難しい判断を要求される場合、例えばものすごく長い約款を読まなければいけないとかもありますけれども、そういった場合には個別化されたAIの説明が必ずしもよい判断にはつながらないのですけれども、自分に対する満足度が上がってしまうので、いい判断をしたと思ってしまうみたいなことがあり得るというわけであります。

さらに、ここでもさっきおっしゃっていただきましたけれども、いわゆる迎合的な振る舞いというのがLarge Language Modelにあるということが確認されています。これの一つの原因は、過去集めたデータの中で、AIからの応答があったときに、この応答はあなたにとってよかったですか、悪かったですかみたいな、○・×をつけるみたいなことがあったわけです。それを大量に学習すると、当然ユーザーにとっていいものが出てくるようになるのですけれども、ユーザーにとっていいというのが、ユーザーが求めていることが出てきたということも含まれていますし、ユーザーにとって心地よい回答が出てきたということも含まれてきて、いいとなっているわけです。それを一個ガッチャンとすると、当然ユーザーにとって心地よいものを間違った形で表示するみたいなことも出てきてしまうというわけです。

これ自体はテクニカルには解ける問題だと私は思います。ところが、Large Language Modelというビジネスモデルはサブスクリプションなわけです。つまり、月々払って何ぼというものです。そうすると、ユーザーに継続的に使ってもらわなければいけないわけです。となると、当然ユーザーにとって心地よい回答をしてくれるものが好ましくなるよねという話に当然なってくるので、そういった社会的な構造として、なかなかこれをなしにしてくださいというのは結構難しいのではないかなと私は思っています。

もう一つ、そういった中でやることとしては、Explainable AIといったように、AIがこうこう応答をしました、それはこうこうこういう理由ですよみたいな形で、AI自身になぜそれが起きたかということを説明させるという話があります。これもやるのは大事なのですが、やり方が結構重要で、これもほかの方の研究なのですが、3種類のAIの振る舞いに対する説明のスタイルを提示して、それがどういうふうにユーザーのAIの振る舞いの理解につながるかということを検証したものです。

ここではAIの振る舞いを正当化する自然言語があるのと、全く正当化しないでこういう動きをしましたみたいなものと、あとQ値というよく分からない数字がただ並んでいるみたいな、そういう条件を表示して、AIの知識がある人となない人とでそれぞれどういうふう

に解釈するかということを実験したものです。

このときに、数字の条件に特にフォーカスして今日はお話をしますが、AIの知識のある人は、この数字を見て、何となく理解しようとするのです。理解しようとするのですけれども、よく分からないデータなので、結局意味づけはあまりできないのです。意味づけはできないのだけれども、数字がちゃんと意味がありそうだなと思って、よく頑張っているよねみたいと思う。一方で、AIの知識がない人は、数字がいっぱいすごみたいな感じになって、ロボットは賢そうだねみたいな感じになって、AIを信じてしまうみたいな感じになるというわけです。

これは結構重要なことを言っていて、例えばよくあるのはAIの確信度みたいな、コンフィデンススコアみたいなやつです。この回答、僕はこれが自信ありますよみたいなものがあるのですけれども、それは技術的には確かに正しいのかもしれないのですが、そこにユーザーの理解が伴わないと、結局それをユーザーがいいように解釈してしまうということがあって、それをここの著者の人たちはExplainability Pitfallsという言い方をしています。なので、AIシステムあるいはその設計者がよかれと思って公開的な情報を出したとしても、消費者に理解されない形だとすると、ある意味では意図されない形で使われてしまうということも起こり得るというわけです。

さらに、AIには必ず何かしらのバイアスが存在します。それは学習データもそうですし、内部のアルゴリズムもあるわけです。その影響を調べた研究がありまして、ここではソーシャルメディアは社会にとって善ですか、悪ですかという作文してもらおうというタスクをお願いしています。その作文をする中で、AIから、自分が書いた文章をベースに、次はあなたはこういう文章を書きたいよねというサジェスションが出てくるようなインターフェースを提供しています。

この実験のトリックは、バックエンドに使っているAIに、このAIはソーシャルメディアはいいものだと思い込んでいるものと、悪いものだと思い込んでいるようなものを入れています。そうすると、ここでもそうなのだけれども、最初、自分の文章はAIというのは、ソーシャルメディアはよくないよと言っているのですけれども、途中からいいよという文章に変わっているのです。それを実際に使ってもらってどうなるかというのを見ると、実際にバイアスに流されて、一定程度いいよあるいは悪いよという方向に流された文章を書くということが示されています。

この研究はほかの人の研究なのですが、私はすごく面白いと思って、この間、とあるグループにグループディスカッションをして、AIに関する話をしてくださいというのがあって、これをやってみました。

どんなことをやったかというと、そこでは30名の社会人の方々、比較のお若い方々です。AIを仕事やプライベートで毎日使っています、複数のアカウントがありますみたいな人も結構いたという状況です。5名1グループで、これからの社会人が持つべき趣味、若い人も仕事ばかりやっていてもしようがないので何かしましょうと。そういう中で、新しい趣味

をぜひ見つけましょうよというのを、AIを使ってディスカッションするというていのグループワークを設定しました。ここで我々がLarge Language Model、自分たちでチューニングしたチャットボットを提供して、これも自由に使ってくださいよという形で提供しました。

これは結構面白くて、いろいろな趣味が出てきたりとか、AIとしゃべって新しい趣味を生み出して、ジャグリングとジョギングを組み合わせでジョグリングとかいうものを生み出して発表するとか、結構盛り上がったりとかしました。それでわーっと盛り上がった後に、実は皆さんに配ったやつの中で、チャットボットにバイアスを加えていましたよという話を種明かししたという感じです。それをやってみると、もちろん何をもってインドアとするか、アウトドアとするかというのは結構難しいのですけれども、おおよそエイ、ヤァと分けると大体こんなような感じで、うまいこと流されるというのが分かったというわけです。

ここで結構重要なのは、自分が私はこれが欲しいですとか、私はこれじゃないと駄目なのですみたいなものがあれば、あまり流されないと思うのですけれども、どっちでもいいかもしれないようなときには、AIが言ったことに恐らく相当数流されてしまうということが起こり得るということでもあります。

さらに、こういったAIを長期的に使っていくと自分たちにどういうふうなことから起きるかを調査した、別の方の研究があって、ここでは23人のナレッジワーカーが毎日ChatGPTを使って、今日はどんな仕事をしましたかとか、どんなことを達成しましたかみたいなことを聞いていくというようなことをしました。

それを聞いて集めていくと、生成AIをうまく使えて仕事が終わったというような、そういうやったという気持ちもある一方で、それをやっていくうちに、自分が何か難しいことにチャレンジしている感覚とか、あるいは自分が仕事を持っているという感覚、そういうものがなくなってしまうというようなことがあるわけです。

これもある意味消費者の考え方に立つと、AIにどんどんお任せして決めていくと、何となくその場はうまくいっているのですが、後から、これは本当に自分が考えたんだっけみたいなことが起こり得るのではないかなと思います。

こういった昨今のAIというのは、人を支援する形で設計されているわけです。これはもうまさに今のAIのトレンドで、ここに挙げていますけれども、人を支援するツールというのは3つの形に大別されるとなっています。既存のAIの在り方というのは非常にこれに当てはまる形で、人を補助する、あるいは人のない能力を補完する、あるいは人がそもそもの今持っている力を増強するといった形で使われるわけですが、支援するAIというのは、その場では人の能力を高めるわけですが、いざそれがなくなると結局ユーザーはユーザーのままなわけです。必ずしもそのユーザーの能力自体に何か特別な変化が起きることはないというわけであります。

さらに、そのようなAIに過度に依存してしまうと、そのタスクを行っていくために必要

な能力を失ったりとか、それだけではなくて、いわゆる主体性だったり関与感、あるいは自分がタスクをやったという達成感みたいなもので下がってしまうというようなことが研究で知られているというわけです。

そういった中で、我々の研究グループでは、促考するAI、考えを促すAIという設計思想を考えています。これは単にこれどうですかと言われたときに、あなたに当たる答えはAとBとCですみたいな感じで出すだけではなくて、その人に対して考えさせる質問を投げかけたり、あるいはほかの視点を示すというようなことをやるという設計です。これもいろいろなやり方というのがあって、それをここでは7つ挙げているのですけれども、そういったものをうまく組み合わせると、人を考えさせることに巻き込ませるようなAI設計ができるのではないかと考えています。

例えば別の選択肢を好む消費者がAIで模擬的に作って、その人としゃべって、どっちのほうがいいかなみたいなことをしゃべってもらうとか、あるいはAIに対する応答に対して、別のAIもそれを見て本当にそれでいいのみたいなことを言うみたいな、横から茶々を入れるみたいなやつですね。そういうものがあってもいいかもしれないし、あるいはAIに誘導されている消費者を横目で見て、何かだまされているのではないかなみたいなことを見るみたいな。第三者で見ると、自分がそこの中に入ると分からないのだけれども、横で見ているとそうかもしれないみたいなことが結構分かったりしますので、そういったことをある意味シミュレーションするみたいなことを考えるとこういうことがあります。

あるいは、よくあるのは正当性バイアスみたいなものもかかるので、買ったらいいことが起きるみたいな感じで考えているけれども、そうではなくて、買ったときに最悪のシナリオみたいなものをAIに見せてもらって、ひょっとしたらそれはあるかもしれないみたいな感じで、少し歯止めがかかるみたいなこともあるかもしれない。

一応私のほうでこういったことを入れ込んで、ChatGPTとかGeminiのほうで作ったものがありますので、もしよければ試していただければと思います。必ずしもうまく動くとは限らないのですけれども、こういったことをやっているということでございます。

それを話した上で、とはいえやはりやっていかなければいけないことはありますし、私もHCIの人間として人に対して貢献したいという思いがありますので、私なりのアイデア、種を持ってきたのを次にお話ししたいと思います。

もちろんこの問題というのは非常に複雑で、単にAIがよくなればいいよねという問題ではないと思います。それはもうこの委員会でも十分にお話をされているところかなと思いますので、ここでは人とAI、それから制度というところに分けて、この3つのタグでうまくやっていくということを考えられればいいのかということをお話しできればと思います。

1つは、AIの利用状況の内製を促すようなインターフェース設計ということで、自身とAIとの会話でどの程度健康的な対話をしているかということを見せるというものです。例えばAIに対して、君の言っていることはここが間違っているじゃないか、みたいなことを言

うとか、あるいはこれではないやつを教えてよ、みたいなことをどれだけ言うかみたいなやつです。

皆さんいろいろスマートウォッチとかアクティビティトラッカーとかをお使いかもしれませんが、歩いていると、この写真にもありますけれども、これだけあなたは歩いたよねみたいなものが出てくれますので、そういった形で、今日AIとしゃべった中でこれだけあなたは突っ込みましたみたいな、突っ込みポイントみたいなものを出すみたいなことができる面白いのではないかなと思います。

実は健康的な身体活動に対してインセンティブを与えるということをやっている政府の活動はあって、シンガポールに行くとヘルスポイントというのがあって、歩くとポイントをもらえますみたいなものがあるのですけれども、それと同じで、AIに対して何か考えたらポイントをもらえるみたいなことをやったら面白いのではないかな、みたいなことを思ったというわけです。

それから、これは私たちが行った研究の一つなのですが、AIとの対話によって認知的な接種を実現することができるということが何となく分かってきています。社会心理学の認知的接種というものの、いわゆる予防接種みたいなものの考え版みたいな感じなのですけれども、それをAIとの対話で行うということをやりました。

どんなことをやったかという、AIがわざと誤解を支持するような立場を取って、ユーザーに対して、いやそれは君が間違っているよ、みたいなことを言わせるというようなことをします。そうすると、ユーザーが自分なりに反論するとか、どうやってこのAIを説得しようみたいなことがあって、それを通して、もしこういう誤情報みたいなことがあったときに、どういうふうに反論できるだろうかみたいなことを、ある意味練習するみたいなことができるというわけです。実際に単に読んだりとか書いたりする、エッセイを書いてくださいというものよりも、参加者の健康誤情報に対する耐性を高めたという結果が出ているというものになります。

これをある意味では全国民にやってもらいたい、いわゆる今、予防接種はいろいろなところで子供さんとかは受けますけれども、その情報学的なものができる面白いのではないかなと思っています。例えばあらかじめ告知した上で、意図的に誘導するとか迎合的な応答をするようなAIサンドボックスを使ってもらって、こんなふうになるんだみたいなことを見てもらったりとか、あるいはAI同士がしゃべっていて、どちらかのAIがだましにかかっているみたいなものを見て、ちょっとそれはだまされているよ、みたいなことをやって、訓練をするみたいなことができれば面白いのではないかなと思います。

AIというのは非常にスケールが大きいので、それでかつ個別化もするところなのですけれども、むしろそれを逆手に取って、先にそういうのを体験してもらえば、次あったときに実はどこかで見たぞ、みたいな感じになることはできたりするのではないかなと思います。

それと併せて、AIリテラシーというものの、これも先ほどのスライドにもあったかと思い

ますが、それをどう計測するかとか、そういったところのアセスメントのやり方、その見直しみたいなものも定期的に行っていくみたいなことが大事なかなと思います。

AI開発は世界でいろいろ行われていますけれども、もちろんテクニカルなところも大事なのですが、今、AI開発に対する私の個人的な意見では、単に技術を開発するだけではなくて、それがどう応用されるかというところも大事で、そこが両輪になっていないと結局よくないことが起きまくってしまうという感じなのかなと思います。なので、どう使うかといったところで、国民が高リテラシーを持っていて、その中でAIをバランスよく使えるような形、そういうフィールドを持っているのは日本だけなのです、みたいなことができれば、ある意味では日本のAIの開発だったり応用の大きい勝ち筋になるのではないかなと私は思っていたりします。

それから、AIのベンチマークなのですけれども、最近のいわゆるベンチマークというのは、ある入力に対してその答えのどれが合っているかということなのですけれども、いわゆる正しいとか好ましいということだけではなくて、情報の多様性とか、あるいは多視点性みたいなものですね。ちゃんといいところ悪いところに触れていますかみたいな、そういうところを評価するようにできるといいのではないかなと思います。

さらに、先ほどのスライドにもありましたけれども、単発の応答だけではなくて、ずっと対話を重ねていくわけです。なので、その中で、結局最初がいいことを言っているのだけれども、だんだん結局一つになっていますではあまり意味がないので、しゃべっていくうちにちゃんとそれを中立的に、こっちもいいのだけれどもこっちもどうですかみたいなことをちゃんと言いながら、バランスの取れた対話ができるかというところを評価できるといいのではないかなと思っています。

この辺り、過去の研究、Exposure diversityとか、search result diversificationとか、情報検索とかの辺りとかにも知られているものはありますので、そういったものをうまく活用できるといいのではないかなと思います。

それから、AIモデルを評価するというのは大事なのですけれども、ユーザーが実際に使うのはシステムなのです。そこにユーザーインターフェースが乗っかってきているわけです。シンガポールのAI Safety Instituteは、AIシステム全体を評価することをちゃんと考えましょうということを実は言っています。これは結構大事で、例えばその応答がよかったとしても、バーッとある特定のAIの出力があったときに、複数の意見がどこに出てくるかでもユーザーの注意力は変わるわけです。長いテキストの中の真ん中にチョロっと書いてあってもそれは読まないだろうみたいな感じになってしまうわけです。なので、実際的な消費者が行うインタラクションに即した検証をやっていかないと、本当にシステムを評価したことにはならないのではないかなと思います。

ここもある意味ではAIである種シミュレーションできるところもあるかなと思いますので、そういったところを活用しつつ、HCIだったりソフトウェア工学はこの辺やっていることはいっぱいあると思いますので、その辺りの知見を集約させて、できることを一緒に模

索させていただければと思います。

最後、制度です。これについては、私は全然専門家ではないので的外れなことを言っているかもしれませんが、ぜひいろいろ御意見をいただければと思うのですが、消費者が特定の購買や契約に誘導されてしまうというのは別に人間でもあるわけです。悪い人間がへへっと言いながら、これを買いなよみたいなことはあるわけなのですけれども、AIならではの難しさを考えてみると、一つは、だまそうと思っているかどうかを立証することが極めて難しいというところがあります。完全にブラックボックスのものが多いです。そういうところと、あと責任分界です。どこまでがAIが悪さをしたことになるのか。あるいは、それがなっただとして、AIを作った業者あるいは企業というのはどこまで責任を持つのかみたいなのところも難しい。昨今のAIモデルというのは、いわゆる皆さんが知っているような大企業、片やこっちは一消費者なので、訴えても相手にされないとかいうことも当然あるわけで、非常にパワーバランスがいびつになっています。

さらに、人が主体的に意思決定したと判断するのはすごく難しい。これはどういうことかということ、AIに頑張ってもらって、実はAIに完全に流されているのですけれども、人は私が決めましたと思ってしまっているかもしれないわけです。そうすると、あなたはだまされたよと言われても、私が決めたのですと言われてしまうと、どうしようもなくなってしまうということになってしまって、その辺りが非常に複雑に絡み合っているという状況なのかなと思います。

とはいえ、これも言っていてもしようがないので、ある程度はやっていきながら探っていくということが必要になってくるのではないかなと思っています。なので、例えば人間AI共生特区みたいなものを作って、アジャイルガバナンスみたいな形でいろいろな先行実験を行って、どういうリスクがあるかということを見ながら、制度であったりシステムだったりというのを作っていく必要があって、この辺りは情報工学だけでやってもしようがない話で、いろいろな社会学、法学、様々な方と一緒に、そういうチームを作ってやっていくことが必要だろうと思います。

それから、ダークパターンというところですね。インターフェースのダークパターンというのは最近消費者庁も対応していただいている非常にありがたいところなのですけれども、こういうものというのは通常単発の操作ですね。ある画面があって、例えば消すボタンがすごく分かりにくいとか、サブスクリプションを解約しようと思って100回ぐらいボタンを押さなければいけないとなっているわけです。

いわゆる単発の操作で固定的な設計があって、通常ある種の悪意を持って設計されている。やらせたくないとか、あるいはこっちに誘導しようというのがあのですけれども、AIとの対話で起こるダークパターンというのは、複数マルチターンの対話があつてどこかで起きる可能性もあるし、それが動的な振る舞い、AさんとBさんとCさんと全然違う振る舞いになっていることもあるし、さらに言うと、悪意がない場合があるわけです。たまたまそうなってしまうていたみたいな、たまたま流されていた、たまたまバイアスのかか

ったものがありましたみたいなことになってしまうわけです。そうすると、結構今のデータパターンの考え方では救うことが難しいのではないかなと思います。

ここも何か答えがあるわけではないのですけれども、企業活動を萎縮させることがないように配慮はしながらも、悪意があるから駄目ですという判断だけだとそこは難しくなってしまうので、ネガティブな効果が起きてしまったらそれをうまく救ってあげられるような、ただし企業さんとかAIを提供している人が過度にペナルティーを受けないようにするよう、そのバランスを取りながら消費者保護と制度設計をしていくみたいなことができるのではないかなと思っています。

というところが私からの話題提供となります。以上でございます。

○小塚座長 ありがとうございます。

私どもにとって大変有益なお話をいただきまして、ありがとうございます。今日の前半でお示した論点整理にもいろいろと関わってくるのではないかなと思います。

それでは、早速、委員の皆様から御発言を募りたいのですが、既にオンラインのほうから野村委員のお手が挙がっているようです。野村委員、お願いします。

○野村委員 矢谷教授、非常に貴重なお話をありがとうございます。

お話の中でリテラシーの重要性をお話しされていたのですが、実はこの委員会の中でリテラシーだけでは不十分だみたいな意見もあったりしますし、最近MITのレポートで見かけたのが、AIがこの間暴走するような件がありましたが、ああいうふうな暴走がこれから起こるよということを企業があえて懸念して、責任逃れに使うという風潮が生まれかけているという指摘があったのですが、そう思うと、AIを展開する側もAIを利用する側も今の状況に既に疲れてしまっていて、そこから新たな規範とかリテラシーとかに取り組むだけのパワーがあるのかどうかというところに少し懸念があるのですが、それについてどのように思われますでしょうか。

○矢谷教授 ありがとうございます。

おっしゃるとおりでなかなか難しい問題で、リテラシーだけで私も解決するとはあまり思っていないのですけれども、複数の防御ラインを設けておくということは大事だと思います。それが個人のレベルであるのと、技術を提供する側というのと、社会全体というところになるということで、今回この3つということで私は書かせていただきましたが、その意味で言うと、リテラシーとしてやれることも多分いっぱいあると思います。AIが非常に大きく、毎月毎月言っていることが変わるぐらい今、進展が速いので、なかなか追いついて1個の教科書を作って終わりという形にならないというところでいえば非常に対応は難しいのですが、とはいえやはりそこはやらざるを得ないところがありますし、ある意味ではその柔軟性をうまく使って、いろいろなレベルの人たちに例えばいろいろな年齢の人たちにAIを試してもらったりとか、こういうケースがあるよみたいなことを試してもらいたいなことをスケール大きくやるというのは、ある意味ではAIの得意なところなので、それはできるチャンスはあるのではないかなと私は思っているところであります。

あと、AIが暴走するというのも、企業としても予期しないことというのはやはりあるので、そこは情報開示というよりは情報共有を積極的にさせるようなインセンティブなのか、メカニズムみたいなものがないと、どんどん隠してしまっておしまいみたいな感じにはなってしまうのだらうなと。今回のようなケースはたまたま抗議を受けた側が言ったから分かったようなところもあると思いますので、そういったことを報告させるみたいなことというのは大事なのだらうなと思います。

○野村委員 ありがとうございます。

○小塚座長 ありがとうございます。

オンラインの委員の方がほぼ皆様お手を挙げておられるようですけれども、以下の順序でお願いいたします。まず大塚委員、その後、唐沢委員、馬籠委員、岡崎委員、丸山座長代理、この順序でございます。

大塚委員からお願いします。

○大塚委員 ありがとうございます。

矢谷教授、貴重な御報告ありがとうございます。大阪大学の大塚と申します。

私からは、特にAIリテラシーについて伺いたいと思います。

この研究会でも、野村委員がおっしゃるように、リテラシーの限界については意識をされておりますが、AIリテラシーを醸成することの重要性はずっと指摘され続けております。予防接種というお話はリテラシーをつけるための非常に有用な手段なのかなと私も感じ、非常に勉強になりました。

その中で、予防接種がどこまで有効なのか、予防接種をすることによって、あるいはリテラシーを高めることによって、どこまでAIに対抗できるようになるのかという点をもう少し伺いたいと思います。

具体的には、13ページ、14ページで説明されていた実験についてです。非常に面白い実験だなと思ったのですが、この対象となっている社会人の方は20代から30代ということですので、ある程度AIについては知っているというか、全然知らないという人はいないだろうと思いますし、もしかしたらその中でも使ったことがあるとか、あるいはそれこそ知識としてはバイアスのことも知っている人もいたのかなと勝手に推察しております。

ただ、そういった方でも、かなりバイアスによって影響を受けているということだと思います。ここでもバイアスを受けていった方々はこういったAIとの距離感といいますか、AIについてどれだけ知識を持っていたとか、あるいは30名の方の中にバイアスをより受けていた人、あるいはあまり受けていなかった人がいるときに、AIについての知識とか、そういった経験なんかが影響の受け方に違いがあるのかとか、その点何か気づいた点がございましたら教えていただければなと思います。よろしくお願いいたします。

○矢谷教授 ありがとうございます。

このワークショップでいいますと、具体的にどれぐらい使っているかをアンケート等で測ったわけではないのですが、多くの方はLarge Language Model的なものを何かしら使っ

ていて、数人ぐらいは複数ぐらいアカウントがありますかねみたいな感じだったと思います。企業の中で使っておられるパターンもあれば、御自身のアカウントで使っているパターンもあったという感じだと記憶しています。なので、全く触ったことはありませんとか、そういう人はさすがにいなかったかなというぐらいです。

ただ、そういった方も、バイアスの話、私、これを持って行って実際に見せたのですけれども、全然気がつかなかったみたいな人もいれば、何かこのチャットボットちょっと怪しくないみたいなことに気がついた人も数人いらっしゃいました。何か流されるなみたいなことを言っていた人も実はいて、その辺りは気づいていればあれなのですけれども、このワークショップをやって、このバイアスの話を種明かしして、こういうことに気をつけなければいけないよという話をしたときに、最後に当然質問を受けたのですけれども、そのときにおっしゃったのが、すごく面白いと思ったのですけれども、同時に怖いと思いましたとおっしゃったのです。いわゆる大の大人が5人1組でやってこれだけ流さるのは相当に怖いなみたいなことが分かって、これから気をつけようと思いますみたいなことはおっしゃっておられたので、まさにさっきの予防接種みたいな話に近いと思うのですが、それがどれぐらい長続きするかはまだ分かりませんが、体験することは結構大事なのだらうなと思います。

これはまさに予防接種と同じで、例えばインフルエンザの予防接種も、したから絶対かからないかというところではないわけです。手を洗って、うがいしてとか、人混みを避けてとかというような基本的なことはしつつ、ただ、リスクは下げられるということだと思いますので、まさにAIリテラシー、ある意味ではその程度のものなのですけれども、そこは結構重要に効いてくるところはあると思いますので、何かしらうまい形でできればいいのかなとは思っています。

ただ、もちろんそれだけではカバーできないところがあるので、それを技術の側面であったり、あるいは制度の側面からカバーして、できる限りゼロに近づけていくということが必要なのだらうなと私は思っております。

○大塚委員　ありがとうございます。

使ったことはあっても、だまされた経験といいますか、そういうのがあるとまた別といいますか、単に使っているだけではなくて、だまされたということを気づかされるという経験があると、その後だまされにくいということになっていくのかなと思いました。どうもありがとうございました。

○小塚座長　それでは、唐沢委員、お願いします。

○唐沢委員　どうも御発表ありがとうございました。大変興味深い研究をお伺いできたなと思っています。矢谷教授のなさった研究について確認させていただきつつ、お考えを聞ければと思います。

1つはグループディスカッションの研究なのですけれども、今まで私たち消費者が個人で物を買う場面をメインに想定していたところ、他の人とある意味相談するというのです

か、そういう場面の問題にも視座を広げていただけるような御研究だったなと思っています。実際は、例えば確かに物を買うときに周りの人がどうなのかと、ちょっと話をしたりしてというのがあると思いますし、そういう場面でのプチ流行みたいなものが、しゃべっている部分の中で出来上がって、ということもあると思うので、そういうところのインプリケーションもあるかなと思ったのですが、矢谷教授の御意見を聞き逃したかもしれないのですが、AIにバイアスがかかっているという操作で、それに引きずられるという結果だと思うのですが、一方で、情報のソースとしてAI以外のもの、例えば普通に広告であれ、リコメンデーションであれ、経緯を何か言うということであれ、同じようにバイアスがかかるという可能性は当然あって、そういうものが時にどうなるかというものと、AIの場合との違いにちょっと関心があって、AIがバイアスがかかっているということの意味と、それから、普通にほかに放り込まれる情報ソースが何かバイアスがかかっているということの意味とに何か本質的な違いがあるのかということについて、矢谷教授のお考えをお伺いしたいというのが1点目の質問です。

それから、もう一点確認させていただきたいところがあるのですが、よろしいでしょうか。

○矢谷教授 はい。

○唐沢委員 促考するAIというお話も大変面白いなと思っていて、考えさせるというのですか、そういうことが大事だというのはそのとおりなのだと思うのですが、主体性とか自立性ということは、矢谷教授のお話の中で非常に大事なコンセプトだなと思うのですが、私たちは物を買うときであれ何であれ、主体性とか自立性は全てを自分で決めるということを意味するというよりも、自分の核に沿って決めるという定義のほうが心理学者はむしろ使うと思うので、そうすると例えば任せるということの意味決定がちゃんとできればいいという考え方も成り立つと思うのです。

なので、この判断はAIに任せるとか、こちらは私がやるとか、必要に応じてAI任せを取り返すという話だと思うのですが、促考するAIの中で委任の仕方そのものも視野に入れて支援するというのですか、自分で全部考えるとすごく認知負荷が高いし大変なので、そういうのというのは情報处理的にも私たちは難しいと思うのですが、うまく委任をするというような形で、消費者の行動をより望ましい方向に持っていくようなことに関しての、何かしら示唆的なお考えというのがここから何か出てくるのかどうかということのもちょっと関心がございましたので、この2つをお教えいただけませんか。

○矢谷教授 ありがとうございます。

まず、研究的というかちょっとやってみたいというぐらいなので論文にはなっていないのですが、このワークショップに関して言うと、意図的にバイアスを加えたというのは、その効果をよりはっきりと見せるためということになります。現実的なLarge Language Modelでは、ある種開発者が意図的にバイアスを加えているかということ、そうではないと信じたいと思います。それは残念ながら分かりません。ただし、当然ですけれど

も、データによってバイアスがかかるということも知られています。恐らく既存のいわゆるアメリカで作られたようなLarge Language Modelというのは、ベースとしては英語のデータが対応、これはインターネットに転がっているものを使っているとする、それが非常に比重を占めますので、そこに書かれているものにある程度引きずられているのはあると思います。もちろんそれをオフセットするようなテクノロジーというかやり方というのがいろいろ研究されていて、それが埋め込まれているのだと思いますが、やはり完全にゼロにはできないわけです。それは人間もそうなので、どの人にしゃべっても、自分はこっちのほうがいいと思っているとか、あっちのほうがいいと思っているというふうにしゃべるわけなのですが、人としゃべるときは、この人が例えばどういう人かということは何となく分かっているとか、あるいはこの人はうそをついているなと思うとちょっと目が泳ぐとか、ある種外形的な振る舞いとかいうところもあって、どれぐらい信じるかみたいなことのネゴシエーションがある程度できると思うのです。

ところが、AIはある意味ではどうあってもすごくエキスパートに返してくる。よくも悪くもなので、そうかみtainな感じになるという効果が、人と接して何か情報を得るということよりは強いのだと思います。なので、何かバイアスがもしかかっている応答が出たときに、それに対して引きずられるという可能性は、人よりはひょっとしたら大きくなるのではないかなというのは思うところはあります。

1つ前の研究もそうなのですが、それを分かりやすい形、実際にバイアスがかかっているかどうかは分からないですけれども、もしかかっているAIがあったとして、人がどう振る舞うと思いますかということをはっきりと明かにして、実際につられますねみたいな話になってしまっているというところがメッセージなのだと思います。

続いて、2つ目の促考するAIに関してですが、それはまさに唐沢委員がおっしゃるとおりで、何でもかんでも考えろというのは結構大変なわけです。やはり楽をしたいのが人間の常なので、当然それはそれであるわけです。唐沢委員におっしゃっていただいているとおり、あるところはAIに任せて、あるところは自分でやってというところをやるというバランスをどう決められるかということが大事だと私も思います。

ところが、私もユーザーインターフェースの研究を長年やっているのですが、やはりユーザーは私も含めてですが楽をしたいので、投げられるものはどんどん投げちゃうという感じになりがちなのです。そうするとよくないことが起きてしまうみたいなことがあるので、そのバランスを自分の中でちゃんと組み立てる。それはリテラシーもそうですし、インターフェースの設計としてもそれが必要なのだと思いますが、そのバランスをちゃんと自分なりに組み立てるということ、歯止めをちゃんとかけながら使うということをしていくということが大事なのだろうなと思います。

その一つのやり方として、我々が促考するAIという形で、少しのインターフェースの設計をまとめるみたいなことをさせていただいているという感じかなと思います。

○唐沢委員 どうもありがとうございました。

○小塚座長 それでは、馬籠委員、お願いします。

○馬籠委員 お話ありがとうございました。電通デジタルの馬籠と申します。

質問をさせてください。

23ページ目について質問をさせていただきたくという感じなのですが、健康的な対話というものがちょっと気になりまして、どのようなものが健康的な対応になるのかなというところがあるなと思っていて、逆に言うと、不健全というか、不健康というところが出てくるのだらうなと思っております。例えばやたら対話しているものの言葉遣いが悪いとかというところなのですが、この場合の不健康みたいなものはどういったものになるのかなというところをお聞かせいただきたくというような形です。それは実際に悪いという基準はどのようなものになるのかなと思ったのが気になる点でした。よろしくお願いいたします。

○矢谷教授 ありがとうございます。

ここでいう健康的な対話というのは、ある意味ではある種の自分の考えをどれだけ実施しているかということだと定義をしています。なので、言葉遣いが悪いとか文句を言っているとかそういうことではなくて、ある種建設的な対話を行っているかということです。例えばAIに対してこれをまとめておいて、これをやっておいてみたい感覚ではなくて、レポートを書くみたいなシナリオを想定していただくと、単にこの課題があって、これを何百字で書いてくださいみたいなものではなくて、例えば自分はこういうふうに思うのだけれども、違う意見を教えてくださいとか、あるいはAIで出てきたものに対して、あなたが言っていることのここはどういう意味ですかとか、これは何か違う意見もあるのではないですかみたいな反論をしたりとか、単に受け取っておしまいみたいな感じになっていないかどうかという意味合いで健康的な対話ということをしました。

もちろんこの定義に関してはいろいろあり得ると思いますので、どれだけ認知的に活発な対話をしているのかというところを様々測れると面白いのではないかなと思ひまして、それは言語的なキューからでも大なり小なりはできる部分はあると思いますので、そういったところをうまく可視化して、これぐらいちゃんとAIと向き合っていますみたいなところが見せられると、アクティビティトラッカーの丸もちょっと埋まっていなかったらもうちょっと頑張ろうみたいなことになったりとか、今日はちょっと少ないなみたいなことがあったりとかしますので、そういう簡単な内省にはつながり得るのだと思いますので、そういうものの認知的なバージョンみたいなものがあると面白いのではないかなと思って、アイデア出しをさせていただいたという次第でございます。

○馬籠委員 ありがとうございます。

私は結構投げっ放しだったりするといいますか、健康的に対話ができていなかったりしますので、勉強になりました。ありがとうございます。

○小塚座長 では、次に岡崎委員、お願いします。

○岡崎委員 東京科学大学の岡崎です。

HCIとAIの接点について、興味深い研究を分かりやすく教えていただきありがとうございました。

2つ質問があったのですけれども、ほかの委員の方から類似の質問がありましたので、1点だけ伺いたいと思います。

今日の前半で御紹介いただいた研究では、AIを賢そうな人間に置き換えても同じような結果になりそうに思えたのですけれども、人間ではなくAIを相手にするとどのような特徴的な差が生まれるのか、もしくはその差を引き出すためにどのように実験されているのかということをお教えいただけないでしょうか。

○矢谷教授 ありがとうございます。

岡崎委員のおっしゃっているとおり、人間でも起きることというのは多分あるかなと思います。例えば同じアドバイスを受けるのでも、権威のある人からこっちのほうがいいと言われると、そうかもしれないみたいな感じで流されちゃうみたいなことは人間でもあり得るわけです。なので、もちろんそういうのというのは必ずしもAIだから起きるかという、そうではないと思います。

私自身がその実験をちゃんとやっているかということ、そこはなかなか実は難しいところで、AIはスケールは出せるのですけれども、それを人で置き換えて実験しようと思うと、当然1対1、人を呼んでこななければいけないとか、権威のある先生を呼んでこななければいけないとかなるとなかなかそこは難しいというところがあって、実際に人とAIをダイレクトに比較したというのはあまり多くはないような気はします。もちろん全くないというわけではないと思いますが、そこら辺を緻密にやったことというのは、例えばAIだと、このAIというのは例えば医療情報に関するエキスパートですとか、あるいは医療をよく知っている隣人ですみたいな感じで、いわゆるキャラクターみたいなものをつけてあげて、それをユーザーに開示した上で対話をして、どれぐらい信頼するかとか、そういったところを測るみたいなこととかはあったりするのですけれども、実際の人でどこまでやっているかと言われると、そこはどちらかというと結構物量的な問題でなかなかできていないところはあると思います。

とはいえ、さっきも私、申し上げましたが、人とAIとの一番の違いは、AIのほうは外形的な手がかりが極めて少ない。なので、そこでちょっとこの人は本当に怪しいなとかいうことを思えば、例えば私は今ここでしゃべっていますけれども、矢谷の言っていることはちょっと怪しいなとか、こっち向いたりとかこっち向いたりとかしていたら多分そう思うと思うのです。本当にこの人はこの研究をやっているのみたいな感じになったりしますけれども、AIというのはそういうある意味ではちょっと動揺した振る舞いとか、しゃべり方のトーンとか、そういうものがないわけです。それが、かついわゆる皆さんが知っているような大企業から来たりすると、大抵の人がそうだろうなみたいな感じに聞こえてしまうところがあって、結構そこでもある意味高いところに持って行ってしまっていて、あとはその上下ぐらいになってしまっていて、疑うということをおそらくしないみたいな、

よっぽどへんちくりんな答えが返ってこない限りは、なかなかそこを疑ってかかれない。

例えば自分がよく知っていることを聞いたときに、ちょっととんちんかんな答えが返ってくると本当にそうなのみたいな感じになりますけれども、AIに聞こうと思うことは、大抵自分はあまり知らないことを聞きますので、それで何か自信満々に返ってくると、そうかなみたいな感じになってしまうというのはあるのだろうと思います。

そこが多分AIのある意味では気軽に聞けるというところのメリットなわけですがけれども、それを単に受け取って、それでおしまいにしてしまうとよくないことになってしまうという問題になっているのだろうなどは、インターフェースの設計からするとそうなのかなと思います。

○岡崎委員 ありがとうございます。

○小塚座長 それでは、丸山座長代理、お願いいたします。

○丸山座長代理 慶應義塾大学の丸山絵美子と申します。

本日は、とても御示唆のある御報告をありがとうございます。

民法、消費者法を専門としております。今回の御報告で言うと、6ページ辺りでしょうか。個別化されたAIの説明が必ずしもよい判断にならないというお話であるとか、それと関係した制度設計のほうのお話について質問をさせていただければと思います。

従来、規制や制度設計を考えるときには、客観的に不合理な行動を取っていても、個人が満足していればその個人は幸せであると捉えられていて、よほどのケースでなければ介入しないという考え方が出発点とされてきたと思います。しかし、個別化されたAIなどが必ずしもよい判断にならないとしますと、制度設計を考えるときに、客観的に合理的な行動となっているのかという点に着目するのが重要となるのか、それとも個人に選ぶ機会や再び考え直す機会を確保していくということが重要となるのか、この辺りの御感触的なものでもございましたら教えていただければと思います。

○矢谷教授 ありがとうございます。

なかなか難しいところかなと思って、私もあまり明確な答えがあるわけではないのですが、もう一度、質問の最後のところだけ繰り返していただけますでしょうか。

○丸山座長代理 2つのことを言いました。AIが迎合していろいろ言うことが、必ずしも消費者のためにならないけれども、満足が高まるという現象が増えていくだろうというところで、制度設計を考えていくときに、そうであれば客観的に合理的な行動や意思決定になっているのかというところに着目して、様々な問題を捉えていくのがよいのか、それとも、こういった特性があるのだから、何か選択肢を消費者に増やしていく、あるいは再び考えるような機会を確保していくというのが重要となりそうなのか、何かこの辺り感触的なものがあったら教えていただければという質問でした。

○矢谷教授 ありがとうございます。

それで言うと、恐らく私の答えとしては後者のほうかなという気はします。多分今の Large Language Model の振る舞いもそうなのですから、ある種の一つの選択肢に偏っ

てしまうというところは結構あると思います。昨今のインターフェースの設計は、一つのAIモデルもしくはエージェントと対話をするという形になっていて、その考えしか見えてこないわけです。そのAIエージェントがある意味ではうまく対応的に振る舞ってくれれば問題ないのですが、例によってパーソナライゼーションであつたりとかおべっかの問題があつて、ある種一つの見方に偏っていくというところがあつて、結局一つの偏った見方というところが強まっていくということがあると思います。

今日は御紹介しなかったのですが、私たちの研究室でやっている研究では、複数のAIエージェントと対話をするみたいな環境を作つてあげて、この人はこういう性格です、この人はこういう性格でこういう意見を持っていますみたいな人と一緒に対話をする、自分と同じような問題を抱えているとか同じような問題意識なのだけれども、自分と違う対応をしたりとか自分と違う考えを持っているというのが見えて、そこから学び取るみたいなことがユーザーに発生するということが実は見えています。それは人間同士でもソーシャルラーニングみたいな形であつたりするのは知られているのですが、それと同じような効果が複数のAIエージェントとの対話でも見られていますので、ここには書かなかったのですが、一つのやり方としては、複数のAIエージェントと何か対話をして、私はこっちのほうがいい、私はこっちのほうがいい、私はこっちもいいと思うみたいな対話をしながら、あなたはどっちのほうがいいのみたいなことを考えるみたいなことができれば、そこら辺はもう少しインフォームドな形になるのではないかなと思います。

○丸山座長代理　ありがとうございました。

○小塚座長　それでは、会議室の皆様、お待たせいたしました。

田中委員からどうぞ。

○田中委員　矢谷教授、どうもありがとうございます。

特にモデルの評価だけではなく、インターフェースも含むシステム全体の評価が重要だという点は私も非常に重要だと思いました。

私のほうからは、2つの立場に関して質問をさせていただけたらと思います。一つはユーザーの立場で、もう一つはエンジニアの立場です。

一つ、ユーザー側については、先ほどのワークショップのように、事前にバイアスを知らされていないけれども、後から実はバイアスがありましたよということで気づきを得てもらって、ある種の予防接種のような効果を期待するというお話がありました。偽・誤情報対策でも、予防接種の理論を応用した介入研究はされていますけれども、その中で出てきているのが効果のスピルオーバーの話で、それだけに対して注意をしてくれればいいけれども、本来は信頼に足るものまで疑い始めてしまうのですとか、正しい情報まで疑い始めてしまうというようなスピルオーバーのエフェクトというものが指摘されています。これは結局バランスの問題だと思うのですが、どういうふうに効果だけを得てスピルオーバーを防ぐか、この辺りについて御知見があればお伺いしたいと思います。

2点目については、スライド10ページ目で、システム設計者の意図しない形で消費者に解

釈され得るというお話がありました。必ずしもシステム設計者が悪意を持ってバイアスを埋め込む場合もあるかもしれませんが、そうではない場合もたくさんあると思うのです。そういったときに、でもシステム側、エンジニアとしてはできるだけ誘導したくないと思っているし、考えを促したいと思って設計するのだけれども、ユーザー側がそうではない挙動をした場合に対して、それに備えてエンジニア側を支援するような策としてはどのようなものが考えられるか。何か御知見があれば教えていただきたいと思います。

○矢谷教授 ありがとうございます。

まずスピルオーバーに関しては、あまりやり過ぎると疑心暗鬼みたいなものが起きてしまうみたいなことなのだろうなとは思います。こういうことに限らず、さっきまとめていたスライドにもありましたけれども、AIのことを信じ切ってしまうと、人を信じないみたいな、AIに事前に相談してこう言われて、言ってみたら全然違うと言われて、AIと言っていること違いますみたいなことというのが医療の現場とかでも結構あって、私はこんなに重篤な病気なので風邪なわけないみたいなことを信じ切ってしまうと、言うことを聞いてくれないみたいなことを相談されたりすることとかは結構あったりします。

なので、もちろんそういう可能性もあるとは思いますが、ただ、そこはリスクと得られる効用とをどれぐらいバランスを取ってやるかというところを見極めながらやらざるを得ないのかなというところはあって、それは予防接種も多分同じで、場合によっては副作用とか、そういうものが出る方もいらっしゃるわけです。

もちろんそれは最小化するようにプログラムというか、そういうトレーニングシステムというか、そういうものを設計するというのが一つ大事だとは思いますが、それをしつつ、この辺でもありましたが、ある種の人間AI共生特区みたいな、ここでは実験的なこともいっぱいやりましょうみたいなものを作って、そこでいろいろやってみてリスクを少し取りながら、ただ、それを実際に広げていくときにはリスクをできる限り下げる形でやるみたいな、そういった最初はスモールスケールで回して、どこまでのトレーニングでどういうことが起きるのかみたいなことを見ながらやるということも大事なのかなと思います。

もう一つ、エンジニアの視点からというのはまさにおっしゃるとおりで、それはここで私が書きたかった、言いたかったことの一つなのですけれども、萎縮し過ぎ、リスクが起きてしまうからやりませんとなってしまうとあまり意味がないので、そういうことが起きたときに、起きてしまったのだけれども、悪意がなければそこに対してペナルティーは加えないけれども、ちゃんと対策はしましょうということをやるということで、エンジニアだったり、その会社の責任みたいなものを過度に追求したりとか、さらし上げるみたいなことがないようにしてあげるということは大事なのかなと思います。

もちろん悪意があってやっている人は駄目ですけども、そうではないのであれば、そこに関しては問わないというか、むしろ守ってあげるという形で、ただし、さっきのAIが暴走してみたいなことがありましたけれども、どういったときにそれが起きてしまったの

かという知見をためていくということが大事だと思うのです。いわゆる何か事故があったときに、例えばこういう不備があったから、こういう事故が起きたみたいなものがたまっていくことによって、社会全体のレジリエンスが上がっていくみたいなことになると思いますので、それをちゃんとためておける社会制度としてそういうものがあって、ある種の作る人にとってのセーフティーネットがちゃんとあるみたいな形にしてあげること、あと、そうなのですよということをちゃんと言うということが大事なのかなという気はします。

○田中委員　ありがとうございます。

○小塚座長　ありがとうございます。

加藤委員、お願いします。

○加藤委員　私は消費者団体から来ているので、ユーザー側の目線でお尋ねしたいのですが、けれども、対話型AIを使っていると、時々なのではけれども、選択肢が1つだったものが2つ出てくるときがあるのですが、この「選択肢」という観点から教えてほしいのですが、AIが迷っていると選択肢が出現するのですか。

○矢谷教授　そこら辺は作った人に聞いてくださいとしか言いようがないのですが、迷っているということもあるかもしれないですね。あるいは、その状況にもよると思います。あるいは、こっちのほうがいいですかみたいな、今、例えば応答の中で、ある種ユーザーが期待するような答えを出せなさそうだから、ちょっと違うこと、変化球を出しましょうみたいなことがあるのかもしれないかなと思います。ただ、私も具体的になぜそうになっているかと言われると分からないし、答えようがない。

○加藤委員　流れの中で、AIがすごく自信満々に回答してきたものが、急にあなたが選択してくださいみたいな形で2つ選択肢が出てきて、じゃあ私はこっちを選ぼうと思って選ぶと、次また会話がつながっていくのですけれども、この選択肢が出てきたときに一種の安心感のようなもの得られまして、自分自身で「選んでいる」という気持ちになるのです。

先ほど複数のAIエージェントを並列させて、それを対話型でやっていくというふうに、それはまだ世の中にはないのですよね。

○矢谷教授　あまりないと思います。私たちは研究でそれをいろいろなところで展開したりとかするのですが、一般にはあまりないとは思いますが。

○加藤委員　そうなのですね。1回のサブスクの契約で、例えば3つのAIエージェントが同時契約できて、それぞれが対立し合う構造になっていて、その議論の中で自分が選択をしていくなんていうことができる時代がいつか来るかもしれないということですか。

○矢谷教授　そうですね。そういうのはシステム上の設計としては十分あり得ると思います。今でも例えば一つのこういう問合せをしたときに、幾つかのモデルに投げて、それを取ってきて、自動的に例えば多数決を取ってこうでしたとかいうのをやったりとか、例えばこういう問いだったらこのモデルがいいよねみたいなことを知っていて、このモデルに飛びましたみたいなことはあるのですけれども、我々はどちらかというとそれをユーザーにむしろダイレクトにしゃべってもらって、私はこう思います、私はこう思います、私は

それはちょっと嫌だとか言ったりとかというのをやると、そうかもしれないみたいなことがあったりとかいうのはうまく生み出せるのではないかなと思っていて、ダイレクトにいわゆるグループでディスカッションするみたいなものを心理支援だったりとか、あるいは論文を読んだりとかというときですね。

例えばそういうのも、今、物すごい量の論文があって、人とディスカッションする時間なんて全然全くないのですけれども、そうすると自分が勝手に解釈して、そのまま書くと全然違うことを書いているみたいなことが結構あったりするので、読みながら、こういう分野の立場の人からこれを見るところこういう面白さがあるのだよみたいなことを言われると、なるほどこの論文はこういうふうに解釈できるんだみたいなことを教えてもらうみたいなことがあったりとかするので、そういうのはうまく設計できるのではないかなと思います。

一方で、あまり数が多いと、いわゆるグループとかでもあまり人数が多いと、あっちからもこっちからも来てしまうので、そうするとまたそれはそれで情報過多になってしまうというのがあって、そこをどう設計するかというのがインターフェースとしてはまだ明確にはなっていないところはあるので、やるべきことはいっぱいあるのですけれども、私自身はそういう設計というのが大事なのではないかな。聞くタイミングで複数に聞いていすみたいな、それでこっちもいい、こっちもいいみたいなことが大事です。

あと、その中でもう一つ大事なものは、AIが同じような意見を持っていると、むしろそれは完全にフィルターバブルを作っているだけになってしまうので、エージェントAさんはこれがいいです、Bさんもこれがいいです、Cさんもこれがいいですだとそれになってしまいうので、完全にもうAしかないでしょうみたいな感じになってしまうとそれはよくないので、どうやってその多様性を保つのだみたいなことが結構重要になってきて、一つのモデルを使って例えばプロンプトだけを変えてエージェントの振る舞いを変えようとしても、AI同士でディスカッションさせていくとだんだん意見が似てくるのです。というのが何となく我々もやっていて分かってきて、同じようなことを言い始めるみたいなことがあって、その多様性はうまく保つのは結構実は難しいのではないかなみたいなことは何となく分かってきていて、それをどうやって、ここにもありますけれども、単に正しいかどうかではなくて、いかに違う意見を言っているかみたいなものをどうやって評価してあげて、それをシステムに組み込むかということというのは、そういう環境だとなおさらやらなければいけないというところはあるかなと思います。

○加藤委員　ありがとうございます。

消費者には選択する権利があると言われてるので、もちろんどのAIエージェントを使うかという選択肢も必要ですけれども、回答を選択できるようにとか、いろいろな議論の中で自分が選択できるようにしていくというのがすごく、そういうのがあったらいいなと感想ですけれども思いました。

ありがとうございます。

○小塚座長　坂下委員、どうですか。

○坂下委員 御発表ありがとうございました。

資料を事前に拝見していて、7ページ目に出てくるアンダーラインが引っ張ってある3行があるのですけれども、これは消費者に好かれるAIと消費者のためになるAIは違うよというメッセージを出しているのだと思いました。

その上で、促考するAIが出てきます。促考するAIというのを社会実装しようとした場合にどうなるかということを考えると、通常のAIは答えを出すものですよね。推奨AIというのは最適解を出すもの、促考するAIというのは多分複数案を出したり、反対意見を述べたり、昭和の時代に同じ町内にいた人のようなものだと思います。

それに対し、促考と反するAIというのは、自分の判断の理由を問いたすようなものを目指しているのだと思うのです。これを社会実装しようと思った場合というのは、判断の質とか、主体性とか、満足度とか、信頼度とか、意思決定の度合い、そういうものを長時間にわたって、長期にわたって測っていかないと基準ができないと思うのです。そこで矢谷教授は27ページ目に全体を評価する枠組みとして、長期的な測定が要るのだということをおっしゃっているのだと思うのです。ただ、これはやらないといけないと思うのですけれども、こういう取組というのは今、検討されているものとか予定されているものは何かあるのでしょうか。

○矢谷教授 私たちの研究室のレベルでユーザー実験をしたりとか、そういうものというのはやってみようと思うところとか、例えば先ほど申し上げたような複数のエージェントとしゃべるというのは、具体的には動機づけ面接という、いわゆる健康診断とか引っかけると、ちょっと痩せましようみたいな話をさせられますよね。あなたはどうしますかみたいなときに、あなたはビールを飲むのをやめてくださいとかというのではなくて、一緒にどうしましょうかということを考えるのです。それで自分なりにこういうことやってみましょうということを発話させるというコミュニケーションスタイルがあって、それは動機づけ面接というものであるのですけれども、それをグループ形式でやるというものがあります。通常は1対1のカウンセリングみたいな感じでやるのですけれども、それをグループでやるということがあります。

それは対面では実はあって、どこであるかということ、薬物依存症の治療プログラムであるのです。そこではいろいろな方々、依存症を持っている方々が来て、今週はこんなふうに頑張りましたとか、今週はどうしても駄目だったとか、いろいろな話をしながらお互いに学ぶみたいなことがあるのですが、当然対面でやるのはなかなか難しいということで、それをオンラインで置き換えた形でやってみるのはどうだろうというのを我々が考えて、そこにAIエージェントでいろいろな役割というか、ちょっとキャラクターが違うものを入れながらということをやったりという感じです。

その実験では、2か月実際に使ってもらったりとかということをしてやったり、ユーザーは薬物依存症の方ではなくて、同じようになかなか人にしゃべりにくいという話で、月経前症候群を抱えているユーザーさんに対して行いました。これも女性が女性に対してもな

かなか月経の話はしにくいとか、女性が男性に対してはなおさらしにくいみたいなのところがあって、どうしても一人で抱え込んでしまうというのがあるので、AIと1対1でしゃべるというのと、グループでしゃべるというのとを比較するみたいなことをデータを集めてやってみたという感じです。

それをやっていくと、実際に例えばポジティブな言葉が増えるとか、あるいはソーシャルラーニングみたいな、AIが言っていることは私もちょっとやってみようみたいなものが増えるとか、出てくるみたいなことがあったりとか、実験の参加者の方は、自分以外は全部AIエージェントというのは知っているのですけれども、そこに居場所を感じるみたいなことというのが実は生まれていて、そういった効果があるということが見えてきてはいます。

なので、もちろん今は研究レベルなので、予算とかいろいろなものの都合で、そんなに1年とか2年とかできるわけではないのですが、数か月の単位でやってみても、何かしら見えてくるものはあるというところはあるので、それをもっと大規模にまたできれば面白いだろうなという気がします。それは多分、世界のどこを見てもやっていないので、日本がリードできたらすごく面白いのだろうなという気はします。

○坂下委員 ありがとうございます。

疫学みたいな調査のようなものになると思うのです。私はBCIの研究もやっているのですが、BCIだとドーパミンとかセロトニンとかが出ている方は赤信号の分析とかにかけるではないですか。ああいうものと連携すれば、介入なのか誘導なのかという区分が作れるような気がちょっとしていて、これは学際を超えて研究するべきものではないかなと思いました。どうもありがとうございます。

○矢谷教授 ありがとうございます。

MITなんかの研究者も今、EEGを使って実際にAIを使っているときとそうではないときとでどんなふうにシグナルが変わるかみたいなことがあって、今は予備的な実験的な扱いではあるのですがやっているものがあって、我々も似たような、脳の活動みたいなことが見えたら面白いねみたいなことというのは議論させていただいているので、ぜひいろいろ意見交換させていただければと思いますので、よろしくお願いします。

○小塚座長 ありがとうございます。

矢谷教授、今日は本当に示唆のたくさんあるお話をいただきまして、質疑も10分も超過してしまいましたけれども、どうもありがとうございました。

《3. 閉会》

○小塚座長 では、本日の議事はここまでとさせていただきたいと思います。

事務局から連絡事項がありましたらお願いいたします。

○三谷参事官 事務局でございます。

本日は、長時間にわたりまして、誠にありがとうございました。

次回の本専門調査会につきましては、確定次第、事務局より御連絡をさせていただきたいと思います。

事務局からは以上です。

○小塚座長 思わず興が乗ってしまいまして、時間を延長して大変失礼いたしました。

これにて閉会とさせていただきます。

皆様、お忙しいところありがとうございました。

以 上