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11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11回）
議事次第

1．日時　令和8年10月8日（木）10:00～11:23

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、大塚委員、加藤委員、坂下委員、田中委員、馬籠委員

【テレビ会議】

丸山座長代理、野村委員

（オブザーバー）

【テレビ会議】

鹿野委員長、大澤委員、柿沼委員、善如委員

（説明者）

【会議室】

秋山　満昭　ＮＴＴ株式会社 社会情報研究所 上席特別研究員

（事務局）

小林事務局長、小堀審議官、三谷参事官、江口企画官

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 消費者によるＡＩ技術の適切な利活用に向けて
有識者ヒアリング（秋山　満昭　ＮＴＴ株式会社 社会情報研究所
上席特別研究員）

（3）閉会

≪1．開会≫

○小塚座長　皆様、おはようございます。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」第11回の会合を開催いたします。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。

本日の出欠状況ですが、専門調査会委員のほうは、大塚委員、加藤委員、坂下委員、田中委員、馬籠委員、それから、私、小塚が会議室におります。

そして､丸山座長代理と野村委員が､テレビ会議システムで御出席くださっております。

消費者委員会の委員の皆様、オブザーバーということですが、鹿野委員長、大澤委員、柿沼委員、善如委員がテレビ会議システムで御出席くださっているということです。

岡崎委員と唐沢委員は、所用により御欠席と承っております。

それから、本日の議事（2）でプレゼンをいただきます、NTT株式会社社会情報研究所の秋山研究員ですけれども、10時半頃会議室にお越しいただけるという予定になっております。

それでは、いつものことですが、本日の進め方について、事務局から御説明をお願いいたします。

○江口企画官　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視聴いただき、報道関係者のみ会議室にて傍聴いただいております。

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

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。

以上です。

≪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　ありがとうございました。

本日の議事ですが、議事次第を御覧いただきますと（1）ということで、まず「本専門調査会の検討状況に関する報告」ということです。事務局で、本専門調査会でどういう議論状況になっているかということをおまとめいただいた資料がありますので、それを御報告いただきたいと思います。10分から20分の間で御報告いただければと思います。よろしくお願いします。

○福島補佐　事務局でございます。

お手元の資料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（案）」を御覧いただければと思います。

次のページをお願いします。

第10回専門調査会までの議論の概要を御報告させていただければと思います。

まず、1、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。

こちらにつきましては、第9回専門調査会で概要につきまして報告を済ませていただいております。

2、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・評価することが必要ではないか。

この点につきましても、第9回専門調査会で概要を報告させていただいております。

3、上記2のリスクへの対応として、環境整備を推進していくことが必要なのではないかという点につきまして、AI技術の利用によるリスクから消費者の利益を保護する側面と、AI技術を利用して消費者の利益の実現を図る側面の双方を推進するような、制度的対応・技術的対応を含む環境整備として、そもそもどのような在り方が適切かということについて議論をいただきました。

次のページをお願いいたします。

まず、前提として、AI技術の利用と消費者の意思決定プロセスの関係性について、主に以下の特徴が指摘できるという議論を行っていただきました。

まず、①につきまして、有用性と有害性、短期的な利点と長期的な悪影響等のように、消費者にとって利益の源泉となる要素が、同時にリスクの要因ともなり得る。このため、同一の技術であっても、システムの提供者や設計・利用の在り方により、消費者への利益とリスクのバランスの態様・程度は大きく異なり得ること。

また、②、消費者の利益の実現に資することが期待されるAIシステムであっても、その設計・運用によって、消費者の利益の実現と事業者の目的とが必ずしも一致しない状況、こちらは目標の不整合と呼ばせていただきますけれども、こちらが生じ得るため、継続的に細やかな調整を実施していくことが課題となること。

また、③、第9回本専門調査会報告のとおり、意思決定プロセスに係るリスクにつき、消費者が自ら不利益に気づき是正することは容易ではなく、消費者個人での自衛が困難といえること。

これにより、次のことがいえるかと存じます。多様な主体による、AIシステムの運用・設計の継続的な検証・調整をはじめとする、種々の手法を組み合わせた環境整備、すなわちAIガバナンスを通じて、消費者にとってのリスクの低減を図るとともに、利益の実現の促進を目指すことが重要であるということが議論されました。

次のページをお願いいたします。

必ずしもAIガバナンスという用語を使用しているわけではございませんけれども、既に消費者政策において、以下のような環境整備の在り方について提言がされているところでございますので、紹介させていただければと思います。

こちらは、令和7年、2025年7月の、消費者法制度のパラダイムシフトに関する専門調査会の報告書より引用、そして、まとめをさせていただいております。

同報告書では、AI技術を含むデジタル化の進展等により、消費者を取り巻く取引環境が大きく変化しており、従前の消費者法制度では、必要な対応ができなくなっているのではないかという問題意識が起点となっております。

そして、生身の人間、生活者たる自然人である消費者について、主に以下の点が摘示されております。

1点目は、内的及び外的要因の組み合わせと相互作用により、誰もが脆弱になり得ること。

2点目といたしまして、多様な脆弱性を有する消費者が取引に関わるための環境整備が重要であることです。

そして、これらの対応として、ハードロー的手法とソフトロー的手法等、種々の手法をコーディネートした実効性の高い規律の在り方が必要であることが提言されました。

右の図におきまして、まとめられておりますので御参照いただければと思います。

これに関連しまして、第8回、丸山座長代理より、消費者政策において、消費者に情報を与えた上で自己防衛を求める情報パラダイムからの転換が議論されるようになったことの御報告がございました。

次のページをお願いいたします。

意思決定プロセスへの影響に関する我が国の民事法制度についての議論をまとめたものでございます。

こちらについては、民法、消費者法の契約成立過程の規律の射程というところが議論されました。

これまで、我が国の民法・消費者法の契約に至る意思決定プロセスに係る規律は、基本的に、こちらの上の図、こちらは大塚委員から御提供されたものでございますけれども、②の選好の部分を射程外としてきたと解されるということ。

しかし第9回本調査会報告のとおり、AI技術について、個別の契約成立過程より前の、曖昧で潜在的な選好・問題認識・ニーズ等、以下「選好等」と呼ばせていただきますけれども、こちらの形成過程への影響力が増しているということ。さらに、選好等への影響については、契約の先行段階にとどまらず、契約に至らないものも含めた消費行動への将来的な影響を含む点にも留意すべきことが指摘されました。

このように、現行の民法・消費者法による契約成立過程の規律のみでは必ずしも十分に対応しきれない場面も想定されることから、こうした場面における適切な対応の在り方について議論を深める必要があるということが指摘されてまいりました。

次のページをお願いいたします。

そこで、第10回専門調査会までの議論の概要ですけれども、消費者利益の保護・実現を図るため、AIシステムの開発・提供事業者が、必ずしも特定の契約の場面が生じることを前提としない段階においても、事前のリスク評価や設計改善に取り組むことを促すAIガバナンスの在り方が求められるのではないかというところが議論されております。

これまで本専門調査会で議論されてきた事項を含め、問題を生じにくいAIシステムの在り方について、継続的な検討・評価・改善を図ることが重要であるといった旨が示唆されました。

次に、ケーススタディについて御紹介をさせていただきます。

想定事例として、対話型AIの助言を過度に信頼した結果、専門家への相談を行う機会を逸したようなケースというものを想定させていただいております。

こちらは、大塚委員より御提供いただきました事例に基づいて、事務局が作成したものでございます。

上記の事例につきまして、AIシステムの開発・設計段階において考慮すべき要素の例、いろいろあると思いますけれども、代表的な例というものを挙げさせていただいております。

まず、過度な信頼を防ぐ措置が取られていたか。当該AIの出力が専門家の意見に代替するものではないことが明らかにされていたか。

また、AIの出力内容自体を完全にコントロールすることは困難であるとしても、消費者の意思決定への過度な介入や過度な信頼・依存を防ぐため、UI等を通じた適切な予防措置を開発・設計段階で講じることは合理的に期待できたかといったようなことが考慮すべき要素として挙げられるかと思います。

この点、ガイドライン等におきまして考慮要素が整理・明確化されることは、法制度としての民法の不法行為法、第709条でございますけれども、こちらにおいても、注意義務の内容の検討や過失の判断に際し参照され得ると考えられるということが議論されました。

次のページをお願いいたします。

最後に、不法行為法につきまして、本専門調査会の議論に関わる範囲で、やや大まかな形ではございますけれども、御紹介させていただければと思います。

上の四角囲いを御覧ください。

民法上の不法行為法は包括的な規定であり、技術の発展に逐次対応が可能であることから、実務上、AI登場以前から新しい技術が実装された際に、事案に合わせて条文の解釈を行う柔軟な解決が図られてまいりました。

ただし、条文の解釈についての裁判所の裁量の余地が極めて大きく、予見可能性の確保が課題とされている点が指摘されているところでございます。

以下につきまして、資料を掲載させていただいております。こちらに掲載させていただいております資料は、全て大塚委員から御提供いただいたものでございます。

特に、右の四角囲いでございますけれども、AIに絡めて、AI等不法行為法の問題点につ

いて整理をいただいております。こちらは読み上げをさせていただきます。

まず、①、AI開発・提供・利用に対する規範的な評価として、過失、すなわち予見可能性を前提とした結果回避義務違反というものが想定される。

これについては、開発者・提供者・利用者がどのような義務を負うかの規範的評価が必要であるということ。

また、②といたしまして、過失・因果関係の証明につきましては、被害者が過失・因果関係を証明する必要があり、AIのブラックボックス性により、過失・因果関係の証明が困難であるということを御指摘いただいております。

また、最後になりますけれども、スライドの5ページ、図のところに戻っていただきまして、こちらの各段階について、大塚委員がまとめてくださっておりますけれども、どの段階でどのような影響があり得るかという実証的な見地からのさらなる細分化という点につきましては、今後の技術や科学的知見の発展を踏まえ、継続的に注視し、検討していく必要があると考えられることが議論をされてきたところでございます。

事務局の報告は以上になります。

○小塚座長　ありがとうございました。

当専門調査会の検討状況を事務局でおまとめいただいたわけですけれども、公開して、いわゆる平場でプレゼンをいただいたり、議論をしたことも、もちろんありますし、それに基づいて、さらに事務局と各委員の先生方との間でやり取りをさせていただいたものとか、それから、前からアナウンスしていますけれども、法律家としての整理をさせていただきたいということで、丸山座長代理、大塚委員と私でディスカッションをしたりしまして、そのような成果も反映させていただいております。

そういう意味で、この専門調査会は、文字どおりの専門の先生方にお集まりいただいて、言わば集合知として議論していますけれども、その集合知を1つのまとまりに落とし込んだものと御理解いただければと思います。

逆に申しますと、こういうまとめ方をするのであれば、こういうことも盛り込めるのではないかとか、これはニュアンスが違うのではないかとか、そのような御指摘もあろうかと思いますので、ぜひ、この場で御指摘、御発言をいただきましたらと思います。

どなたからでも結構ですが、いかがでしょうか。

坂下委員、お願いします。

○坂下委員　坂下です。ありがとうございました。

この資料を拝見していて、5ページ目の枠囲みの中で「個別の契約成立過程より前の」というのがありますね。これは、昔の広告からずっと起きていることだと私は思います。

パーソナライズ広告とか、あと、ケンブリッジ・アナリティカとか、これもパーソナライズされている話です。

問題は、AIの場合というのは、個人をリアルタイムで把握して、継続的にデータを対話で取って、そこで判定していくところが重いところだと思うので、その辺を皆さんの集合

知として集めていただいて、EUは消費者法ガイダンスの中で、たしか低所得者の方にリアルタイムの広告を出すときに、こういう配慮をしろというのは定めていると思うのですけれども、ああいうことも考えないといけないのではないかなと思いました。

以上です。

○小塚座長　ありがとうございます。

広告というものが、だんだん変質してきて、もともとは個別に働きかけるのと、広く一般の人に訴求する広告というのが全く別の世界だったのが、だんだん境界が曖昧になってきている。

それが、いよいよ顕著な問題になって、問題というのは、いいか悪いかは別にして顕著に取り上げるべき論点になってきたということだと思いますが、坂下委員、もう少し踏み込んで、そこに政策的対応をしている海外の事例もあるとか、そういうことにも留意すべきだという御趣旨でしたでしょうか。

○坂下委員　そうです。EUなどだと、ガイドラインをつくっていますので、そういうものも参考にしながら、ここをまとめていくと骨太になるのではないかなと思いました。

以上です。

○小塚座長　ありがとうございます。

ぜひ事務局としては受け止めていただいて、そういう趣旨も入れていただきたいと思います。

これからの時代、広告というものをどう考えていくかというのは、制度的にも、消費者政策的にも非常に大きな課題ですので、ここで頭出しするというのは非常にいいことだと思います。

オンラインから丸山座長代理の御発言の御希望があります。丸山座長代理、お願いします。

○丸山座長代理　ありがとうございます。

私のほうからは、今後整理をして文書化をしていくに当たって、少し精緻に整理をしていったほうがよいのではないかと思った部分について、発言をさせていただければと思います。

まず、5ページでしょうか。意思決定のプロセス図が載っているものと、それ以降の6ページ、7ページの不法行為やガバナンスの話が問題となる局面なのですけれども、まず、留意点としましては、この意思決定のプロセス、ずっと後ろのほうで扱っているガバナンスのケーススタディなどは、異なる場面というのを取り上げておりますので、この点はパワーポイントのほうにも、少しその点は言及しているのですけれども、誤解を与えないような整理というのは、必要ではないかと思いました。

どういうことかといいますと、6ページのほうのケーススタディでは、消費者が意思表示をしようとしている場面ではなくて、AIを利用している最中の場面でございまして、いわゆるAIの助言サービスの在り方というのが、専門家のサービスのような誤解を与えるとい

う問題なので、AI利用の継続中の問題と位置づけられると思うのですけれども、AIサービスの品質標準であるとか、法的には不法行為のみならず、サービスの提供契約のところの契約不適合の問題とか、あとは開発の場面では、PLの問題も発生していくような問題が取り上げられておりますので、そういった場面の違いがあるということは意識したほうがよいのではないかと思いました。

その一方で、このケーススタディに関しましては、その望ましいベストプラクティスというのを探求していくという問題と、不法行為責任を帰結する結果回避義務を認定できるかという問題は、基準がずれてくる可能性があると思います。

不法行為責任の活用ということについては、賛成はしているのですけれども、公的規制がない場面で709条の責任を肯定するには、結果回避義務があるのかということを回避コストなども考慮して、個別の事例で検証していく必要がございますので、文章化するに際しては､あまり広範に断定的な文章にならないよう､注意が必要ではないかと思いました。

最後に、情報パラダイムからの転換ということを言う場合には、従来は、新たなダークパターンとか、中毒性のある設計というのが登場してきたことを契機にしまして、衝動惹起とか、あとは、バイアス、認知のゆがみの利用であるとか、中毒という意味での依存とか、プライバシー侵害という、そういったものが消費者契約の場面で発生するようになったので、情報開示だけでは十分ではなくて、設計による対応とか、新たな被害を視野に入れた消費者法の設計が必要となるという、そういった議論がされてきました。

そのような文脈で、このパワーポイントの最後のほうにある、開発とか提供とか利用の各段階で、どういったガバナンス、設計というのが要求されるのかというのは、つながっていく重要なポイントになるのではないかと思うのですけれども、その一方でやはり、そこでの行為規範というのが、不法行為責任に直結するかどうかというのは、行為規範の設計次第というところもあると思いますので、ここも慎重に文章化していく必要があるのではないかと思いました。

文章の全体では、消費者の意思決定ということにフォーカスしているのですけれども、広い意味での消費者の選択とか、行動の自律性というものを問題としていると考えられます。

他方で､5ページのほうの意思表示の図というのは､民法で言うところの一定の法律効果を欲する意思の表示と定義されるところの意思表示というものを念頭に置いて、分析をしていただいているものであると思いますので、問題としている意思決定というものが、民法で言うところの意思表示を問題としているのか、選択ないし行動の自律性確保といったことを問題としているのかという点は、これは、後者であれば、消費者契約法4条が関わらないというのは当然のことにもなってきますので、その辺りの、何を取り上げているのかという点は、少し意識して整理をしていったほうがよいのではないかと思いました。

すみません、長くなりましたが、私からは以上です。

○小塚座長　ありがとうございました。

非常に緻密な御指摘をいただいたと思いますので、事務局は、この先、文章にしていくときに、いろいろ注意をしていただきたいと思います。

それで、できれば後で同じ法律家として大塚委員にもコメントをいただきたいところがあるのですが、私は、今、丸山座長代理がおっしゃっていただいたことで、2つ大事なことがあって、1つは、不法行為の事例ですね、不法行為、とりわけ特別法をつくるわけではなく、民法709条という議論をしますと、法律が、あまりお詳しくない先生方もいらっしゃるかもしれませんが、非常に抽象的に故意または過失があって、権利とか利益を侵害したら損害賠償責任だという非常に抽象的な書き方ですので、その具体的な事実がどういうものかということが非常に重要になってきます。

そういう意味で言うと、この3、4行で書ける話ではないのですね、それは間違いなくて、ただ、他方で、やはりこれは国民に対して、社会全体に対して、AI時代の消費者問題というものを考えていただく文書として、何ら具体的な状況がないと、それはそれで何だこれはともなりかねませんので、そういう意味で、分かりやすいシチュエーションを入れるということであって、そこに書かれた数行から、こういう場合に必ず責任が発生するということを言う趣旨では全くないということです。これは、疑いないと思いますけれども、大塚委員、何か補足あれば、お願いしたいのですが、事務局の書き方も気をつけていただきたいと思います。

他方で、意思決定プロセスの話は、これも丸山座長代理がおっしゃることは、そのとおりなのですが、大塚委員からいただいた図を非常に便利に我々は使ってしまっているのですけれども、契約をしようとするときの意思表示のプロセスというものの背後には、実は人間の意思決定の仕方というモデルみたいなものがあって、それ自体が1つの古典的近代の人間像なのですね。十分な能力を持っていて、情報さえあれば、きちんと判断ができるという人間像があって、それを法律に落とし込むと、契約をするという場面での意思表示の仕方ということになるということで、それに対して、今まで民法とか消費者契約法でどう対応してきたかといえばということを図示していただいているわけですけれども、その背後にある近代的人間像自体が脅かされているというのが、脅かされるという言い方はいいのかな、何か微妙なニュアンスがあるかもしれませんが、それ自体が生成AIの出現によって大きく変わってきているというのが、この専門調査会での認識で、その人間の判断の仕方というもの自体の展望、これ自体は、法律的に言う意思表示とか、契約成立という場面を超えた大きな射程の議論でもあって、そういう文脈で一応ここは書いている。その上で、今までの法律ではという話が出てきている。この2つを混同しないようにきちんと書き分けるようにという御指摘はそのとおりで、これは、また事務局に神経を使っていただきたいのですけれども、そういうことですので、全体として、このおまとめいただいたものが、今の議論の趣旨とか、法制度を誤解してつくっているわけではないというのが、私の理解ですけれども、大塚委員、補足があれば、お願いできますでしょうか。

○大塚委員　大塚です。

座長と座長代理のおっしゃることは、もっともなところが多いかなと思います。その上で、少し補足をさせていただきますと、まず、7ページ目の、AIと不法行為に関する論点についてですが、最後に、AIと不法行為の問題点ということで、私の資料を引用していただきましたけれども、最後の過失と因果関係の証明が困難であるといった点は、さらに続きがございまして、このAIの出力について、過失や因果関係を問うという場合には、やはりブラックボックス性がどうしてもつきまとってきて、証明困難という状況になるのですが、それ以外の部分を問題にするということは十分可能かと思います。それが6ページ目のケーススタディの下に書いてある部分でして、AIの出力の内容自体をコントロールするわけではなくて、その出力を見た利用者、消費者が過度な信頼を抱かないようにする、そのようにUI等を設計していく、この点については、特にブラックボックス性が効いてくる場面ではございませんので、その点について、何らか行為規範を設定する、この行為規範というのは、座長代理がおっしゃっていたように、望ましいものというよりは、法的に、ここまではやらなければいけないものという、そういった意味での行為規範ですが、そういった行為規範を定めることができれば、その違反について事業者の責任を問うていくということが、必ずしも困難なことではないように思います。

この後、文章としてまとめるとしても、この文章で、ここまでやらなければいけないという法的な拘束力を持つような文章はつくれませんので、あくまで、その後、裁判所等によって参照され得るという程度ではございますけれども、その際にも望ましい、やったほうがいいという規範なのか、そうではなくて、ここまでは、それはやらなければいけないねという規範なのかということを区別するというのは、重要な整理なのかなと思います。

もう一つ、5ページ目の意思決定プロセスに関する図についてです。あるいは、これと6ページ以降の不法行為との関係についてです。

私がいただいた課題が、初めの部分は、意思表示に関する議論をまとめてくださいというもので、次の発表では、不法行為についてまとめてくださいという課題でしたので、別のトピックとして御報告したわけですが、おっしゃるとおり、両方を全体の体系の中でどのように位置づけていくかというのは、結構重要なポイントなのかなと思います。

この意思決定プロセスの図については、まさに意思表示が生じる、特に契約が成立するという場合の問題点を指摘したものでして、それ以外のパターンについては、不法行為あるいはAI事業者との間の契約の話に行き着くのかなと思います。

ただ、意思決定プロセスの図のように、意思表示がされた場合でも、この意思表示の有効性を争うというのは、その意思表示の相手方との関係の問題ですが、そうではなくて、AIを提供している人に対して責任を問う場合には、不法行為なり、AI事業者との間の契約なりというものの問題が生じてきますので、この意思表示の問題と不法行為の問題が、1つの事例で同時に出てくるということは十分あり得るかなと思います。

そういった意味で、やや複層的な関係に立ちますので、この辺り整理ができると、なおよい資料に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。

責任があるとかという言い方をすると、事業者の方々は、非常に緊張されますので、そこはあくまでも慎重に、しかし、同時に消費者委員会として消費者の利益ということを実現していくということが、政策的なミッションですので、そういう意味で言うと、法的な責任の有無に限らず、どうしていくことが望ましいかという視点は、やはり大事だと思いますので、その辺り誤解のないように書いていただければと思います。

そのほか、委員の皆様から、田中委員、お願いします。

○田中委員　丸山座長代理から御指摘のあった、より精緻にしていくという方向性について、1つ観点を補足させていただけたらと思います。

スライドの6ページ目で、今、大塚委員が御説明いただいた、UI等を通じた適切な予防措置を開発設計段階で講ずることは、合理的に期待できたかのところについて、先ほどやらなければいけないということと、やったほうがいいことは分けたほうがいいということでした。

ここをもう少し精緻にすることができるかなと思っていまして、やらなければならないの中でも、必要条件なのか十分条件なのかは大分違うと思うのです。最低限として、ここはやったほうがいいけれども、それをすれば、予防が完全にできるかという話は、結構グラデーションがあると思うのですね、そこは書き分けるというスタンスが必要かなと思います。

それから、現時点での開発者側も、やりたいけれども、やることができないということもあると思うのですね。それは、技術が発展していく上で、どうしても生じるフェーズですので、そこも分けたほうがいいかなと思っていまして、今やったほうがいいということと、今できること、やったほうがいいけれども、今、何ができていないかということも明記するということは非常に重要だと思っていまして、これはアメリカのNISTのAIリスクマネジメントフレームワークでも、測定しているものと、測定できていないものを記録することが非常に重要としていまして、そういったところも、もう少しグラデーションを分けて説明することができるかなと思いました。

以上です。

○小塚座長　ありがとうございました。

御指摘いただいたことは非常に重要だと思いますので、ぜひこれは明示的な形で、今後、言語化していただけますでしょうか、事務局にはお願いいたします。

これも法律家的に言うと、それは、予見可能性とか、結果回避可能性とか、そういう言葉に吸収されていくわけですけれども、そのようにしてしまう以前に、やはり開発している方あるいはそのサービスを使う、使うというのは、事業者の立場で使って市民向けに提供していく皆さんのために、分かりやすい言葉で言語化していくというのは大事なことだと思いますので、非常に重要な点です。ありがとうございました。

それでは、このまとめ方について、ほかにもまだ御意見等ありましたら、適宜事務局にお知らせいただければと思います。議事1につきましては、このあたりとさせていただければと思います。ありがとうございました。

≪2.　② 消費者によるＡＩ技術の適切な利活用に向けて
有識者ヒアリング（秋山　満昭　　ＮＴＴ株式会社 社会情報研究所 上席特別研究員≫

○小塚座長　さて、議事2に進みたいと思います。「消費者によるAI技術の適切な利活用に向けて」ということです。

本日は、NTT株式会社社会情報研究所の秋山満昭上席特別研究員にお越しいただきました。秋山様は、先ほど会議室に入られました。お忙しい中、ありがとうございます。

秋山研究員は、サイバーセキュリティを専門とされており、人間の認知に関わる脅威を扱うコグニティブセキュリティ等の領域において研究をしておられます。

本日は「対話型生成AIと消費者の意思決定をめぐる最近の研究動向」というお話をいただけるということになりました。

これから大体20分程度お話をいただきまして、その後、私どもで、質疑応答、意見交換と進めたいと思います。

それでは、秋山様、どうぞよろしくお願いいたします。

○秋山上席特別研究員　御紹介ありがとうございます。NTT株式会社の秋山満昭と申します。

本日は、このようなタイトルで、私の研究に関係ある部分、私が把握している範囲ではありますが、最近の研究動向について御紹介させていただきます。そして、幾つかの論点を最後にまとめさせていただけたらと思っております。

なお、本日の発表は、私の所属組織の公式な見解を示すわけでありませんので御注意ください。

先ほども説明させていただきましたように、既に公開されている研究内容を中心に、本日は、お話をさせていただきます。

メインテーマが、対話型生成AIに関わる消費者の意思決定において、自律性というものを共通のキーワードとして幾つかトピックを挙げさせていただきます。大きく分けて5つのトピックを挙げさせていただきます。

それでは、最初のトピックになります。

皆さんもダークパターンは御存じだと思います。これまでのダークパターンというのは、左上の図にありますように、画面上で人が注意を向けやすいようなデザインにして、おの

ずとその方向性に意思決定をゆがめると。これが、サービスの事業者にとってメリットがあるようなものになっているということで、こういう視覚的、構造的なデザインによる操作だとか欺瞞というものをダークパターンと呼ぶとされております。

ただ、最近、LLMが急速に発展しておりまして、こういうダークパターンをLLMの対話の中で実現できるのではないかという研究がされております。LLMと人が自然な対話をしている中で、人の信念をゆがめるような情報をインプットするだとか、ユーザーの個人情報を必要以上に抽出するような質問をLLMが投げかけると、こういうものがどこまでできるのかという研究が、今年発表されております。

このようなLLMによるダークパターンは、普通のLLMが人に対して支援的な対話をするのと見分けがかなりつきにくいというところで、技術的な難しさがあります。

この研究をやっている論文では、課題として、通常の支援的な会話との区別が難しくて、ユーザーの認識だけに依存した対策には限界があるだろうと指摘しております。

そして、LLMダークパターンというものが、実験的には実現できることが実証できたのですが、これを検出したり、評価する方法だとか、基準というのが、現時点で十分に確立されていないことも問題として指摘しております。

そして、こういう問題が生じたときに、誰の責任になるのかということも、この研究の中でユーザー調査を行ったところ、これはサービス企業の問題だと考える人もいれば、LLMの開発者の問題、もしくはユーザーの自己責任だと考える人もいたということで、この責任の所在が曖昧だということが指摘されておりました。

ですので、これをまとめると、ある種のアカウンタビリティー、責任主体だとか、責任の範囲とガバナンス、それを管理して統制する仕組みが、課題として浮き彫りになったということが示されております。

論文の中では、幾つか詳細なLLMダークパターンのカテゴリーと、その中で使われるテクニック、具体的なLLMが人に働きかける方法について説明されております。本日は、詳細を割愛させていただきます。

2番目、このようなダークパターンがある中で、これは法的にどのように解釈されるかというところが、実はそれほど明確ではなかったということです。その問題を解決する取組も行われているということで紹介させてください。

有名な法規範としては、EUのDigital Services Act、DSAの25条ですかね、オンラインプラットフォームのインターフェースについて、利用者を欺く・操作する、または自由かつ十分な情報に基づく意思決定の能力を、実質的にゆがめるだとか、損なうような設計というものをここでは禁止しております。これは、利用者の自律的な意思決定に関わる規定だと捉えられます。

条文では、このように書かれているのですけれども、これを技術的にどう捉えるかというのは、実は非常に難しくて、この問題に取り組んでいる研究者がいまして、ここで言う自律性の侵害というのはどういうものかというのを、さらに解釈を深めることを取り組ん

だ研究があります。そこでは、これは具体的にはデセプション、欺くこと、誤解を招く情報だとか、隠蔽をすること、そして、マニピュレーション、操作をすること、心理的バイアスを利用すること、そして、ディストーションおよびインペアメント、ゆがめて損なうこと、つまり、これは選択の余地を狭めるような設計をするということ、このように分解して解釈されると説明されております。

そして、それと並行して、ダークパターンについて、個別の事例が大量に出現している状況において、これを体系化する研究も出てきております。抽象度の高い概念から具体的なパターンへ階層的に体系化するような研究が出てきております。

この2つを束ねることによって、ダークパターンと、この自律性の侵害がどのようにマッピングされるのかということを分析した研究が、今年発表されました。

具体的には、DSAの3つの自律性侵害の種類に、ダークパターンをマッピングするのですけれども、そのときに、自律性侵害の判断に関わる8つのデザイン要因というのを特定したということが大きな貢献になっております。

このように、法的な理念というものと、具体的な技術、ダークパターンの技術的な実現方法というのをマッピングさせたということは、ダークパターンの規制だとか、法規制の設計指針の検討に資する知見になっていると感じました。

こちらが具体的なマッピングになります。自律性侵害の種類は3種類ありました。欺くこと、操作すること、ゆがめ・損なうこと、これと、自律性侵害の判断に関わるデザイン要因、これが、いろいろ分析する中で8種類あったとされていて、これをマッピングしたのが、こちらの表になります。

これで何がうれしいかと言いますと、例えば、選択肢の隠蔽をするようなダークパターンがあったときに、これは、選択空間において、ゆがめて損なう行為であると、ということは、これは自律性の侵害に当たると解釈できます。

同様に、うその情報を提示するようなダークパターンがあったときに、これは情報空間において欺くことになります。ということは、これは、自律性の侵害であろうと解釈できるます。このように使えるツールとして、この成果を活用できると考えられます。

この論文の中では、もう少し詳しく、具体的なダークパターンと自律性侵害の種類だとか、デザインの要因へのマッピングをしている詳細な情報も提示されております。本日は、割愛させていただきます。

これまでは、そのLLMが人間の認知に対して直接的に害を及ぼすようなものとしてダークパターンを説明してきたのですけれども、それだけではなくて、逆に、AIが人間の認知をサポートするようなものとしても、今後使われ得るという話をしたいと思います。その中でも、まだまだ問題が生じ得るだろうということを説明させていただきます。

近年、特に偽誤情報だとかが世の中に蔓延し始めた、そういう背景を受けまして、AIと人の対話によって人間の判断支援をするという研究が活性化しております。

偽誤情報を見分けるときに、AIが何かしらの参考情報だったり、見分けるポイントとい

うのを教えてくれると。それによって、このAIとの対話中に見分けるパフォーマンスが向上するという研究が幾つも発表されております。

これ自体は非常に望ましいことではあるのですけれども、一方で、AIとの対話による、この判断支援の効果というものは持続するのかという疑問を持つ研究者もおりまして、その調査がされております。

具体的には、AIの支援を一定期間行います。その後にAI支援を除去すると何が起きるか。その実験をやったときに、効果は、実は持続せずに、長期間AIを取り上げると、どんどんパフォーマンスが下がっていくということが示されました。

つまり、AI支援下において、人間の判断力が改善したということは、人間自身の判断能力が向上したということを意味しません。これは別の話になってくるということで、この研究では、人間自身が推論する、その過程を支援するようなAI設計上の論点があるだろうと主張しております。

ということで、これらの研究に基づくと、AIによる支援だと、支援下においての人間の判断の質が注目されがちなのですけれども、それだけではなくて、支援を離れた後の人間の判断能力、ここまで考慮する必要があるのではないかと示唆されます。

そして、本人にとって望ましい選択というのは何なのだろうかということを考える必要があります。

これは、古くから主に行動科学の分野において、いろいろな議論があるということを紹介させていただきます。

いろいろな考え方があるのですけれども、1つは、パターナリズムというものです。父権主義的なものなのですけれども、これは、本人の利益を守るために介入するというもので、これは、選択の自由を制限してでも何か助けてあげようという、そういうものになります。

それと全く逆で、リバタリアンという考え方もあります。自由主義的と日本語で訳されますけれども、これは、強制だとか命令をせずに個人の選択の自由を尊重するものになります。ですので、自由に選択していいのですけれども、自己責任ですよという、こういう真逆の考え方がある中で、行動科学の分野では、この中間の部分、リバタリアン・パターナリズムという考え方が広まっているというか、一般的になっております。

これは、自由な選択肢を残したまま、より本人の利益につながる選択をしやすいように、仕組みだとか環境を整えるというものです。ナッジという技術がありますが、これの基礎理念が、これに当たります。ナッジを提唱したキャス・サンスティーンが、このようなことを言っているということです。

そのナッジの理論を提唱したキャス・サンスティーンらが、さらに言っているのは、本人にとって何が望ましいか、そういうことを助けてあげたいのですけれども、これを直接把握するのは、実際は難しいと。ですので、そういう場合に考えられる幾つかのリバタリアン・パターナリズム的なアプローチがあるということも説明されております。

それが、3つぐらいありまして、いろいろナッジとしてよい方向に行動できるように、何

か手助けをしてあげるのですけれども、仮に明示的に全部の選択肢を出して聞いたときに、多くの人が選ぶだろう方法というのをナッジとして仕向けてあげるような、そういうものを採用しなさいと言っています。

もしくは、本人自身が明示的に選択する方法を採用すべきではないか。そして、本人が選択した後にオプトアウトする人が最も少なくなる、そういうものに対して、行動を仕向けるべきだと思っております。

こういう行動科学的な基本的な理念に基づいて、本人にとって何が望ましいかというのを考えて、技術だとか法制度を設計していく必要があるのではないかと考えられます。

最後に、法規制と研究・技術開発の関係について説明させていただきます。

GDPRが採択された前後のプライバシー研究が非常に大きなインパクトがあったということで、事例を紹介させてください。

このGDPRには、透明性、具体的には、データ処理を本人が理解できることですけれども、これと介入の可能性、本人がデータ処理に何らかの介入ができる、こういうデータ処理はやめてほしいだとか、そういうことですね。

そういうものが、GDPRの理念の中に入っているわけですけれども、このようなGDPRが採択されて運用開始された前後において、このプライバシー系の国際会議において、研究動向がどう変化したかというものを調査した報告があります。

2016年から2018年前後でGDPRが採択されて運用を開始されたのですけれども、それ以前は、このような透明性だとか、介入可能性の議論というのはほとんどなかったのですけれども、2019年以降ですと、こういうことに言及したものが、発表される論文の10％を超える水準になっているということで、非常にこの法制度が、セキュリティーとかプライバシーの技術的な後押しをしていると考えられます。

さらに、最近の人の意思決定に影響を与えるような研究に関して、幾つかの提言がされております。こちらは、私の今年発表した研究でもあるのですけれども、人に意思決定を与えるようなものに対して、設計として何を取り入れるべきかということを考えております。

GDPRで言われていたような透明性だとか、介入可能性だけではなくて、それ以外にも第三者による検証可能性だとかも重要ではないかと、この論文では言及されております。

つまり、人の意思決定に影響を与えるような、何かデザインだったり、サービスを実施するときは、透明性、つまり働きかけをユーザーが認知できるべき。そして、介入拒否可能性ですね、働きかけを、変更、拒否できるべき。そして、第三者による検証可能性、第三者が働きかけを検証できるべき、こういうものを設計の段階で入れ込むべきではないかという議論がされております。

これらの研究に基づくと、法規制によって保護するべき価値だとか、要求というのが具体化されることによって、それらを、測定、検証、実現する技術研究が促される可能性があるのではないかと示唆されます。

論点を最後に整理させていただきます。

対話型生成AIにおいても、消費者の意思決定に影響する設計上の問題が、実証的な研究で示されております。

そして、消費者の意思決定の影響というものを、自律的な意思決定の侵害という観点から整理して、具体的な設計と法的概念を接続するような試みも進められております。

一方で、意思決定を支援する設計を考える上で、本人にとって何が望ましいかというものをどう捉えるかというものも、これまでの行動科学研究上の様々な議論があります。

そして、AIによる意思決定の支援については、その場の効果だけではなくて、AIから離れたときの人間自身の判断能力への長期的な影響というのを検討する、そういう研究も現れております。

最後に、法的要請が具体化されることによって、それを測定、検証、実現するための研究だとか、技術開発が促される可能性があるということも指摘されております。

ということで、これらをまとめると、対話型生成AIと消費者問題を考える際には、情報の正確性だとか安全性だけではなく、こういうものは、以前から指摘されていたことですけれども、それだけではなくて、AIとのインタラクションが消費者の選択・自律性・判断能力に与える影響というものを捉えて、それを保護するための要件だとか、具体的な設計・技術上の課題につなげていくという視点が重要になると考えております。

最後のページは、今回の発表で取り上げた論文の詳細情報になります。

私からの発表は以上になります。

○小塚座長　ありがとうございました。

まさに最新の研究動向について御紹介いただきました。大変参考になるお話でございました。

それでは、委員の皆様、また、御質問、御発言、御意見等ありましたらお願いいたします。

馬籠委員が最初ですかね、では、馬籠委員、お願いします。

○馬籠委員　ありがとうございました。非常に勉強になりました。

私、電通デジタルというところで広告代理事業者として働かせていただいているのですけれども、その中で、正しいコミュニケーションデザインを消費者にどう設計していくのか、みたいなところは、企業側でも問題になってくるかなと、お聞きしながら思っています。例えば、正しいコミュニケーションデザインをするときに、こう使わないべきであるという具体的なチェックリストとか、何かしないほうがいいねといったことがあるのか、ないのかというところを、お聞きしたいなと思っていたところです。

もうひとつ、例えばLLMのダークパターンによって、消費者が、被害を被ったのが分かるのか、分からないのかというところもあると思うのですけれども、その後、例えば、それを行った企業に悪影響があるのでしょうか。要は、このようにだまされたっぽいぞということが分かったときに、非常に嫌になるとか、それとも、もともとそのブランドにはあま

り影響がなさそうであるとか、そういったものが、もし、あるとしたら教えていただきたくというところでございます。よろしくお願いいたします。
○秋山上席特別研究員　ありがとうございます。
現時点で、設計上こうすべきではないというものが何か確立されているというわけではないかなと思っております。
1つヒントになるのが、2番目に説明させていただきました、このダークパターンと法規制における理念とのマッピング結果になります。例えば、情報空間上でどのような制御をするのかとか、選択空間上でどのような制御をするのかによって、それがDSA的な自律性の侵害にどう関係するのかというところから、こういうデザインはやめておくべきだということが参考になるのではないかと考えられます。
そして、LLMにおいてダークパターンを行ったときに、これは非常に分かりにくいと思っております。この研究でも指摘されているのですけれども、LLMが支援的に何かおせっかいかもしれないのですけれども、いろいろあの手この手で何かアドバイスをしてくれることに対して、それがダークパターンと判断できるような行為なのかという、その線引きというのが、現状、全くそういう基準がないところが、技術的には問題だと思っておりますので、まずは、そういう判断する技術をつくることも重要だと思いますし、このページでのダークパターンのマッピングというのは、従来のダークパターンに関するマッピングなのですけれども、LLM上でのダークパターンが表出するような現象に対して、こういうマップをつくるということも、まだ、技術的にも進められなければいけないと思っております。
そして、LLMでダークパターン的なことをやったときに、ユーザーがその行為に気づいたときに何が起きるかということに関しては、まだまだあまり研究はされていないのですけれども、一部私の研究でもありまして、うそをついて推薦を行う研究をやるのですけれども、あの手この手で人間の認知を悪用することによって人の行動というのは制御できてしまうという結果が得られました。その後に我々は、実はあなたの行動を制御するために、こんなことをやっていたのですよとネタばらしをするのですね。そうしたときにユーザーがどう考えるかということも聞きました。そうすると、ものすごく反感がありまして、当然、うそをついて行動を左右されることは、ほとんどの人が嫌だと言うのですけれども、あなたのためを思って、うそだけれども、こういう行動をさせてあげたのですよ、これについてはどう思いますかと聞いたら、実は過半数の人が、それでも嫌だと、うそをつかれるのも、何か気分はよくないと答えていました。ですので、ひとたびユーザーに知られてしまうと、非常にネガティブな印象を持たれるということが分かったので、サービス開発をするときにも、そういうリスクを考慮した上で、ばれなければいいということではなくて、ばれたときにブランドの毀損を含めていろいろダメージがあるので、そういうことを考慮した上で、デザインすべきではないかと、研究結果がそういうことを示唆していると思っております。
○馬籠委員　ありがとうございます。非常に勉強になりました。

○小塚座長　ありがとうございます。

その話は、まさに消費者の自律ということで、結果がいいかどうかではなくて、自分で決めたというところが消費者の評価ポイントの1つになっているという示唆のような気もしますが、私のつぶやきはともかくとして、オンラインで野村委員から御質問があるそうです。野村委員、お願いします。

○野村委員　どうも貴重な御研究の調査をありがとうございました。

今回の御発表の中では、LLMダークパターンの技術的な特性からいろいろと分析をされているという研究が主だと思われるのですが、逆にLLMダークパターンにはまった人間側に認知行動のゆがみとか、その結果、QOLの損壊とか、そういう人間工学的な研究というのはなされているのでしょうか。

○秋山上席特別研究員　現段階では、LLMにおいてダークパターンというものが、どの程度実現できるかというところが検証されたという段階であって、今後、それを使ったユーザーへの評価、ユーザーがLLMダークパターンに遭遇したときにどういう判断をして、認知にどう影響を与えて、wellbeing的にもどういう影響があったのかということを評価するような研究がなされていくと考えられます。

○野村委員　ありがとうございました。

○小塚座長　では、田中委員、お願いします。

○田中委員　田中です。御発表ありがとうございました。1つコメントと、1つ質問をさせていただきます。

1つは、スライド10ページ目のナッジの話ですけれども、こちらは、いろいろなアプローチの中の1つとして、こういった研究があるということで御紹介いただいたと理解しました。

ここで挙げられているのは、もともとたしかセイラーとサンスティーンたちの話だと思うのですけれども、こういった何が望ましいかの1つの幾つかの基準があると、ガイドラインの評価項目とかに入れることができて、そういった指針として役に立つと。ただ、オプトアウトする人が最も少なくなる方法が万能かということを言っているわけではなくて、オプトアウトすることがいい場合というのも幾つか条件があるということも言われていたように思います。例えば、変更する代案が理解されているとか、変更するときのコストが少ないとか、いろいろな条件があって成立する1つの方法として、こういったアプローチがあるという御紹介のページだと理解しました。

質問なのですけれども、スライドの4ページ目で最新のLLMダークパターンの研究を御紹介いただきました。これは非常に重要な研究で、こういった研究が今後も必要になってくると思っています。

著者たちは、たしかカーネギーメロンの研究チームだったと思いますけれども、英語環境で、結構、三十数名ぐらいのサンプルを使った実証研究になっていたかと思います。

こういった研究結果が、それで、ダークパターンをすることが発生し得ると、それによ

って誘導することができるということを示した、ほぼ最初の研究になると思うのですけれども、こういったものが日本での情報環境で、どれぐらいユーザーがダークパターンで影響を受けるかということに対する示唆をどれぐらい持つのかということと、恐らく日本では、ほとんど研究がされていないのではないかと思うのですけれども、海外のこういった研究のみに依存して、日本の情報環境での実証研究がされていないということはどういったリスクがあるのかということについて、お考えをお聞かせいただけたらと思います。

○秋山上席特別研究員　ありがとうございます。

このような先進的な研究というのは、大体、米国で行われることが多くて、英国だとか、欧米での研究者が牽引することが多いのですけれども、欧米での文化的な背景と、日本の文化的な背景というのは違う部分も多くありまして、その文化的な背景に基づいた意思決定のやり方というものも異なるものが多くあるかなと思っております。

ですので、アメリカでユーザー調査が行われて、ユーザーはこのように考えるのだ、このように行動して考えるのだと結論が出たとしても、それがそのまま日本においても同様の状況になるとは限りません。

ですので、米国と日本と、それ以外の国といろいろな国で比較分析をすることによって、何が共通的なもので、何が違うのかということを明らかにするような研究というのも重要だと考えられております。

例えば、日本だと、個人主義の度合いというのは、アメリカとか欧米の国よりも低いとされております。周りの意見を聞きやすいということが言われておりますので、そうすると、例えば、そういう日本の人をターゲットにした意思決定のゆがめ方として、多くの人がこういうことやっていますよというような、そういうフレーミングを使うことによって影響を受けやすくなってしまうということが考えられるので、そういう文化的な特性を使った攻撃というのも今後考えられると思いますが、そういう特性というのを研究されていないと、問題も分からずに対策もされないということで、そういう研究が日本の中でも促進されるべきだと考えております。私もそういう考えのもと、日本とほかの国との比較研究というのを積極的に、今後もやろうと考えております。

○小塚座長　ありがとうございます。

そのほか、いかがでしょうか。大塚委員、お願いします。

○大塚委員　御報告ありがとうございました。私からは2点質問がございます。

1点目は、スライドの3ページ目のダークパターンに関する話です。

ここでは、LLMのダークパターンについての御研究を御紹介いただきましたけれども、ここでいうダークパターンというのは、開発者等が意図してユーザーをだまそうとか、操作しようというものに限られるのでしょうか。

逆に、開発者は意図していなかったのだけれども、LLMの出力の結果、意図せず、ユーザーがだまされてしまうとか、一定の自律性の侵害が生じてしまうといった場合について、何か御研究があるでしょうかという点について御教示いただければ幸いです。これが1点

目です。

2点目につきましては、8ページ目の支援するAIについての御研究です。

こちらでは、7番の引用の御研究において、AIの支援下では判断能力が改善しているけれども、それをやめてしまうと、必ずしも判断能力の向上が持続しないということが御紹介されておりましたが、ここで言う判断能力というのは、偽情報、誤情報を見分けるという点に限られるということになるでしょうか。

自律性という点で言いますと、そういった偽情報、誤情報にだまされないという点に限られず、何かもう少し自分で決めたという、まさにその点が問題となり得るようにも思いますが、そういったより広い自律性を支援する仕組みとしてのAIというものをどのように評価すればよいのかという点について、お考えがありましたら教えていただけると幸いです。よろしくお願いいたします。

○秋山上席特別研究員　まず、1点目ですが、今回紹介した文献2の研究においては、意図してつくり込まれたものだと認識しております。ただし、意図せずにLLMが開発者の想定外というか、想定していない害を生むことも知られておりまして、その1つが、シコファンシーと呼ばれるような現象で、LLMがユーザーと対話をする中で、ユーザーのある種の信念を過度に助長するような、媚びを売って、あなたの言っていることは正しいですよ、そのままでいいですよとか、そのように助長させてしまうようなことも問題だと言われております。

本来ならば、人と対話をする中で、それは、あなたは間違っていますよとか、こうしたほうがいいのではないですかと、適切なアドバイスを受けるところが、LLMがユーザーの信念を過度に助長してしまうことによって起きる問題というのも生じておりまして、それが、極端な話でいうと、自殺とかまで追い込まれてしまうような事例とかが報告されております。

2点目の質問ですが、文献7の研究では、偽誤情報に対しての判断力になります。ただし、人が判断しないといけないことは偽誤情報だけではなくて、様々生きていく上でのものがあります。例えば、自分のプライバシー情報の扱われ方というものも正解はなくて、人それぞれの生き方の話だと思っております。自分の名前を、あるサイバー空間上で公開していいのかどうか、別に自分の名前なんて多くの人に知ってもらって、知名度を獲得したいという人もいれば、いや、それは匿名でこっそり自分の趣味をエンジョイしたいという人もいるし、それは生き方の問題であって、正解はないと思うのですけれども、それをAIによって、あなたはこうしたほうがいいですよと勝手に決められるものではないと思いますし、ですので、何がリスクとしてあるのかということを適切に理解させた上で、自分が判断できるようなAIの支援というのは、今後も必要になってくるのではないかと考えられます。

○大塚委員　ありがとうございました。

1点目について少し補足の質問なのですけれども、そういった意図せず迎合的になって

しまうとか、そういった自律性侵害のパターンについては、これはまた、ある種マッピングというか、類型化というのが必要になってくると思うのですが、その場合、6ページ目のマッピングとは別のカテゴライズをすることになるのか、あるいは両者併せて何かマッピングを試みるほうが望ましいのか、恐らく今後の研究ということになると思うのですが、何か方向性について御見解がありましたら伺いたいと思います。

○秋山上席特別研究員　完全に別物ができるわけではないかもしれないのですけれども、足りない軸があるような気がします。単なるデザインの話だとか、選択肢がないとか、そういう話ではなくて、その対話の中で表出するものなので、ここでは表現できていないものがある気がします。具体的それは何なのだと言われると、まだ、十分に分析できていないところがあると思いますが、この分類の中に当てはめることでは足りないかなと考えられます。

○大塚委員　ありがとうございました。大変勉強になりました。

○小塚座長　ありがとうございます。

その辺りは私も非常に関心がありますね。そのほか、いかがでしょうか、御質問、御意見等どなたでも、坂下委員、お願いします。

○坂下委員　どうも御説明ありがとうございました。大変参考になりました。

この2番の引用文献とかを読んでいると、結局、UIだけの問題ではなくて、個人プロファイルとか対話の中でアウトプットが出てくると。

そうすると、悪意を持った事業者ではなくても、つくり出したものがダークパターンLLMになる可能性もあるというように何か解釈できたのですね、この論文を読んでいると。

そうなってくると、その事業者だけの問題ではなくて、つくるプロセスみたいなものも誰かが見る必要があるのではないかと。

それで、もし、この御研究の中で、第三者検証みたいな、こういうポイントは検証したほうがいいとか、そういうような議論がもしありましたら教えていただきたいと思います。

なぜかというと、自律的な決定というのは、契約をする自由だけではなくて、私はこれがほしいと思う自由もあると思うのですね。今回のAIというのは、後者のほうを、もしかしたら誘導してしまうかもしれないところがあると思うので、教えていただきたいと思います。お願いします。

○秋山上席特別研究員　大変難しいコメントになるかなと思います。私の研究の中でも第三者的な検証が重要ではないかと言及したものの、それが、いろいろなサービス形態だとか、いろいろな人への作用がある中で、まだ、具体的に何の情報を、どのようなプロセスでやることによって検証ができるのかという具体的なところまで、何か確立されたものがあるわけではないです。

さらに、LLMダークパターンのような話になってくると、LLM自体のプロセスもかなりブラックボックスなところもあって、より問題が難しくなるのかなと考えられます。よって、答えるのが難しい、というのが答えになります。

○坂下委員　ありがとうございます。

私もいろいろ調査していて、ブラジルが、例えばAIのインパクトアセスメントをやるというのを決めていたりとか、あと、ISOの中でも、AIを利用する組織に対する認証を始めたりとかというのが幾つか出ていますから、そういうものを何か組み合わせなくてはいけないのではないかとは考えてはいますが、先生方の御研究を、また、アウトプットを待ちたいと思います。ありがとうございました。

○小塚座長　ありがとうございます。

そのほか、いかがでございましょうか。

もし、今のところないようでしたら、これは、座長がよく理解していないだけなのですけれども、EUのDSAができたことで自律性の侵害とはどういうものなのだということで研究が進んで、それで、このマッピング、8類型が出てきたというお話だったと思いますが、8つに分けていますけれども、結局いずれも自律性の侵害になるものを挙げているわけですね。この8つの類型が分かったことで、そこでは、どういう政策的なインプリケーションがあったのでしょうか。

○秋山上席特別研究員　まず、こういう分類ができたことによって、ダークパターンというのは、今後出続けてくると思うのですけれども、新しいダークパターン的なものが出てきたときに、これは、DSAの自律性侵害に当たるものなのかと判断をするときのツールとして使えるというインプリケーションです。

○小塚座長　なるほど、つまり、この研究論文で取り上げられているものは、これは当たるだろうというものを挙げているのだけれども、枠組みができたことで、今後は当たるか、当たらないかの判定に使えると、そういうことであるわけですね。

○秋山上席特別研究員　そうですね。

○小塚座長　ありがとうございます。

非常に興味深いと思うのです。つまり、法律上は抽象的な形でしか書かず、日本の法令の普通の常識で言うと、政省令とかで具体化しましょうという発想になっていくわけですが、そうではなくて、抽象的な法令のままで、あとは、むしろ技術的な研究をしていくことで、それが具体化されていくという、何か官民の動きが合わさって規律の内容が充実していくと、そういう非常に面白い事例ということですね。

○秋山上席特別研究員　はい。

○小塚座長　ありがとうございます。

何か日本にとっても参考になるところがあるかもしれません。

そのほか、何かございますでしょうか。そろそろ時間は近づいていますが、よろしいですか。

ありがとうございます。それでは、質疑は、この辺りにさせていただきたいと思います。最先端の、また、いろいろと示唆のあるお話をいただきましたので、これは、我々の専門調査会として受け止めて、さらに議論をさせていただきたいと思います。

秋山様、お忙しい中、今日は、どうもありがとうございました。

○秋山上席特別研究員　ありがとうございました。

≪3．閉会≫

○小塚座長　それでは、予定した議事は以上ですので、事務局から何か連絡事項等ありましたらお願いいたします。

○江口企画官　本日は、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、事務局より御連絡させていただきます。

以上です。

○小塚座長　ありがとうございます。

本日は、これにて閉会です。皆様、御多忙のところお集まりいただきまして、ありがとうございました。

以　　上