

## デジタル空間における情報流通の健全性確保の在り方に関する検討会（第4回）

1 日時 令和5年12月15日（金）10時00分～12時00分

2 場所 オンライン開催

3 出席者

（1）構成員

宍戸座長、山本座長代理、生貝構成員、石井構成員、越前構成員、奥村構成員、落合構成員、クロサカ構成員、後藤構成員、澁谷構成員、田中構成員、増田構成員、水谷構成員、森構成員、安野構成員、山口構成員、脇浜構成員

（2）オブザーバー

一般社団法人安心ネットづくり促進協議会、一般社団法人新経済連盟、一般社団法人セーフティーインターネット協会、一般社団法人ソーシャルメディア利用環境整備機構、一般社団法人テレコムサービス協会、一般社団法人電気通信事業者協会、一般社団法人日本インターネットプロバイダー協会、一般社団法人日本ケーブルテレビ連盟、一般社団法人日本新聞協会、日本放送協会、一般社団法人MyData Japan、一般財団法人マルチメディア振興センター

（3）総務省

湯本大臣官房総括審議官、西泉大臣官房審議官、田邊情報通信政策課長、大澤情報流通振興課長、恩賀情報流通適正化推進室長、内藤情報流通適正化推進室課長補佐、上原情報流通適正化推進室専門職

4 議事

（1）構成員からのご発表

（2）意見交換

（3）その他

【宍戸座長】 それでは定刻でございますので、デジタル空間における情報流通の健全性確保の在り方に関する検討会の第4回会合を開催いたします。本日も御多忙のところ会合に御出席いただき、誠にありがとうございます。

議事に入る前に、事務局より連絡事項の御説明をお願いいたします。

【内藤補佐】 本日、事務局を務めます、総務省情報流通行政局情報流通適正化推進室の内藤です。

まず、本日の会議は公開とさせていただきますので、その点、御了承ください。

次に、事務局よりウェブ会議による開催上の注意事項について案内いたします。

本日の会議につきましては、構成員及び傍聴はウェブ会議システムにて実施させていただいております。

本日の会合の傍聴につきましては、ウェブ会議システムによる音声及び資料投影のみでの傍聴とさせていただいております。事務局において、傍聴者は発言ができない設定とさせていただいておりますので、音声設定を変更しないようお願いいたします。

本日の資料は、本体資料として資料4-1から4-3、また、参考資料4-1から4-2を用意しております。万が一、お手元に届いていない場合がございますら、事務局までお申しつけください。

また、傍聴の方につきましては、本検討会のホームページ上に資料が公開されておりますので、そちらから閲覧ください。

なお、本日は、江間構成員が欠席でございます。また、生貝構成員、後藤構成員、増田構成員は会議途中から御出席予定と伺っております。

事務局からは以上でございます。

【宍戸座長】 ありがとうございます。それでは早速、議事1に移ります。構成員からの御発表シーズンの続きということになりますけれども、本日は田中構成員、それから水谷構成員、遅れて御参加になりますけれども、生貝構成員からプレゼン御発表をいただくということを考えております。それでは早速でございますが、まずは田中構成員から20分ほどで御発表いただければと思います。よろしく申し上げます。

【田中構成員】 それでは始めさせていただきます。名古屋工業大学の田中です。本日はデジタル空間における情報流通の健全性確保の在り方について、それと関連する心理学領域の動向を中心にお話ししたいと思います。

まず、簡単な自己紹介ですけれども、専門は認知科学で、心の働きの中でも特に推論や意

思決定といった高次認知に対して実験心理学的なアプローチを用いて明らかにする手法をとっております。近年は多くの意思決定がデジタル環境を介在して行われるようになっていくことから、人の認知活動とデジタル環境との相互作用に対して、工学系の研究者と共同で研究に取り組んでおります。本日も最後のほうで、その取組についても少しだけ御紹介できればと思います。

デジタル空間における情報流通の健全性確保という、この本検討会のテーマに関連して一つ、心理学では最近大きな動きがありました。先月11月にアメリカの心理学会が誤情報対策に対する共同声明というものをしています。この共同声明とともに、心理学を誤情報の理解と対策に活用するというタイトルの報告書が出されています。この報告書では、なぜ人が誤情報を信じるのかや、誤情報に対抗するにはどのような介入が効果的か、などの問いに対して人の要因に焦点を当てながら論じられています。

専門的なところで言えば、記憶や推論のメカニズムや介入の長期的効果検証などは、まだまだこれから解明が必要な部分もあるんですけども、この報告書の最後に数百本の実験論文が引用されているように、学術的な知見が急速に蓄積されていることも事実でして、現時点で明らかになっている心理学的知見を社会問題の対策に活用する観点から、この報告書では制作者や実務家の方々に向けて誤情報による脅威に対処するための推奨事項というものが整理されています。本日は、この報告書の中でも出てくる幾つかの重要な知見に焦点を当てて、少し具体的に御紹介したいと考えております。

ちなみに、誤情報という用語の定義についてですが、御承知のように偽情報やフェイクニュースなどの関連した用語が使われることもあります。偽情報は発信者の悪意が仮定される用語ですが、ユーザー側や情報受信者の立場からはそのような悪意を判別することが難しいことも多いですので、心理学では誤っている情報という意味のみを示す誤情報が使われることが多いです。今回の発表でも、そのような意味で誤情報という用語を使わせていただきたいと思います。

誤情報に対する介入の捉え方は、いろんな捉え方があると思うんですけども、今日は個人レベルとシステムレベルに分けて捉えて、捉える考え方に沿ってお話したいと思います。まず、下のシステムレベルにはアルゴリズムによる介入や法的介入が含まれています。心理学が対象とするのは上の個人レベルによる介入で、BoostingやNudging、デバンキングなどがそこに含まれています。これらは、それぞれ目的ですとか背景にある理論的なものですとか、期間やタイミングが異なるもので、例えば、ブースティングのところにあるこのプ

レバンキングというのは、誤情報に対する事前の備えという位置づけになります。一方で、デバンキングは誤情報が既に拡散した後の事後対応という位置づけです。

基本的な考え方は上にあるようにnot a one-size-fits-all solutionという考え方で、単一で万能な手法があると仮定するよりは、まずどのような介入手法があり得るのか、この10年ほどで急速に研究が発展する中で複数提唱されてきた流れがあります。同時に効果の検証が行われていまして、ようやくこの図で表されているようなカテゴリーに整理されてきたという認識しております。

全て御紹介したいところですが、時間も限られておりますので、ここではプレバンクとデバンクに焦点を当てながら具体例を御紹介いたします。まず、プレバンクです。プレバンクは、事前に誤情報に対する心理的な耐性を築くことを目的とした介入手法になります。もともとは1960年代に社会心理学者が提唱した接種理論というものを土台としたものになっているのですが、現在はケンブリッジ大学の研究チームを中心に心理的予防接種という介入方法に発展していきまして、WHOやGoogleなどとの共同研究による大規模な実証研究が行われています。

このプレバンクの主な内容は、誤情報に近い将来、出くわす可能性のある人々を対象にその可能性を警告すること、また、誤情報に対抗するためのあらかじめ反論材料を提供する構成になっております。

その具体的な手法から、能動的なプレバンキングと受動的なプレバンキングに分けられています。ここでいう能動的というのはインタラクティブだとか対話型とか、ユーザーの積極的関与が前提とされている手法を指します。例えば、この黄色のBAD NEWSというゲームでは、ユーザーはあえてフェイクニュースを拡散するお題が与えられます。その目的のモットーは、よりフォロワーを獲得するためのメッセージを選択するというような構成になっています。その中で、このスライドの右側にあるような、誤情報を広めるためによく使われるテクニックがどのようなものであるのかというものを学んでいきます。ゲームではこのテクニックを理解して使うと、ポイントの代わりにフォロワーが増えていたり、各テクニックを理解するごとにバッチを与えられたりとか、あと、ゲームの中でも言葉遣いが平易で会話調にするなど工夫されていまして、比較的ユーザーが飽きにくく使い続けるような工夫がされているのが特徴です。

BAD NEWS以外にも政治やコロナなど特定のトピックに関連したゲームが後続で開発されておりまして、WHOと共同開発された一番下のGO VIRALは、できるだけ利用しやすいよう

に5分程度という短めに設定されているというものもあります。BAD NEWSゲームもGo Viralも残念ながら日本語版はないですけれども、ヨーロッパを中心に複数の言語に翻訳されています。また、効果検証もメタ分析が可能になる程度までは蓄積されていまして、次のスライドで述べる受動的なプレバンキングと比べると、比較的效果が持続する報告がされています。

能動的なプレバンクは効果が得られやすい一方で、時間が長かったり、ユーザーの積極的な関与が求められるために広くリーチしにくいというような特徴があります。そこで受動的なプレバンキングという手法では比較的短い動画ですとか、ポップアップ形式のテキストメッセージを提示する形で誤情報で使われるテクニックに対抗するための情報を提供する方法をとります。

例えば、ケンブリッジ大学のチームがGoogleと共同で実施したフィールド調査では、各テクニックごとに30秒から90秒程度の短い動画を作成し、YouTubeで視聴させたところ、誤情報の共有意図を低減させた結果も報告されています。このような能動的なプレバンキングも受動的なプレバンキングも誤情報に対する心理的耐性を向上させる効果が見られてはいるんですけれども、その効果は時間経過とともに減少する傾向があるということですとか、また、この効果検証そのものが北米やヨーロッパ中心で行われている現状で、日本を含めて他の文化圏でどの程度効果があるのかについては、研究が不足しているのが現状です。

続いて、デバンキングです。既に誤情報が発生していたり広がっている場合は、事後的にその影響力を弱める介入方法が必要になってきます。これがデバンクと呼ばれる手法です。例えばCOVID-19に関する誤情報への対策としてWHOやCDCが特設ページを開発して、それを訂正しようとした試みですとか、多様な誤情報を扱うファクトチェックサイト、マスメディアによる訂正情報を含むニュースですとか、あるいはSNSユーザーによる集合的な訂正など、様々な形式でデバンキングというものが日常的に行われています。

このデバンクは一般的に用いられる手法であるんですけれども、一方で社会科学研究を中心にその効果検証が行われておりまして、1度、誤情報の影響を受けると、それを事後的に緩和することが必ずしも容易ではないことも示されています。

なぜ訂正情報を出しているのに、まだ誤情報が拡散し続けるのかとか、なぜ誤情報を信じ続ける人がいるのか、こういった問いと密接に関わる心理現象が誤情報持続効果と呼ばれるものです。これは単に訂正情報を視聴していないとか、それを知らないから誤情報をうの

みにしているパターンもちろんあるんですけども、それに加えて、訂正情報を視聴して、訂正情報が出されていることを知った後もなお誤情報を信じ続けたり、その影響を受け続ける心理現象を指す用語です。

これまでの心理学研究では、この訂正情報に視覚的な注意を払っていても、また訂正情報の内容を記憶していてもこの現象が生じることが確認されていまして、認知的な処理のレベルの中でも高次認知レベルの観点からこの現象の解明が進められています。つまり情報を健全に伝達するといった時に、単に物理的に情報が遮断されることも重要な観点ですけども、ここでフォーカスが当てられているのは、物理的にはスマホのニュースアプリを開いていて、そこに訂正情報の記事が表示されていても、その後で心理的な障壁が生じることによって、その訂正の効果が制限されることに関する研究になります。

この誤情報持続効果が生じる心理的な要因は、複数明らかにされています。その一つに、真実錯覚効果という人の認知的な特徴があります。これは、同じ情報に繰り返し接触するとその情報の、正しく感じられるようになっていく現象です。この背後には、人が同じ情報を繰り返し接触していると、親近性だとか情報処理の流暢性が高まることによって、それが正しさのシグナルとしてヒューリスティックに活用されていくというような認知的なメカニズムがあります。

例えば、右側の図は縦軸が誤情報の影響を表していまして、上に行くほど誤情報の影響力が強いことを意味します。横軸は誤情報への接触頻度を被験者間で統制したものですけれども、例えば接触頻度が1回の場合より3回繰り返し接触すると、その影響力が強くなります。これが真実錯覚効果の1例です。

もし、繰り返し接触することで真実錯覚が生じるのなら、訂正情報も同じように繰り返し行えばよいのではないかと思われまますけれども、実験ではそのようにうまくいきませんでした。例えば、誤情報を1回提示した後で訂正情報を1回提示した場合と3回提示した場合は、有意な差は見られませんでした。つまり、誤情報の3倍の頻度で訂正を行ったとしても、誤情報の影響は誤情報に触れなかった0回のレベルまでは戻らないことを示しています。このような現象は真実錯覚効果の非対称性と呼ばれていまして、誤情報の受け入れられやすさだとか信じられやすさに対して、それを事後的に訂正することの難しさの間にギャップがあることを示唆します。

このような認知的な特徴に基づいて、できるだけ訂正の効果を上げるにはどのような具体策があるのかという問いに対して、2020年にデバンキングハンドブックという報告

書が出されています。心理学における学術的な知見に基づき、誤情報に関する心理的な特徴 17点と実行可能な対策 10点が整理されています。残念ながら、こちらも日本語版はないんですけども、日本での対策を考える上で、先ほどのプレバンキングのガイドブックですとか、冒頭で述べたようなAPAの推奨事項と併せて、この辺りの知見はもう少し日本でも注目されても良いのかなと思って御紹介した次第です。

ここまでは、主に心理学の中での動向について御紹介してきましたが、最後に私たちの研究チームが最近関心を持っていることについてお話ししたいと思います。先ほどの図を用いますと、私たちが関心を持っているのはこの個人レベルとシステムレベルの相互作用についてです。人間の情報の認知処理ですとか、意思決定がデジタル環境ともう完全に独立して行われることが少なくなっているのです、このデジタル環境と心理の相互作用を理解することの重要性が今後さらに高まっていくと予想しています。

そのような研究はヒューマンコンピューターインタラクション研究と呼ばれてますが、例えば次のような研究があります。検索エンジンが返す情報のランキングが人の検索行動に影響を与えて、結果的に人の意思決定に影響を及ぼす現象があります。例えば右上のグラフはランキングごとのクリック率を示したものですけれども、ランキングが下がるごとにクリック率が顕著に下がっていく傾向が、この研究以外でも一般的に観察されています。デバンキングのための訂正情報に関してもデジタル環境を経由して伝達されることが多いので、このようなクリック行動は訂正情報へのアクセシビリティと関わってくるでしょうし、それが結果的に人の心理にも影響することが考えられます。

ところが、このクリック行動という重要だと思われる要因については、これまであまりよい測定指標が開発されてこなかったために、単純なクリック数だとか、クリック率だとか、そういった分析しか行われてこなかったような背景がありました。

そこで私たちの研究チームでは、クリック行動の個人差を測定する指標を新たに開発しました。そして訂正情報が人をどの程度、訂正情報をクリックしているのかを客観的に示すための実験を行いました。

簡単に説明しますと、デバンクで重要になってくるのは多くの人々が誤情報を信じている時です。このスライドでいうと、赤色で表している部分になります。実験では誤情報に、誤情報というラベルが貼られていまして、そこをクリックすると訂正情報が表示されるデザインになっています。クリック行動の特徴を分類したところ、信じている誤情報に対する赤色の訂正情報を積極的にクリックする集団と、選択的にそれを避ける集団がいることが分

かりました。

例えば、10個の誤情報を信じている場合に、左側の集団は平均すると4個以上の訂正情報をクリックしているのに対して、右側の集団は1個にも満たない7%という数字でした。この右側の集団は青色の、こちらにも誤情報というラベルが貼られているんですけども、これはそれなりにクリックしているので、ラベルで避けているというよりは、自分が信じているものに対する訂正を選択的に避けていることが分かります。

このようなデジタル空間上で生じる心理的、かつ選択的な回避行動は、誤情報が拡散して社会問題となったときにどのように訂正情報を届けるのかという問いと密接に関わってきます。例えば、専門家の方々やジャーナリストの方々が事実と照らし合わせて丁寧にファクトチェックをしたとします。そして、それをデジタル空間上に公表して誰でもアクセス可能な状態にしたとしても、この心理的な、この選択的回避という障壁によって誤情報が信じ続けられることが生じ得るということを示す実験結果でした。

現在、私たちの研究チームでは、このような選択的回避傾向を持つ集団の特徴を明らかにすることですとか、どのような介入方法が考えられるのかについて研究を進めているところです。

こちらは、まとめのスライドになります。

私からの話題提供は以上です。

**【宍戸座長】** 田中先生、本当にありがとうございました。それでは、ただいまの御発表につきまして御質問、御意見のある方はチャット欄で私にお知らせいただきたいと思いますのですが、いかがでございましょうか。

それではまず、越前先生お願いいたします。

**【越前構成員】** 田中先生、心理学的な観点から興味深い発表ありがとうございます。非常に興味深かったと思います。

一つ質問ですけど、11ページの真実錯覚効果という図がございしますが、誤情報への接触頻度と訂正への接触頻度ということがございしますが、一番右の3回誤情報に言及した際に、3回訂正への接触頻度があれば0回、何も触れてない人より下がるということでしょうか。要するに、誤情報の言及頻度と同等程度、訂正への接触頻度があれば、この3回の部分と比較しても下がるということでしょうか。ここについて、もう少し御説明いただければと思いますが。

**【田中構成員】** これが接触していない、一番左のグラフが誤情報に触れてない人です。

【越前構成員】 なるほど。

【田中構成員】 誤情報の影響は、これは誤差程度に言及することがまれにあるようなレベルです。これが触れる前の影響力がほとんどない場合と比べると、もちろん訂正を行うことはまず効果があるのでやったほうが良いというのは、ここの差がそれを表しています。ただ、ここで例えば3倍やったとしても、ここに戻ることはないですし、3回触れた場合でも、ここと比べると顕著に下がっていくわけですが、ここまで戻ることはないというようなことを示す結果になります。

【越前構成員】 ありがとうございます。私が誤解していたようです。これはまず、誤情報に接触した後に訂正情報に接触したということによろしいでしょうか。リポートナーという。

【田中構成員】 そうです。最初に誤情報に触れて、3回繰り返しやった後で0回、1回、3回の訂正情報に接触するという。

【越前構成員】 分かりました。そうすると最近ではソーシャルメディア等だと、例えばXのバードウォッチみたいに、誤情報があった時に訂正に近い情報も同時に提示する場合がありますけど、そういった場合の影響、そういった場合の心理的効果というのは何か検討とかございますか。それ同時提示することでほとんど拡散がないような状態になるとか、そういった意味での研究ってのはあるんでしょうか。最近ではプラットフォームもすぐにそういった訂正情報、最初はどうか分かりませんが、しかるべき後に同時に情報を出すケースが結構あるかなと思いましたので、そこについてお聞かせいただければと思います。

【田中構成員】 たしか、あると思います。今すぐ具体的な結果を厳密に言うことはできないんですけども、むしろ訂正のパラダイムで研究されることがありまして、誤情報に触れたことがない人が最初に訂正から触れるというパターンが想定されると。その時に誤情報を引用しながら訂正に触れることがあると。そうすることによって、訂正の効果ではなくって誤情報がかえって信じられてしまうんじゃないかという、バックファイア効果と言われている現象があるんですけども、それについての検証がされていまして、結構研究が蓄積されていまして、そういったバックファイア効果は発生する可能性は今のところ、支持されていないような結果になっています。

【越前構成員】 ありがとうございます。よく分かりました。以上です。

【宍戸座長】 ありがとうございます。さすがに田中先生の御報告が面白いので幾つも御質問が来ていますので、できるだけ簡潔に御質問、御意見いただければと思います。

それではまず、奥村先生お願いします。

【奥村構成員】 ありがとうございます。私はどちらかというと情報を出す側で、それを認識する側のことについては非常によく整理されて、よく分かりました。ありがとうございます。

お聞きしたいのは、そのような今、バックファイア効果に少し触れられましたけれども、特定の国では国際会議などではそういう認識というか、何ていうんでしょうか、間違っただけの情報信じ込んでいる人たちの集団が例えば暴力的な行動に出るとかというような、要するにニュースを信じない、もしくは誤情報を信じてしまっているがために何か行動を起こしてしまうというようなことがあるわけですが、そのような行動を起こすか、起こさないかというようなきっかけになるようなところは、心理学の分野ではどのような形で今、知見がためられているのかということを知りたいということです。よろしくをお願いします。

【田中構成員】 ありがとうございます。つまり、この部分に関する事かと思えます。認知プロセスがあって、信じ続ける人がいて、そこから行動に移るかどうかの、このところに関する知見についての御質問だと思うんですけど、この辺りは様々な別のファクターが絡んでくるので、私は押さえきれないのが正直な報告になります。

例えば、ワクチンに関する誤情報を信じ続けている人がワクチンを打ちに行くとか、行かないとか、逆に海外ですと妨害しに行くみたいなパターンってあると思うんですけど、そういったものも近くにあるかどうかですとか、どれぐらいの数があるかどうかですとか、どのようにWHOとかが介入しているかによっても変わってくるので、また、そこには複雑な要因が絡んできていると思います。

【奥村構成員】 ありがとうございます。いや、御質問したのは、アメリカで連邦議事堂の襲撃事件がありまして、その時に各ニュース機関がいろいろ選挙結果についての報道をして一生懸命やったわけですが、結局それが大規模な人たちの行動を招いてしまったと。おっしゃるような政治的ないろんな働きかけも要因としてあったので、どこまでニュースメディアができたことは分からないんですけど、ということで問題提起をさせていただきます。ありがとうございます。

【宍戸座長】 ありがとうございます。それでは次に安野先生、お願いします。

【安野構成員】 中央大学の安野です。田中先生、大変興味深い、また示唆に富む御報告どうもありがとうございました。私もこの分野、とても関心がありますので、とても勉強に

なりました。私からは2点ほど御質問というか、コメントも入っちゃうかもしれないんですけども申し上げます。

まず一つは、illusory truthの効果、私もこれ、すごく面白いなと思って関心を持っているんですけども、そこで訂正情報がどうしても信じられないというのはすごく重要なことだと思うんですね。その時に、先に誤情報に接しているから訂正情報の効果がないのか、それとも誤情報が信念に合致してしまっているから訂正情報の効果がないのか。illusory truthは、もうとにかく、たくさん接触すれば認知的に処理しやすいから真実だと見えてしまうということです。それは事実だと思うんですが、同時に信念に合致してしまっていることが問題の部分もあるのかもしれないとも思いました。もし、どちらを先に出すかということで効果が違う可能性の御知見や御意見ありましたら、よろしく願いいたします。これが1点です。

もう一つは今のことと関係しているんですけども、セレクトティブアヴォイダンス、これもすごく重要かつ面白い現象だと思うんですが、これも何というか、ワクチンの割と強い信念を持ったとか、そうですね、強い信念とか、あと利害関係に関するようなものと特にアヴォイダンスが起きやすいのか、トピックによって何らかの起きやすさが違うのかどうか、もし御知見ありましたらお願いいたします。

以上2点です。ありがとうございます。

**【田中構成員】** ありがとうございます。それでは、2点目から先にお答えいたします。このセレクトティブアヴォイダンスについては、本当にこの研究で初めて明らかになった現象なので、まだ複数のトピックで検証する段階までは行っていません。一応この研究ではワクチンを接種したかどうかですとか、周りにコロナに感染したかどうかみたいなことも聞いているんですけども、それとの有意な関係は見られていないのが現状です。

もう一つ、真実錯覚効果について、訂正情報を先に出したときにどうなるのかというのは非常に重要な御指摘で、今、記憶研究で研究が進められているんですけども、そのメカニズムについてはまだ幾つか諸説あります。アップデートがうまくいっていないのか、訂正情報と誤情報が両方保持されて検索の時にうまくいかないのか、その辺りについては幾つかのモデルが提唱されていますけれども、まだ一つに収束するほどまでは至っていない現状です。

**【安野構成員】** どうもありがとうございました。

**【宍戸座長】** ありがとうございます。それでは続きまして山本先生、お願いいたします。

【山本座長代理】 ありがとうございます。非常に重要なお話をいただいたかと思えます。私から2点コメントですけれども、1点は、私も鳥海先生なんかと健全な情報空間の鍵になるのは情報的健康、インフォメーションヘルスだというようなことを申し上げてきたんですけれども、要するに健康とのアナロジーで捉えてきたところがありますけれども、今回、お話を伺ってプレバンキングというのが例えば心理的な予防接種であるとか、心理的な耐性というものをつける。しかも、そういった研究がWHOと共同でやられているというところで、非常に健康問題のような形で議論されているんだなということを知りまして、大変、何ていうんでしょうね、方向性がまた私自身、見えてきたなと感じたのが1点目でございます。

それから2点目ですけれども、現状の情報空間というのを健全化することの難しさですけれども、これは非常に個別の現象面というよりも構造的な問題ですとか、これは例えばアテンションエコノミーというビジネス構造にも関わりますが、さらに個人のレベルでいくと認知システムのような、我々がふだん意識しないような、そういう非常に認知のレベルで一つ原因というんですかね、根源的な問題があると。この難しさを感じたところです。

そう考えると、単純にディスインフォメーションなんかをモデレーションしていけばよいとか、あるいはファクトチェックを増やしていけばよいということ、これも非常に重要ですけれども、こういったことのやり方ですとかが非常に重要で、私もリテラシーについてはICTリテラシーの検討会、座長やらせていただいておりますけれども、リテラシーの場面でも今日、能動的な接種というところで認知科学の知見というものを使っていくというお話がありましたけれども、リテラシーでもこういった認知科学の知見を積極的に参照していくことが求められるなと感じたのが1点と。

それから訂正情報、ファクトチェックのようなものをどう届けるかということも単純に届けばよいということではなくて、いかにそういったものを効果的に届けるのか、例えばプロメントのような話とか、このやり方についても、こういった認知科学との連動というのは非常に重要だなと感じたということです。

雑駁な感想ですけれども、以上です。ありがとうございました。

【宍戸座長】 ありがとうございます。田中先生から何か追加でコメントとかございますか。

【田中構成員】 山本先生が今、おっしゃったように心理学はインディビジュアルレベルの中での研究が主になりがちですけれども、誤情報という社会問題を捉えていく上では、こ

の2つのレベルがどうやって連携していくのかというのが今後の課題になっていくのかなと考えております。

【宍戸座長】 ありがとうございます。若干お時間も押しておりますので、できるだけ簡潔にやり取りを、このラウンドはさせていただければと思います。

森先生、お願いします。

【森構成員】 森です。御説明ありがとうございました。本当に勉強になりましたし、驚くべきことをいろいろ伺ったと思っています。

一つ感想、一つ御質問で、一つは個人レベルとシステムレベルの相互作用、Human-Computer Interactionのところですけども、13ページのところですけども、これ、本当に今後そういう、何ていうんでしょうか、人間の認識を形成するといいますか、考え、信念、ここでは信念とお書きいただいていますけれども、そういうものがコンピューターとのインタラクションに非常に影響されるといいますか、深く関わってくる。ここではsearch engine manipulation effectとありますけれども、恐らくはもっと生成AIを頼るようになっていくだろうと思いますので、そうするともっとこの問題は深刻に展開するんじゃないかなと思いました。それが一つです。

あともう一つは14ページの、先ほどからお示しいただいている訂正情報へのアクセスの仕方ですけども、これをどのようにされているのかということをお聞きしたいと思えます。どのように表示されているかということです。例えば、私がイメージしたのは、「この情報は誤っているとの指摘がありますが、より詳しく見ますか？」みたいな、そういうバナーを出してクリックするかどうかを観察されているのかという、そういうやり方のことを教えていただければと思います。

もう一つはケンブリッジアナリティカの文脈で、ウェブの閲覧履歴を使って脆弱性のある人をあらかじめ選ぶというところがあったと思います。それはどっちかという仮説を先に立てて、そこにアプローチしているわけで、先生がここで御紹介いただいた研究というのは、まずはクリックコードがどうなのかということを見るということだと思わんですが、ケンブリッジアナリティカは衝動的怒りに流されるとか、陰謀論に弱いという、そのように推測される人たちの集団をあらかじめ選んで、そこに対してアプローチするとやっていたけれども、そういうこととの関係ですね。こちらの選択的な回避の御研究との関係について、もし何かありましたら教えていただきたいというのが2番目の質問です。

以上です。

【田中構成員】 ありがとうございます。まず、1点目については、最初に事前信念を測定した時に誤情報などを見せるんですけれども、その後で、先ほどお見せしたのものの中にはファクトチェックの結果、誤っているというものが含まれていましたという教示文を出します。その上で、aと書かれている画面をお見せして、リード文は誤情報の冒頭部分になるので参加してくれた人たちは1度、目にしている情報になります。それに対して誤情報というラベルが貼られていたり、正情報というラベルが貼られているようなデザインになります。

2点目については、ケンブリッジアナリティカの点で問題になったのは、パーソナルな情報を無断で取得していた点になるかと思います。それに基づいて、ビッグファイブのような指標を使って脆弱性のある人というものを、マイクロターゲティングのような感じで狙ったというところに問題があると記憶しています。私たちのこの研究は、そういったパーソナル情報はまず使わない前提でおります。あくまでも実験環境としてクリックコードを測定していますし、個人情報というものを使って狙っていくようなスタンスとは明確に線引きをしているような入り方でっております。もう少し、実験ではこういった特徴を見るんですけれども、より一般的にはそういった情報を使わなくてもアヴォイダンス群がアヴォイダンスを少し緩和して、ファクトチェック見てもいいかなというような気持ちになるにはどういった介入が可能なのかというような方向性で研究を進めていくつもりです。

【森構成員】 なるほど、よく分かりました。ありがとうございました。

【宍戸座長】 ありがとうございます。それでは落合先生、お願いします。

【落合構成員】 どうも御説明ありがとうございます。簡潔に御質問させていただきます。

一つが、今回の御発表の内容、非常に興味深い内容だったと思っておりました。一つお伺いしたいのが、地域ですとか人種だったりだとか、あと何か受け手の教育の程度だったり社会的身分だったり、こういうものによって傾向が違うような部分というのが出てきたりするところがあるのかということです。また、これが仮に差異があるのだとすると日本においてどういう研究がされているのかも、また重要なかなとも思いました。この点について、もしお分りの範囲があれば教えていただくとありがたいなと思いました。

もう1点が、訂正情報について、誰がこの訂正情報を流しているのかによって、もちろん御説明いただいたように、そもそも最終的に届いていなければということもあると思うのですが、主体によって訂正の効果が変わる部分があり得るのかどうかや、また、その訂正においてどういう内容を示しておくとのよいかの方が重要かと思っております。

御説明いただいた中で、没入感があるような取組のほうが、より訂正の効果はありそうであると受け止めましたが、内容面でも何か考慮すべき部分があるかを教えていただければと思います。

以上です。

【田中構成員】 ありがとうございます。まず、3番目から逆順になってしまうんですけども、やり方、訂正をどういうふうにすれば効果があるのかについては、先ほどのこのハンドブックの中でも具体的な説明があります。私たちが書いたこの展望論文の中でも幾つかありますので、もしよかったらそちらを御参照ください。訂正する情報ソースの影響についても、こちらにも研究が蓄積されていまして、どういった情報ソースだとどういった訂正の効果があるのかとか、弱まるのか、そういった研究も蓄積されています。

一番最初の御質問で、人種だとかそういったものが関係するのではないかというのは御指摘のとおりで、例えば受け手側と誤情報の人種が違う場合に、敵対的な人種的なステレオタイプを持っている場合に訂正情報が効きにくいですとか、訂正情報のほうが敵対的な人種、ステレオタイプを持っていると効きにくくなるか、そういった研究もあります。

【落合構成員】 ありがとうございます。よく分かりましたし、また、頂いた論文や、ハンドブックを参考にすることが非常に重要だと思いました。

【宍戸座長】 ありがとうございます。それでは、田中先生とのやり取りはここまでとさせていただきます。貴重なインプットをありがとうございました。

【田中構成員】 ありがとうございました。

【宍戸座長】 それでは続きまして、水谷構成員から御発表をお願いいたします。

【水谷構成員】 関西大の水谷でございます。よろしくをお願いいたします。資料は事務局から投影していただけるということですが、お願いできますか。ありがとうございます。

先ほどの田中先生の御報告の中の個人とシステムの相互作用という話は、法学の分野でも、今後、非常に重要となってくるんじゃないか、今日の私の報告にも関連してくるのではないかと思いつつ興味深く聞いておりました。ではさっそくお話しさせていただければと思います。

私は現在、関西大学の社会学部におりますが、憲法とメディア法を専攻しております、準抛国というか、主として参考にしているのはアメリカですので、今日はその観点からプラットフォーム・ガバナンスのお話をさせていただきます。

目次を飛ばして3ページ目に進んでいただけますか。私がずっと興味関心を抱いている

のは、デジタルプラットフォーム事業者、単にDPF事業者と呼びますが、そうした事業者が私たちユーザーの情報の発信や閲覧や需要に対して、環境のデザインをもって非常に大きく作用しているのを憲法やメディア法の観点からどのように考えていくべきかという点です。こうした作用は、大きく分けて3つの諸要素から構成されていると思います。アーキテクチャとアルゴリズム、それから今日のお話の主要な部分になりますが、コンテンツ・モデレーションですね。こういった各種機能を有しているDPF事業者は、最近ですとニューガバナー、新たな統治者と呼ばれたり、あるいは、これは山本龍彦先生が御指摘をされていますけれども、国家をリヴァイアサンと例えた場合の対抗存在としてのビヒモスというような存在として語られる、つまり単なる大企業・社会的権力を超えるものとして語られることが多くなっております。

ただ、この後お話ししますが、これはDPF事業者が国家と同一視される存在であるということを示しているわけでは必ずしもなくて、国家に類似する機能を持ち始めて匹敵する存在になりつつあるだけけれども、ただやはり、国家とは違う側面も持つ、そういう特殊なアクターとして位置づけられているとも思います。

次のページをお願いします。特にコンテンツ・モデレーションに関しては詳しいプロセスとか、アルゴリズムを用いた自動的なコンテンツ・モデレーションがどれぐらい使われているかは参考資料を後ろにつけてありますので、そちらを御覧いただければと思います。

研究レポートとかを参考に記述的に描き出してみたんですが、次に挙げるような特徴がDPF事業者の情報管理にはあるかなと考えています。一つは、コンテンツ・モデレーションについてもAIが非常に重要であると。コンテンツの種類によっては9割がた、AIによってフラグ立てが行われている現状があると。これは1日当たりで膨大な量のコンテンツをチェックしなきゃいけないので、人力ではもう到底手が回らないということでAIが活用されている側面ももちろんあります。

である以上、このコンテンツ・モデレーションの世界というのは昔ながらの法学がイメージしてくるような、法廷でこの言論は虚偽かどうかとか、あるいはこの言論は権利侵害がされているかどうかというのを緻密に時間かけて議論して判断するというような仕組みというよりは、どちらかというとパブリック・ヘルス的世界、例えば誤情報でいうと誤情報にユーザーが接触する確率を減らすといった世界、むしろ感染症対策に似ているかなと思いますけれども、そういう世界観になってきているわけです。

当然ですが、そうした接触をゼロにすることはできないので、コンテンツの削除において

も、日々、膨大な量のエラーが出るわけですね。ここでいうエラーは過剰にコンテンツを消してしまったり、あるいは本来消されるべきルールに反するコンテンツが残っていたりというような状況を指しますけれども、これらが膨大な量のエラーが毎日出るということ为前提に考えなければいけないということですね。

またよく言われるのは、D P F 事業者に対する規律は、この後お話ししますが、基本的にアメリカなどは自由放任でやってきたわけです。けれども、ではD P F 事業者は野放しになっていて無秩序なのかというと全然そんなことはなくて、まずD P F 事業者は国家と違って企業ですので、ユーザーや、特にソーシャルメディアに関しては広告主ですね。これらの存在が非常に大きいと考えられています。なので、ユーザーや広告主からそっぽ向かれるような状況は彼らにとって好ましくないので、規制がされていなくても彼らは自主的にコンテンツ・モデレーションを行うわけです。このようにD P F 事業者は、背景に非常に複雑なステークホルダーを抱えている。それによってD P F 事業者のガバナンスが決められている側面があるというのが第二点です。

そして、そうした背景があるからこそ、ケイト・クロニックという、D P F 事業者をニューガバナーと位置づけた研究者が、自身のブログでここ数年のビッグテック改革、まさに透明性の確保とか、そういう部分については、直感に反するかもしれないけど実は黄金時代だったと指摘しています。確かにD P F 事業者は透明性レポートも自主的によく出しているし、それこそ、フェイスブックは監督委員会という最高裁みたいな外部組織を、非常に多額のお金を出して作ったりしていて、実はそうした改革が自主的に進んできた時代でもあったわけです。

しかし、昨今の経済不況でビッグテックも経済的にその力が衰えてくると、社会に対する影響力は弱まっていないけれども、トラスト&セーフティだとか、公共政策の部門などから人とか資金が引き上げられてしまうと。そういう自主的な取組の後退が見られるようになるわけです。

次のページをお願いします。そうした前提のもとで、アメリカのデジタル空間の規制に関する話をさせていただきますけれども、基本的には先ほどお話ししたように政府は介入しないし、介入しにくいといったような状況です。特にソーシャルメディアに関しては、アメリカには憲法修正1条、これは表現の自由を定めた条文ですがけれども、この保護が日本なんかよりもはるかに強力な保護を私的アクターに与えます。ですので端的に言えば、例えば偽情報の発信者に制裁を加える法規制なんていうのはアメリカでは即違憲になると思います

し、DPF事業者に対する法規制を行うことも原則として非常に困難であるという状況だと思います。

さらにアメリカ憲法の特徴は、この後、生貝先生がEUのお話をされるとは思いますけれども、ヨーロッパ圏と違って憲法というのはあくまで国家権力を縛るものであると。そしてDPF事業者は、あくまで国家ではなく私的アクターにすぎない。私的アクターに憲法の統制を効かせる場合は、アメリカではステイトアクションの法理というのがあるんですけども、その非常に厳しい条件を満たさないと憲法の縛りというのは基本的に私的アクターに及ばないという側面があります。また先ほどもお話ししましたが、表現の自由が世界的に見ても強力に保護されている側面がアメリカにはあり、一部の論者がこれを修正1条の「武器化」というような批判をしたりしていますが、連邦最高裁の判例や多くの修正一条理論では、今でもそうした強力な保護を後押ししているところがあります。

それに付け加えて、通信品位法230条という非常に強力な法律の存在もあります。次のページをお願いします。アメリカのDPF事業者の成長を支えた一側面として、このCDA、通信品位法230条というのがよく注目されます。これは、ある論者に言わせると憲法の修正1条のステロイド強化版だと評されたりするわけですが、日本のプロバイダ責任制限法3条に相当するような条文が書いてあるわけですね。ただ条文では、DPF事業者も含め、双方向型コンピューターサービスのプロバイダまたはユーザーに対して、発信者や権利侵害情報の被害者からの責任追及に対して非常に広範な免責規定を設けています。時間がないので省略をしますが、もともとCDA230条ができる以前、インターネット草創期のオンライン企業に対してどの程度、情報媒介者として法的責任を負わすべきかというのが下級審の判例で議論されたんですけども、そこでいろいろ課題が生じました。端的に言うと、モデレーションをしっかりとやった人たちは責任を負わされ、逆にモデレーションをやらないという人たちは免責される可能性があるという、いびつな状況が生じそうになったわけです。その課題を解消するためにCDA230条は、96年に制定されました。

96年というので皆さん、お分かりかと思いますが、この頃は今のDPF事業者のような巨大な存在はなくて、むしろネット企業は新興産業だったので、ここに強力な法的責任を負わせるような状況が生じると企業の発展可能性を阻害してしまうことが懸念されました。なので逆に広範な免責を与えて、むしろ事業者にコンテンツの管理は自主的にやらせる、そういう自主的なモデレーションを促進するようなことを目的にこの条文というのは作られたと言われていています。

ただ、特にトランプ政権前後からアメリカではD P F 事業者に対する規制の議論があがっています。しかしそれらの議論の中には、E Uとは方向性の違うものが散見されます。特にトランプさんがツイッターからBanされたように、保守派は、D P F 事業者が保守派を狙い撃ちしてモデレーションしていると勘ぐっています。そして、コンテンツ・モデレーションを狙い撃ちでやるなんてけしからんと。実際にそうかどうかは私個人としては疑義があると思いつながりながら見っていますが、そういうふうに保守派の人たちが反発をして、むしろD P F 事業者はコンテンツ・モデレーションをすべきでない、コンテンツに手を触れるなどというような規制、いわゆるコモンキャリア的な規制をしようとしている州などが見られます。

次のページをお願いします。その典型例がフロリダとテキサスの州法です。これら2つがどちらも保守派の知事のもとで出来たんですけれども、基本的にソーシャルメディアプラットフォームがユーザーに対してアカウント凍結とか、コンテンツの削除とか、あるいはシャドバンといったことを行うのを制限しようという法律を作っています。ただ、これはD P F 事業者がコンテンツをモデレーションする自由も表現の自由としてアメリカでは考えられている節があるので、こうした州法はそれを規制することになるとして、差止めを巡って訴訟で争っている最中です。そろそろ連邦最高裁で判断が出ると思いますけれども、そういう状況があるということですね。

次のページをお願いします。D P F 規制について、E Uはむしろ、コンテンツの管理をもっとやれ、デジタルサービス法でもシステミックリスクをちゃんと分析評価して、軽減措置をちゃんととりなさいというような規制が入っていたりするわけですけれども、それとアメリカの規制の議論は全く異なる側面があるわけです。一方でD P F 事業者の透明性、不透明さ、彼らのコンテンツの管理のプロセスの不透明さをめぐる議論というのは、実は世界的に共通した話題としてアメリカやそれ以外でもあがったりしています。

例えば、国連の特別報告者のデビット・ケイさんが、D P F 事業者のコンテンツ管理の仕組みを、プラットフォーム法と評していて、非常に不明瞭な仕組みをしいていると指摘しています。デジタル空間の秩序形成の主導権が公的機関から私的機関であるプラットフォームに委ねられていると、これが憲法が大切にしてきた法の支配という仕組みを後退させるのではないかというような懸念があったりとか、あるいは、もうちょっと怖い話でコンテンツ・カルテルというのが形成されているというのをアメリカ研究者が指摘をしたりしています。

これはD P F 事業者間やD P F 事業者と政府の間で、有害な情報への対策等についての

協調体制がとられるわけですね。その中で、不透明な中で適切なモニタリングもなくコンテンツやアクターをサービスから排除するようなことを、一致協力してプラットフォーム間  
が内々で取決めをしてしまっていると。まさにカルテルのように。さらにそこに政府関係者  
の関与を強めると、DPF事業者が検閲代理人化するんじゃないかという懸念が指摘をさ  
れていたりします。

次のページをお願いします。今、言ったような、DPF事業者内部のものと、DPF事業者  
と政府間のものという、二つの不透明さを背景に、アメリカの連邦議会で、プラットフォー  
ム説明責任及び透明化法案というような、内容的に見ると部分的にアメリカ版DSAと言  
ってもいいかなというような法案が出てきたりしています。ただ、これは法案レベルなので  
実際に成立するかどうか、かなり怪しいところだと私自身は思っていますけれども。この法  
案は、以前から議論されてきたんですけど、今回もまた出ているようです。

次のページをお願いします。もう一つ、アメリカでさっきの検閲代理人化の観点で今、議論  
が沸騰しているものとして、jawboningと呼ばれるものがあります。jawboning、日本語で何  
て訳すか、今、迷っているんですけども、取りあえず「口先介入」としてあります。要す  
るに、DPF上で事業者が合法的な言論を、つまり法には反してないけどプラットフォーム  
のポリシーには違反しているというようなものもモデレーションしているわけですが  
も、そういった本来合法的に許されるはずの言論の取締りを行うために、政府関係者がDP  
F事業者に対して非公式の圧力をかけるというようなことがアメリカで問題になっている  
わけですね。

例えば、バイデン政権はCOVID-19対策の一環で偽情報対策をしなきゃいけなかつ  
ただけけれども、先ほどお話ししたとおり、アメリカでは偽情報の規制はかなり困難なわ  
けで、法規制でなかなかできないもんだから、それこそFBIとか政府、ホワイトハウスの  
関係者がDPF事業者と綿密に協力して、プラットフォーム上から例えばワクチンに対す  
る偽情報とかウイルスに対する偽情報をモデレーションするように圧力をかけていたんだ  
と指摘されています。これが一部、表に出てきたことによって、モデレーションされた側の  
人たちが裁判を起こしてしまして、DPF事業者に政府関係者が接触をすること、これを表  
現の自由の観点から差止めを求めるといった訴訟が現在係争中です。

これも連邦最高裁に上がっているんですけども、恐らく今年も終わるんで来年ぐら  
いに判断が出るかなという感じです。ただ難しいのは、政府によって事業者に適切な説得をす  
る行為は、これ自体は本来許されてしかるべき行為なわけです。憲法上もこれは問題ないは

ずとアメリカも考えているんですけど、そこに強制が入るとまずいと。

例えば有害な言論、違法じゃない言論をDPF事業者に消させるために、事業者の担当者に政府機関の警察が、これを消さないんだったらお前も逮捕されるぞ、とか言って脅迫をする場合、こういったものはまずいだろうというのは何となく分かるわけですけど、ではどこで線引きができるのかというのは非常に難しいということで、アメリカでも議論になっています。

なので、むしろここにも透明性確保のための規制を入れろという提案が在野からあがってきていて、ソーシャルメディア検閲透明化法というものを作ったらどうかというのが在野の研究者からあがっているところです。むしろ、これは単なる在野の提案ですので、今のところ連邦議会にはあがっていないと思います。

次のページをお願いいたします。さてここまで、透明性に関する議論をいろいろあげてきましたけれども、透明性というのはあくまで目的ではなくて手段だと考えております。ではそれで何を達成するのかという点で4つぐらい考えられるわけですけども、一つはDPF事業者がユーザーや市民社会に対するlegitimacyを醸成する役割や、あるいは、有名な裁判官の言葉ですけども、日光は最高の消毒液であるというようなことが情報公開とかの文脈でよく言われるわけなので、透明性を高めることによって内々で行われている不誠実な対応に対する抑止効果が出るだろうというようなことが考えられます。

またもう一つ、規制を何かしら入れた場合の効果測定のため、モニタリングするためにも透明性は要りますし、あるいは、AI等の技術利用がどういうリスクをもたらすかということも実は分からないところがあるので、これもモニタリングするためには透明性が重要になってきます。

次のページをお願いいたします。ということで最後、まとめですけども、今日の報告では直接触れられませんでしたけれども、この検討会で言われている健全性確保という、その健全性というのが何を指すのかというのがはっきりしないといけないと思いますので、そこに向けてアメリカ、EUの議論を参考にしながらデジタル空間をどのように嚮導していくのか、そのための概念はまずしっかり立てる必要があるかと思えます。

つい先日、朝日新聞で北九州市立大の山本健人先生がデジタル立憲主義というヨーロッパの議論を紹介されておりましたけれども、そこでの議論が参考になるのではないかと。日本では最高裁が自由に様々な意見、知識、情報に接し、これを接種する機会というようなことを指摘していますので、これを確保することが一つポイントになると思います。また透明性

確保についてもプラットフォーム・ガバナンスとして、市場とか、いろんな圧力にプラットフォームはさらされているわけですが、それらがうまく機能しないようであれば、何かしらの法的な義務というのもしっかりと議論しなきゃいけないですし、もう一つ、透明性レポートをあげてもらったら終わりではなくて、透明性は手段ですので、その目的を達成するために外部からしっかりとモニタリングを入れることも重要だと思います。

それからアメリカの議論で参考になるのは、DPF事業者を検閲の代理人化させてはいけないので、事業者への政府機関のコンテンツ削除要請なんかに関しては、政府機関側の、政府機関サイドの透明性をいかに確保していくかが今後、非常に重要になってくるかと思っています。

私からは以上です。

【宋戸座長】 水谷先生、ありがとうございました。それでは、ただいまの水谷先生の御発表につきまして御質問、御意見のある方はチャット欄で私にお知らせいただきたいと思っています。大体10分から15分程度を予定しておりますけれども、いかがでございましょうか。

バタバタと来ましたね。まず山口先生、お願いします。

【山口構成員】 山口です。御説明いただきありがとうございました。いつもながら大変参考になる、そして勉強になる御講演でした。私からは何かもう以前にお伺いしたような内容も含めてですけれども、改めて水谷先生にお幾つかお伺いしたいんですけれども、まず、今日の透明性のお話の中で政府側の透明性の議論が出ていたと思います。米国では特にこれ、物すごく議論されているところだと思うんですけれども、日本、まだそこまで、何でしょう、その議論が活発になってないのかなとも思っているんですね。

そういった中で、この政府がプラットフォーム事業者と連携する時に、それをしっかりと透明性を持って公表するとか、そういったことを一体どうやって確保するのかという、何かそういう方法について御存じでしたら、ぜひ教えていただけると幸いです。これが1点目です。

2点目が、米国の取組いろいろある中で、日本の政策として参照になる点というのは特にどの辺りなのかというところ、水谷先生のお考え、頂けると幸いです。

3点目が、今日はあまり議論出てなかったんですけれども、偽情報を使ってもうけるとか、あるいはアテンションエコノミーの中で広告収入を得るために過激なタイトルつけたりとか、それも偽情報、流したりとかということが起こっているわけですが、こういった

広告収入がそういう悪意を持った集団に流れてしまうことが結構、問題の根底の一つにあって。それをどうやって対策すればいいのかという話で、以前もこの点は恐らくお伺いしたと思うんですが、結構何でしょう、プラットフォーム事業者が頑張っても半分も減らないようなデータも出てたと思うので、例えばGoogleが対策したら、まだ60%ぐらい残っているみたい話でしたっけ。なので、ここって回答、解がないのかもしれないですけども、もし、こうしたほうがいいんじゃないみたいなことが水谷先生の中であれば、ぜひ教えていただけると幸いです。

以上です。

【水谷構成員】 ありがとうございます。まず1点目ですけども、政府側の透明性を確保する方法としては、もちろん色々方法はあると思いますけれども、一つは法律で、こういう部分についてはきちんと透明性レポートを政府が逆に出すというのを定めるということが一つだと思います。

例えば、さっき御紹介したアメリカで提案があったソーシャルメディア検閲透明化法を参考にしてみますと、政府の職員が表現の自由とか、そうしたことに関係しそうな案件でDPF事業者と接触した場合は、それをきちんと記録として残せと。それを事後的にチェックできるようにしろ、というようなことを規律で定めるということですね。そこまでできるかどうかは別にして今、DPF事業者側が政府からの要請についての記録を透明性レポートで出してくれているんですね。だから、あれに対応するものを逆に政府側が、例えば白書とかできちんと出すだけでもすごく変わってくるので、個々のコンテンツの削除について、どういうものを行っているのかということもきちんと明示していただくための仕組みというのを、何かしら作るというのが重要なと思います。

アメリカの議論で日本で参照できそうな点というのは、なかなかこれ、難しいところがあって。私自身は、結構、アメリカを反面教師で見ているところがあるんですね。基本的に政府が規制を入れないのがアメリカでは前提になっているんですけども、さっきのアメリカ版DSAと呼んでいいかわからないですけども、プラットフォームの透明化法みたいなのがアメリカでも議論されていて、あそこでは研究者がデータにアクセスすることを結構重要視しているんですね。そこはDSAでもそういう規定が入っていますけれども、研究者がプラットフォーム側のデータにアクセスして、それこそ偽情報の影響とか、そういうものもきちんとリスク評価できる、そういう状況にするのが一つ、参照できる点かなと思います。

最後のデマネタイゼーションの話は、山口先生とずっとこの話をお互いに交互にキャッチボールさせていただいているんですけども、なかなかいい案が浮かばないんですよ。一つ単純な方法としては、デマネタイゼーションをきちんとDPF事業者側に法的に義務づける方法があるわけですけども、DPF事業者のデマネタイゼーションもモデレーションの一種なので、それには限界があると思います。また基本的にコンテンツ自体を直接制限するわけではなくてお金を止めるというだけだから間接的にはだけども、直接的な表現規制とは違うという議論はあり得るわけですけども、果たして法規制として国家がそこまで介入していいのかというのが、まだ私も憲法的に検討しきれてないところがあるので、今後の宿題としてそういう規制が憲法上可能なのかというのを考えていきたいなと思っています。

私から以上です。

【山口構成員】 ありがとうございます。まさに何でしょう、削除までいなくても、例えば収益を奪うという行為ですよ。こういったことも私はすごく有効なんだろうなと思いつつも、例えば動画共有サービスとか、そういったところで収益化をいきなり止められて、それで生活できなくなるみたいなことというのは十分起こり得るわけですね。そうすると、かなり巨大なプラットフォーム事業者が1人の生命を結構握っているみたいな状態にもなるので、私もその辺り、すごく最近考えているところです。

ありがとうございます。では、お戻しいたします。ありがとうございました。

【宍戸座長】 ありがとうございます。それでは越前先生、お願いします。

【越前構成員】 越前でございます。私のような技術分野の人間にも分かりやすい御発表、どうもありがとうございます。私から2点ほど。まず、透明性の質問ですけども、米国でもAIにおけるコンテンツ・モデレーションというのは極めて重要だと理解いたしました。そこで透明性ですね。AIに関するコンテンツ・モデレーションをした際に、私の講演でも少し問題提起したんですけども、例えばAIを使ったモデレーションについてどの程度まで開示するのが妥当なのかという議論はされていますかというのが1点でございます。

もう1点が真贋判定する場合に、誤判定というのは必ずあることだと先生も仰っていましたが、その際の責任の所在ですね。例えば今後、選挙等でクリティカルな情報が流れた時にプラットフォームがこれ、間違っていますよみたいな情報を誤って流した際の責任の所在、そういったところに関して興味がございます、その2点について御知見を共有いただければありがたいと思います。

以上でございます。

【水谷構成員】 ありがとうございます。私も透明性と一口にしているけれども、どこまで開示すれば十分といえるのかというのは、これは多分、論文とか読んでいてもいろんなレベルで議論されているところがあるんですよね。あくまで私個人が考えているところですが、特にAIがどの程度、プラットフォームのモデレーションに食い込んでいるのかというのをきちんともっと明示していただく。

恐らくコンテンツの種類によってAIが適切に使える、使えないといった差があると思うんですね。例えば、難しいのはヘイトスピーチだと思うんですよね。ヘイトスピーチは国によって文脈が全然違う。ある国ではヘイトスピーチになるものが、ある国では全然ヘイトスピーチとして通じないというようなことがあるので、その問題がおそらくある。です。まず、どこまでAIが入っているのかというところ。もう一つは、2点目の責任の所在と関連しますが、絶対エラーは一定程度、出ているはずなので、このエラー率ですよね。どれぐらいエラーが出てしまっているのか。AIを入れたけど、特定の分野ではむしろエラー率が高いといったことが分かれば、どこに課題があるかわかるので、偽陰性、偽陽性がどの程度、出てきているのかというのをきちんとか明瞭に開示していただくことが、ポイントなんじゃないのかとは思っています。

もう一つは、透明性っていうと全ての人に公開しろということが前提になっているようにも思ってしまうわけですが、そうじゃなくて、まさにアメリカのPATAは資格を持った研究者にのみアクセスさせるみたいな、そういうことを前提にもしていたりしますので、透明性で出す情報でアクセスできる人たちというのの範囲を限定しながらやる方法もあると思いますので、その辺を複合的に考えていくべきと考えています。

2点目の誤判定の責任の所在ですが、まずコンテンツ・モデレーションの仕組みは、先ほどもお話しした通り、だいぶシステム的な仕組みになっていますので、個別のエラーについて事後的に責任追求するというのが、非常に難しいわけですね。そのうえで、DPF事業者の場合は、コンテンツ・モデレーションの部門と経営部門というのが基本的には報道機関のように分離していない、ピラミッド型でモデレーションの部門、トラスト&セーフティ部門等がある。つまり報道機関で言われるような編集と経営の分離といったエシカルな議論を基本的には採っていないわけです。しかも、モデレーション実務の多くを外注しているといった側面もあり、これがモデレーションの責任の所在を外からますますわかりにくくさせていると思います。ですので、責任の所在を議論するためにもコンテンツ・モデレーシ

ョンのルール形成とか、コンテンツ管理のオペレーションに関する政策といった点に関する年次計画みたいなものをまずは示してもらった必要があるのではないか。そこにどういう人たちが関わっていて、どのようなプロセスでできているのかという、ここも結局、透明性の話になるんですけども、責任所在をはっきりするためにも透明性の確保が重要になってくるかなと思います。

私からは以上です。

**【越前構成員】** 大変参考になりました。AIの透明性の観点ですが、全て公開してしまうと、それを迂回して攻撃する攻撃者も考えられるので、そのバランスをどう取るかというのは結構、重要な課題かと思いました。どうもありがとうございました。

**【宍戸座長】** ありがとうございます。だんだん時間が押してきていますので、簡潔にお願いをいたします。

落合先生、お願いします。

**【落合構成員】** どうも御説明ありがとうございます。非常に、アメリカの貴重な御報告をいただきましたが、大変参考になりました。

今回御説明いただいた中で、アメリカの議論の拝見をしていて、いろいろな主体がそれぞれ透明性を持っていくことについて、分け隔てなくといいますか、何らかのアクションをするものについては、それぞれ透明性を持っていくというのが一つ、意味がある対応なのではないかという示唆もあるようにと受け止めた部分もありました。日本の中でもそういう意味では情報発信者ですとか、情報の媒介者、または、それに対して影響を与え得る政府と、それぞれの主体が透明性を高めていくことも一つ大事なのではないかとも思いましたが、どのようにお考えになりますか、というのが私からの質問です。

**【水谷構成員】** ありがとうございます。まさに、落合先生がおっしゃるとおりだと思います。私の研究では、透明性の確保についてDPF事業者が主軸になっておりますけれども、今日のお話だと政府側も当然必要ですし、そういうコンテンツを生成する側ですよね。コンテンツを生成する側がどういうプロセスで自分たちがコンテンツを作っているのかと、こういうところにも透明性、もっと言うと説明責任ですかね。自分たちのコンテンツの生成プロセスの説明責任が、まさに生成系AI使い始めると重要になってくるかなと思っております。

私から以上です。

**【落合構成員】** ありがとうございます。

【宍戸座長】 ありがとうございます。それでは石井先生、お願いします。

【石井構成員】 大変勉強になる御報告ありがとうございました。私からは、1点目、表現の自由の議論の仕方と申しますか、どこに帰属することを前提に議論されているのかという点を確認させていただこうと思います。

先生の御説明の中に、プラットフォーム事業者によるコンテンツ・モデレーションの自由というような御説明があったように思いますが、コンテンツを生成してそれを乗せる人に権利が帰属する、そういう表現の自由ではなくて、プラットフォーム事業者自身もコンテンツ・モデレーションを行うかどうかのところに、基本的な自由があり、それはまさに表現の自由な話なのか、というところの確認が1点目です。そういう議論がアメリカでなされているのかということでしょうか。

それから簡単に2点だけ。2点目はコンテンツ関係の話ですが、これはアメリカでの競争法の規制は基本的には及びにくいという議論がされているのか、そういう前提でのお話になっているのかという質問です。

3点目は、透明性についてほかの先生方から御質問があったところにかぶりますので簡単な確認ですが、透明性を高める趣旨は個別のユーザーに届くかどうかではなく、抽象的に社会全体に向けて情報を発信することによって責任の所在を明確にし、信頼性を高めていくという趣旨でしょうか。個別のユーザーにどう届くかではなく、抽象的なレベルで透明性の仕組みを入れることによって、社会全体の信頼性を高めていくという文脈の議論という理解でよろしいですか。

質問は、その3点になります。よろしくお願いします。

【水谷構成員】 ありがとうございます。まず、1点目ですけど、コンテンツ・モデレーションの自由という言葉が安易に使ってしまったんですが、少し詳細に言いますとアメリカではDPF事業者の行うコンテンツ・モデレーションを、私自身はイコールで結ぶのはあまり賛成ではないんですけども、報道機関の編集の自由、編集判断の自由と重ねて議論している論稿や下級審の判例などがあったりするんですね。

また逆に、それに対しては、プラットフォームをコモンキャリアと位置づけて、だから編集判断はしてはならないという議論をしている下級審判断もあって、今、ここが対立したりしています。また研究者の中でもコンテンツ・モデレーションを表現の自由の観点からどう位置付けるか、様々に議論されているので、今後の推移を見守らなければいけないと思っています。ただ少なくとも修正1条がDPF事業者がコンテンツをモデレーションする行為

に関係してきそうだというのはアメリカの現状ではあります。

2点目のコンテンツ・カルテルと競争法の議論ですけれども、競争法の話とどこまではつきりリンクしているかは私も追い切れていないところですが、ただ、ソーシャルメディアの企業が今、寡占状態にある、それによりカルテル的な存在になりつつあるというところは確実に議論の背景にはあるように思います。

3点目ですけれども、これは個人・個別のユーザーよりも社会全体に発信する透明性が特に重要というのは、私の今日の発表の趣旨はそういうつもりでしたので、まさに石井先生がおっしゃられたとおりです。

私からは、すいません、手短ですけど以上です。

【石井構成員】 ありがとうございます。大変よく分かりました。

【宍戸座長】 ありがとうございます。それでは森先生、お願いします。

【森構成員】 ありがとうございます。大変勉強になる御発表ありがとうございました。

私も2点お伺いしたいと思っております、1点目は先ほどの裁判で、プラットフォームはコンテンツ・モデレーションしてはいけないんだという判決というものですけれども、何かアメリカは表現の自由を非常に重視する、そしてそれに対する規制を恐れる、いろんな規制を恐れることによって反面教師といいますか、危険なところにいるのではないかと私も感じているところですが、その下級審判決、モデレーションしちゃいけないというのがケンブリッジアナリティカの事件よりも前のことなのか、それとも後のことなのかということが、もし分かりましたら教えていただきたいと思います。これが1点目です。

それから2点目は、山口先生の御質問と同じですけれども、私も一定の広告について、それを非収益化する、さらには禁止すべきなんではないかと考えているわけですが、

一つは表現の自由と言っても、広く皆さんの前にパブリックに示される表現でなければ、それはその個人に対するメッセージとして見せられた場合には、検証の不可能性によって個人がうっかり信じてしまう、「それは間違っているよ」と誰も教えてくれないというようなこともありますし、また他方で、広告ですのでお金があればどんどん行けてしまうと、幾らでも出せてしまうこともありますので、そういうことを考えると政治的なメッセージをマイクロターゲティングで出すようなことは禁止されても憲法上、問題がないんじゃないかなという気もしております。

山本先生は公選法の戸別訪問の禁止とのアナロジーを使っておられたと思いますけれども、何か私もここまで弊害が大きくなってしまうと、何か、それでも表現の自由だから、そ

のかつての表現とは随分表現の形が変わってきているかなとは思いますが、その辺りについて、山口先生の御質問と重複するんですけども御意見伺えればと思います。よろしく申し上げます。

**【水谷構成員】** ありがとうございます。まず、1点目ですけども、このプラットフォームのコンテンツ・モデレーションの規制に関する裁判も含めて、この手の議論というのは本当につい最近ですね。ですからケンブリッジ・アナリティカの事件よりも後です。むしろケンブリッジ・アナリティカ事件以降、さっき黄金時代と言いましたけれども、まさにDPF事業者はコンテンツ・モデレーションをかなり強化してきたわけですね。フェイクニュースを含めて。それに対する反発として出てきているのではないかというのが私の見立てです。つまり、保守派がヘイトスピーチとかフェイク、偽情報で消されることが多いんだと思います。恐らく。トランプさんのツイッター・アカウントの凍結、デプラットフォームがかなり保守派の間では緊張感を高めることになったというのは報道レベルとかでも見ますので、その流れかなという気がしています。

2点目ですけども、まずデマネタイゼーションの話とマイクロターゲティングの話に分けて考えなきゃいけないのではないかと考えています。デマネタイゼーションの議論については、先ほどの山口先生とのお話に加えて、参考にしようと考えているものがありまして、アメリカでサムの子法という法律が、表現の自由との関連で議論されたことがあります。これは、犯罪告白本で収益を上げることへの規制だったわけですが、これが修正一条上の議論になっています。そのあたりをアナロジーで捉えて今後議論していかなくちゃいけないのかなと思っています。

もう一つのマイクロターゲティングに関しては弊害についての議論ももちろん存じておりますし、まさにあれはより広告を見させることが背景にあり、広告に関しては表現の自由としても規制が可能ですので、そこをベースに規制を考えるということはあると思う一方で、マイクロターゲティングが、政治的なマッチングとしての側面をもつという点も無視できないと思うんですね。つまり自分の考えと近い政治候補者とか、そういうところとどうやってつながっていくかという民主政治の活性化の問題とも、ターゲティングの話は関連してくると思うので、弊害とメリットの部分と両方見据えながら規制の枠組みを議論していったほうが良いのではないかというのが、現状の私の考えです。

以上です。

**【森構成員】** ありがとうございます。どちらも大変よく分かりました。前半の判決が

最近のことであるというのは驚きを感じましたけれども、いずれにしましてもよく分かりました。ありがとうございました。

【宍戸座長】      ありがとうございます。それでは、まだまだ議論されたいところもあるかと思いますが、このラウンドはここまでとさせていただきたいと思います。水谷先生、ありがとうございました。

それでは続きまして、お待たせしました、生貝構成員から20分程度で御発表いただければと思います。よろしくお願いします。

【生貝構成員】      ありがとうございます。冒頭出席できず、失礼いたしました。お時間もあるかと思うので、少し手短にお話をさせていただきます。今、水谷先生のお話の中でも度々出てまいりました、EUデジタルサービス法を中心に、特に、わけてもこの偽・誤情報対策というところに対してEUがどのように、言ってみれば法に基づくアプローチをとっているのかということを中心に御紹介したいと思います。

それとまず、デジタルサービス法そのものについては、皆さんももう既にいろいろな機会でご覧になられているかもしれません。もともと、日本のプロ責的な規定を中心に置いていた電子商取引指令をもとに、違法・有害情報に対するプラットフォームの受動的な責任と、そしてそれからその社会的影響力等を鑑みた、よりポジティブなレスポンシビリティ、そして透明性等の在り方を全面的にアップデートするものであるといえるだろう。

そして、非常に様々規律あるわけですが、わけてもディスインフォメーション等の舞台になってしまいやすい、この特に巨大なデジタルプラットフォーム及びオンライン検索エンジン、これをVLOPとVLOSEと略しますが、第3章でそれらに対して偽・誤情報対策を含む、かなり他の事業者さんたちよりも追加的な義務を課するような作りになっている。

既に17VLOPと2VLOSEが指定されており、既に特に巨大な部分に関わるものは8月から適用開始されていて、来年の2月には全面適用されるようなスケジュールで、これ自体が非常に巨大な法律であるんですが、特にディスインフォメーション等々関わるところではこの3点、コンテンツ・モデレーションの話、それからデータ保護とプロファイリングの話、そしてVLOPの偽情報に直接関わる話といったようなところで簡単に分けて御説明いたします。

その前にデジタルサービス法は、まず、媒介サービスを導管、キャッシング、ホスティングと3つに電子商取引指令と同じように分けた上で、その中でコンテンツを預かるサービ

に追加的な義務が課されて、その中で特にコンテンツを配布するようなサービスはOPという形でまた義務が増えてきて、一番大きいものは4,500万人以上のユーザー数ということで、EU域内の10%以上の人口といったところですね、まさにこの後、御紹介するシステムリスクの評価と軽減といったようなことの義務等々がかかってくる。他にも取引OPに関わる特別の規律なんかもあつたりしますけれども、それは今日は割愛いたします。VLOSEは、ほぼVLOPと同様の義務がかかると言っているかと思いますが。

まず今日、度々お話にも出てまいりましたコンテンツ・モデレーション、このことについて端的に申しますと少なくとも違法コンテンツではないもの、有害情報ですとかディスインフォメーション等に対して、少なくともこの部分では積極的に削除せよということはない。何を言っているかということ、透明性と、そしてそれによって影響を受ける利用者等の救済という形でございますね。

まずDSA、コンテンツ・モデレーションというものに対して恐らく初めて明確な法的定義を与えておまして、上でございますとおり、まさにさっきお話あつたような、まずは違法情報の削除です。あるいは利用規約に適合しない情報、これがまさしく有害情報ですとか、ディスインフォメーションが広く含まれる。しかし、その定義はあくまでデジタルプラットフォームの側がする形になっている。

そうした時にそれを識別したり、まさに削除のほかにも非常にいろんな手段がある。降格、まさにお話があつた収益不能化・デマネタイゼーション、アクセス不能化、可視性に影響を与える措置など、非常に広くとっているわけですね。この中ですと、例えばしばしば話題になるようなシャドーバンのようなものも含まれてはくるわけでございます。

まずはそれを徹底的に透明化する、利用規約にこういうコンテンツ・モデレーションをどうやっているのか、どんな人員体制でどんなモデレーションをアルゴリズムを使ってやっているのかといったことを記載する。特に大きいこれらの方々は各加盟国、少なくとも27プラス、EEAの言語で提供しなければならない。そして、それをどのように実施したかということについて、少なくとも年に1回、VLOPは年に2回、透明性レポートというのを提示するわけでございます。

ここでは、まさしく政府の透明性という意味だと各国政府からどんなコンテンツ削除の要請があつたり、情報請求の要請があつたのか、あるいは違法・規約違反別の対応件数、それから大体対応1件あたりに中央値としてどのぐらいかかつたのか。それからコンテンツ・モデレーションの人員では、大体何語ができる人がどのくらいいて、どんなトレーニングを

しているのか。または、先ほどお話あった自動処理のエラー率指標と、それからそのエラーがあった時にどういうセーフガードをとっているのかといったようなことを事細かに出さないといけないということになっている。

本物を見ていただいたほうが分かりやすいと思うんですけども、これは例えばXさん、まさに国ごとに例えば大体このくらいの時間が中央値としてかかっている。まだ、データの欠損が非常に多いですけども、あるいは大体エラー率は今、分かっている範囲ではこのくらいである、ちゃんと反論を受け付けられる体制を作っているのですとか。

特に、さらにこれ、巨大プラットフォームさん皆さん出されているんですけども、ディスプレイフォメーション対策等で重要なのは、この何語ができるモデレーターをどのくらい抱えているかといったようなことです。Xさんは、言語の偏りが大きいといったようなことも言われているところですけども。例えばオランダ語ができる人は1人確保しているのですとか、イタリア語ができる人は2人確保しているのですとか、そういったことがしっかりと透明に公開をされることになっているわけでございます。

もう一つの透明性の側面として面白いのは、コンテンツが削除されたりしますよね。それはしばしばまさに今、大体自動的にやるので間違えることもあるといった時に、コンテンツのモデレーションをやったら、その影響を受けた情報発信者等にちゃんと一々、1件1件理由の通知をしないといけないことになっております。そしてその通知を受ける、そうすると初めて反論の確保、機会が確保されるわけですよ。GDPRのプロファイリングのアナロジーで言えば、まさに機械の決定だけに服さずに人間の関与を求める権利と言っても良いかもしれない。

こういった苦情があった時は、これには別のルートで個人で削除してくれと言ったのに、例えば誹謗中傷等、削除してくれないといったようなものの苦情処理も、ちゃんと理由を通知した上でこういうシステムの対象にしないとけないんですけど、ちゃんと公平な判断をしないとけない。それでも納得いかなかったら、例えばADRを使うことができるといったようなことを定めていたりするわけです。

このデータベースもすごく面白くて、例えば今、既に1か月ぐらいで8億件ぐらい理由通知データベース集まっていますけれども、この中で種類ごとに生データが当然匿名化された形で出てくるわけです。例えばミスインフォメーションというタグで検索をすると、タグがまだ整理されてないので100万件という比較的、数は少ないですけども出てきます。

例えば、T i k T o k とピンタレストが今のところ上に出てくる。その中身を見てみると、

これはGoogleの無理やり翻訳でございますけれども、例えばこういう理由で利用規約のこれに違反するから削除されました、利用規約、Terms of Serviceの根拠はミスインフォメーションであると。コンテンツタイプはビデオである。そしてカテゴリーは市民の議論や選挙への影響だといったようなこと。誰かから言われたんじゃないで自分で探して消した。そして意思決定は人間がやったのか、それともアルゴリズムがやったのかというと、これはフリー・オートメテッドだといったような、こういうデータが非常に事細かく出てきて、こういったデータ一つ一つ確認できるだけではなくて、今、まさにこういうビッグデータを使ったコンテンツ・モデレーションの全体的な研究みたいなのもだんだん進んでいたりするみたいでございます。

そうした時に、要するに利用規約が大事である、なんだけどコロコロ変わる、フォローできなければ後で確認可能性がないということで、利用規約が変わっていくのをちゃんと時系列でもしっかりと把握した上で、変遷も後でチェックできるようなデータベースも作っている。こういうテクノロジーに基づく、ある種の透明性の確保というのもいろんな形で重要なんだろうなと思うのが1個目です。

それからデータ保護、プロファイリング、これは先ほど森先生からもございましたけれども、ケンブリッジアナリティカ事件以降、いろいろな広告ですとか、あるいはレコメンド、特にそうですね、ターゲティング広告なんかが使われたりもしたみたいですがけれども、そのプロファイリングに基づくものが問題だということで、2つ目の柱としてこういうものも偽・誤情報等に深く関わるのかなと思います。

EUではGDPRでプロファイリング規制を置いているのは御承知のとおりですがけれども、それに加えてオンラインプラットフォームなどはターゲティング広告のパラメータを出さないといけない、レコメンダーもパラメータを出さないといけない。正確に言うと、コンテンツ・モデレーションの定義の中にこの2つは入らないんですよ。同じに扱うのではなくて、ユーザー投稿コンテンツへの対応というところとは違うので。しかし、ここまで視野に入れた形での規律というものを考えていくことが包括的な対応にはなるということなのだろう。センシティブデータを使ったプロファイリング広告禁止、青少年保護と未成年個人データのプロファイリング広告禁止ということで、青少年の脆弱性に特に配慮したような規定なんかも置かれている。これが②のカテゴリーです。

そして、特に直接的に偽・誤情報対策に関わる場所では、まさに先ほども言及ありましたVLOP、VLOSEの偽・誤情報を含むシステムリスクへの対応です。今までの申

上げたところは、どちらかという一つ一つのミクロなレベルでの対応の透明性といったようなものを中心としている。それを集計した透明性レポートというのも重要ですけども、まさにシステミックなリスクが問題ですよ。それはしかし、ディスインフォメーションだけに限らないが、しかし当然、市民言説ですとか選挙への影響というのがデジタルサービス法の立法経緯としては非常に大きくある。その他いろいろなリスクを自ら評価し、そして、それをちゃんと緩和する義務を取らないといけないという、非常に重要な包括的な規律が置かれているわけです。

しかし、そういうリスク評価と軽減ってどうすればいいのかというと、なかなか包括的過ぎる、抽象的過ぎるというのがありますので、まさに、これは欧州委員会がある種、非常に強く後押しする形で形成するコードオブコンタクトですとか、あと、もう一つはこの36条の上に戻るんですけども、特にディスインフォメーションは、どちらかという平時というよりは、突然何かが起こったことが問題になったりすることわけですよ。自然災害ですとか武力紛争でありますとか。そうした時には緊急で、これはかなりデジタルプラットフォームの裁量というのを重視した形にはなっているようですが、緊急対応、危機対応メカニズムというのが置かれていて、こういう行動をこんな形で取ってくださいという要請を、まさに法に基づいて出すということになっている。その部分はまた別途、危機プロトコルという形で決められていたりするところです。

さらに、これらがちゃんと守られているかどうか、なかなか公開できる情報と公開できない情報もあるし、しっかり中に入ってデータを見ないと分からないということで、必ず独立した監査を徹底的に受けて、その修正要求にどう対応したかといったようなところもまさしく義務として入っているというわけであります。そして、まさにここの中でも、さっきもアメリカの文脈でお話のあったデータアクセス提供義務といったようなことも、外部の、これは信頼できる研究者等に限った形で認めたりということもやっているわけです。

そうした時に、このデジタルサービス法の行動規範がどういうものなのかといったようなものかについての、御参考ですが、まさにリスク低減義務というものを行動規範で具体化していく、法と行動規範の関係のさせ方が非常に工夫された形になっていて面白いと思います。招請を適切な説明なしに拒否した場合ですとか、あるいは、ディスインフォメーション対応としてこういうリスク評価をやらないといけないような説明もあつたりします。

ここはまた、危機対応メカニズムというのがどうやって機能するかについての、デジタルサービス法の前文等を引用、仮訳しておりますので御覧いただければと思います。すごく限

られた状況というふうに限定はしているんだけど、しかし、例えばコンテンツ・モデレーションのリソースはちゃんと増やしてくださいねとか、利用契約ですとかアルゴリズム、システムを調整するようなことなんかも検討してくださいといったようなことが含まれているということです。

それからもう一つ、A Iの話に触れておきたいと思います。最近A I法案がトリログで、つい先週ですか、合意をして非常に関心を集めているところですけども、A I法案を含めて今までのA I規制論って大体2つの問題意識に分けられるなと思っています。ビフォア生成A Iって、要するに焦点リスクは製品安全とプロファイリングの2つだったんですよね。それが情報を生成するA Iになって、本格的に情報法の領域に関わるようになってきた。それは今、デジタル情報法で問題にしてきたような偽・誤情報への対策、情報環境全般への影響ということ。それがまた、直接的に国家安全保障の問題になってきたりするわけですけども、言ってみればかなり議論の内容というのは、デジタルプラットフォームに関して議論されてきたものと近づいてきている部分が、少なくともアフター生成A Iについてはあるのかなと思います。

そうした時に、このA I法案とデジタルサービス法の関連性をちょっとだけ補足しておきたいんですけども、まずA I法の中でディープフェイクの場合はラベルづけをしないといけないような義務が含まれる予定であることは御存じかと思うんですけど、それってまだ今後、正式に成立して施行されるまで大分時間がかかりますよね。

ですけども、デジタルサービス法ができる前から、適用開始される前から、まさにこのデジタルサービス法の共同規制における行動規範になることを前提にして偽情報行動規範というものを作ってきた。去年6月の段階でまさにA Iが生成したり、操作、マニピレートしたコンテンツに対しては、これは注意喚起ですとか、そういうものに対抗するための措置というのを様々ならないといけないということです。ディープフェイクや偽情報は結局、この文脈だと生成するA Iがあって流通するプラットフォームがあり、両方から考えていけないといけない。まず、法律がある後者からやったということだと思います。

あとそれから、これもいろいろなところで報じられているとおり、A I法案、最初ハイリスクと4分類といったような、2021年の最初の段階では生成A Iを意識しない形で法案を作っていたんですけども、去年末から急遽、生成A Iが重要になってきたということで、それとはまた別に生成A Iなどの、彼らは最終的に汎用A I、general purpose AI、GPAIと呼ぶようにしたわけですが、その内容が大きく含まれることにな

った。

これは、テキスト自体はまだ公開されていないんですけれども、トリログ合意の直後に欧州委員会が詳しいQ&Aを出してしまっていて、まず、全てのそういった汎用AIモデルに課せられるような透明性の義務というものに加えて、わけても10の25乗フロップ以上の計算能力を持って計算されるような非常に大きなモデルは、取りあえずはですけれどもシステムックリスクがある。そのシステムックリスクを持つモデルのプロバイダはリスクの評価と軽減、まさにシステムックリスクの評価と軽減、その他、最近AI規制について言われるようなことをやらないといけないということで、まさにデジタルサービス法のVLOP、VLOSEに課せられている義務というのをスライドさせてきた。まさにAI議論のプラットフォーム化というところの一つの大きな象徴かなと思います。

ざっくりの御説明ですけれども、日本でこれからどういうことを考えていくべきなのかということを経つかの論点という形で挙げています。あくまで論点ですけれども、上の3つはデジタルサービス法そのままですけれども、モデレーションの透明性と救済というのは果たしてどのような情報空間が健全であり、望ましいのかということを経社会で全体で議論する、それに対して個人が異議を申し立てる、そういうのを大前提と考えるべきであろう。

そしてデータ保護、プロファイリング、このことというのはまさにモデレーションと同じか、場合によってはそれ以上に重要なルールというものにもなってくるだろう。そしてVLOP、VLOSEに関しては日本の立法技術上、こういったような仕組みがどの程度可能かというのは難しいんですが、しかしこの個別のディスインフォメーションを削除せよというのではなく、まさにシステムックリスクを管理してもらい、軽減してもらい、そして、その具体的な形は行動規範等を通じて具体化していくようなことは、今の時代のテクノロジー分野のルールで不可欠なのかなと僕は感じているところです。同じようなことが多分、汎用・生成AI、特に大きなものについては必要だという議論も出てきている。

そして最後に、これは水谷先生の御発表とのある種のリンクというところなんですけれども、まさにお話にあったとおり、アメリカは修正1条と、それから特に巨大なプラットフォームですとかAIの会社さんが地元にいるということもあるので、非公式なお願いによってフェイクニュースを、いろいろと戦争のときですとか、消してもらったりすることが常態化しやすい。あるいは、することができる。そうした時に、しかしそれは、公法的な民主的なコントロールや透明性というのがなかなか及ばなくなってしまう。そうした時に、そのルールづくりというのは極めて様々論点ありますけれども、非公式な要請と法に基づく要請とい

うもののどちらが望ましいのかといったようなことは、まさに様々な観点から考えていく必要があるんだろうなと思います。

私からは以上でございます。御清聴ありがとうございました。

【宍戸座長】 生貝先生、本当にありがとうございました。

それでは御質問、御意見のある方はチャット欄でお知らせください。時間が限られていますので、簡潔にお願いいたします。

まずは落合先生、お願いします。

【落合構成員】 生貝先生、どうもありがとうございます。アメリカ以上にヨーロッパのほうが、見ていて日本では参考にしやすいのかなとも思いながら伺っておりました。

私も幾つかありまして、一つが本日、DSAについて御説明いただきましたが、メディアの関係で欧州報道自由法といったようなものもあります。その中でメディアへの多元性、独立性の確保であったりですとか、先ほど一つ前の議論で米国において議論されていたところとも若干、オーバーラップするような内容も議論されているところがあると思います。これについても、位置づけや、重要と思われるような点をよろしければお話いただけないでしょうか、というのが1つ目です。

2つ目としましては、直近で、12月12日にディフェンス・オブ・デモクラシーパッケージというのが出ております。欧州民主主義行動計画というのに基づいてDSAなども議論されている部分もあったかと思いますが、直近でそういう比較的大きい補完の動きなどがあつたりしたとも受け止めております。これの意味合いだったりですとか、それと関連して2022年には実践規範の強化というのがされていて、その中でコンテンツ・モデレーションなどが強化されている部分があったと思います。この辺りのDSAや欧州の対策も比較的早くアップデートされている部分があるかと思いますが、その辺りの経緯や、直近での視点を教えていただければと思いました。

私からは以上です。

【生貝構成員】 ありがとうございます。まず一つは、メディア自由法については、一つはコンテンツ・モデレーションとまさに重要な情報を発信するメディアの関係というものをどう考えていくか。これ、取り上げませんでしたけれども非常に重要であることに加えて、コンテンツ・モデレーションだけに限らず、いかに信頼できる、まさにジャーナリズム、情報というものを多元的な形で欧州全体でしっかり確保していくのか。今日の文脈からは私自身は触れなかったんですけども、そのことというのは、まさに健全な情報空間といった

ようなものの実現の中で非常に重要だろうなと思っているところでもあります。

そして2つ目として、欧州民主主義行動計画等との関わりという意味でいいますと、まず一つは、特に来年、欧州議会の選挙があるんですよね。その防衛パッケージといったようなことで、特に第三国からのリプレゼンティティブや介入ですとか、そういったところに非常に力点を改めて置いた行動を加盟国にも促す部分が大きいのかなと思います。

それから今日、申し上げられなかったですけども、政治ターゲティング広告規則というものもそろそろ終盤に差しかかっているところで、政治と言っても全然選挙だけではなく、まさに特定の政策 이슈に関わるような活動なんかも含めた規律や透明化といったようなものも同時並行で作られていまして、まさに水平的なDSAと、それから政治・民主主義といったようなところのバーチカルな問題というものも併せて出てくるのかなと思います。

最後に、行動規範のアップデートに関しては、もう本当に頻繁にアップデートできるのが行動規範の良いところですので、わけても例えばディープフェイクの問題は、あの頃に改めて初めて出てきたところもあったんで、そういうものも含めたアップデートかなという認識しております。

取りあえず以上です。

**【落合構成員】** ありがとうございます。

**【宍戸座長】** ありがとうございます。それでは水谷先生、お願いします。

**【水谷構成員】** ありがとうございます。生貝先生。私からは2点ですね。1点は、DSAの背景にある基本的価値と申しますか、そういった点についてお伺いしたいと思います。最近の論文で、『ブリュッセル効果』を書いたアニュ・ブラッドフォードが、EUのDSAとか、AIAも含むかもしれませんが、EUのデジタル憲法だと評しているのを見ましたので、そこについてお伺いしたいと思います。

2点目は、異議申立ての仕組みがDSAには入っていて結構、目玉だと思うんですけども、この個人の異議申立てという仕組みというのは、法律家からするとすごくなじみ深いわけですよね。事後的な救済という意味で。ただどさっき、データベース自体がもう8億あるという話だったわけなんで、あの量的な観点から考えても、それを利用するというのは非常に限られてくるユーザーになってくると思うんですね。それと関連して、個別事案の事後的な救済は、例えばコンテンツ・モデレーションに用いられているアルゴリズムのバイアスの問題だとか、根っこの部分に関係してくる問題には対処し切れないのではないかなと思うん

ですけれども、そこの兼ね合いというかはDSAとかで考えられているのかどうかという点をお伺いしたいと思います。

以上です。

**【生貝構成員】** ありがとうございます。そうですね。まず一つは、基本的価値という意味ですと、まさしく先生方が御議論されているようなデジタル立憲主義は、憲法的価値というものをデジタル空間でも私的なプラットフォームも含めて実現していくかというところは、デジタルサービス法の中でも非常に出てきているのを僕自身も感じているところです。

わけても、触れませんでしたけど非常に興味深いのが、利用規約に基づくコンテンツ・モデレーションポリシーの実施に当たっては基本権に配慮しなければならないといったような、大変真正面のことなども含まれていたりして、まさしく憲法的価値というものとデジタル空間を改めて考えるきっかけにもなるのかなと思います。

それから、アルゴリズムのバイアスといったような意味ですと、まさにこれが多分システムリスクとか、そういうものの軽減措置の中にも今後、恐らく様々な形で具体的に組み込まれてくるのだと思います。システムリスクの対象って非常に広い、基本権ですとかジェンダーですとか、そういったような問題も含めてということなので、まさにちゃんとそれに合わせた形でデータのバイアス、アルゴリズムのバイアスみたいなのもちゃんと調整していかないといけないということが、今後その条項の具体化の中でいろいろ出てくるんだろうなと思います。

取りあえず以上でございます。

**【宍戸座長】** 水谷先生、よろしいですか。次は山口先生、お願いします。

**【山口構成員】** 山口です。御説明いただきありがとうございました。大変すばらしい講演で大変勉強になりました。以前も学会で生貝先生のお話をお伺いした時に本当に面白いなと思ったんですけども、先生の最後のスライドの内容にかぶるかもしれないし、また、水谷先生にした質問にすごく似ているんですけど、私の1点目が、先生がEUの対策を非常に見ておられる中で、これは日本でも導入すべき良い方向性だなというの、並びに、いや、これはリスクがあるから慎重であるべきだというようなものが、もしあれば教えていただけると幸いです。

2点目が、DSAとかってかなり強い法律だと思っているんですね。何か経済学的に考えると、市場規模が大きいからそれ、できるというようなこともあるのかなと思っている。つまり、5億人市場だからプラットフォーム事業者もそれに従うわけですよ。一方、例えば

日本というのは1億何千万人ですし、また、さらに人口が少ない、あるいは経済規模の小さい発展途上国がいっぱいある中で、プラットフォーム事業者に対してこういう法律を施行していくって、そもそも難しい国のほうが多いんじゃないかなと思っているところもあるんですけども、その辺りのところ、生貝先生、何か御意見あればぜひ教えていただけると幸いです。

以上です。

**【生貝構成員】** ありがとうございます。そうですね。全体的に特に透明性、救済について、10億人ですとか、20億人使うようなサービスについてリスク評価して緩和せよというのは、僕は社会的なレスポンスビリティとして、それがデジタルプラットフォーマーであってもGeneral purpose AIであっても、あるべきなんだろうなとは思っているのですが、まさに今日、御紹介したようなところというのは、我が国でもバランスのとれた形で前向きに考えていく価値があるのかなとは感じております。また他方で今日、比較的本当に水平な部分を御紹介したんですけども、バーチカルには例えばテロの情報に関して、当局の要請後1時間以内でちゃんと対応しなきゃいけないですとか、またCSAM系ですと非常に難しい問題であるだけに、強い措置というのもEUレベル及び各国レベルで行われている。まさにバーチカルにどうしていくかといったような問題は一つ一つ、その問題の深刻さとの兼ね合いでしっかり考えていかなければいけない部分が多いなと思います。今日はあくまで水平な部分を中心にお話したということです。

2点目、答えましたでしょうか。

**【山口構成員】** 2点目が、市場規模の大きいユーザーというのができるという。

**【生貝構成員】** 失礼しました。全くおっしゃるとおりです。プラットフォーマーは多くの国や地域を対象にしておりますので、基本的には何か全く違ったものを作ろうとするのはプラットフォーマーさんの対応としても大変ですし、例えば、どんなことを透明化してほしいかみたいなことは、G7ですとかOECDなどでもいろいろ議論をしながら、しっかり標準化した形で我が国含めて各国、法制を考えていくことが現実的なんじゃないかなと考えるところでございます。

以上です。

**【山口構成員】** ありがとうございます。

**【宍戸座長】** ありがとうございます。それでは石井先生、お願いします。

**【石井構成員】** ありがとうございます。大変勉強になりました。私からは1点のみ、大

きな観点の質問をさせていただければと思います。

EUの今日、御紹介いただきましたDSAですとか規制法の話、御説明を伺っていますと、そのEUの法制構成全体は、偽・誤情報もそうですし、ターゲティング広告、プロファイリング、ダークパターン、それから生成AI、全て関係するところだと思いますけれども、健全な情報環境を実現するために社会の誘導を、社会が変なほうに誘導されないようにする、そういう視点を持った規制に動いているのかなというように感じたところであります。

マニピュレーションという言葉が使われていたかと思いますが、個別の個人情報保護ですとか消費者保護とか、そういうプレーヤーの話ではなくって、もっと大きな、もう一つ大きな観点の守るべき価値というものですかね。それは社会全体が何かしらのものに御誘導されないようにするという、そういう方向性の議論がなされていると理解しているのかと思っています。

仮にそうだったとして情報環境の健全性を維持、実現していく上で、日本語としてその基本的な価値といいますか、守るべき価値に対してどういうアプローチが適切なのか。その辺りについても、御説明いただいたことも含まれていると思いますが、可能な範囲で教えていただければと思います。

**【生員構成員】** ありがとうございます。そうですね。まさにおっしゃるとおり、比較的これまで、個別の救済ですとか、そういったところに焦点を当ててきたところから、まさにもうプラットフォームが社会全体に影響を与えるものになってきたところ、システム的なマクロな問題にこのデジタルサービス法というのが西側の中だとある種、初めて真正面から取り組んでいるルールになってくるのかなと思います。

そうしたときに、リスクってすごくいろいろあり得るんですよね。イノベーションが起き続けるからこそ。そうした時に、まさにマニピュレーションや操作というのは間違いなくあるリスクですし、あるいは、誰もその意思がなくてもカスケード的な現象って起こったりすることもあるし、あるいは我々一人一人の本性というのがいろいろな差別を引き起こすといったようなこともこれ、当然であろうといった時に、それをある種、アンプリファイさせないでありますとか。まさにそういった、ある種、フューチャーブルーフィングにというんでしょうか、リスクに対処しているのかなといったようなところであります。

そして我が国がどうするかというのは難しいんですけども、一つは我が国の最も重要な価値をどう実現するかといったときに、これはまず一つは前提として、議論ができる前提として透明性というものをしっかり確保するといったようなことが一つ。何が起きている

るか分からないのでは、まさに議論のしようがない。そして、もう一つの論点としてのリスクの評価と軽減という意味では、これはできる限り公表された情報に基づいて公論をして、議論をできるだけメディアを含めてしていく中で、そういったようなことをデジタルプラットフォーム様がちゃんと取り込んでいけるような、そういったプロセスというものが極めて重要だろうなと思います。少し抽象的で恐れ入ります。

【石井構成員】 ありがとうございます。大変よく分かりました。

【宍戸座長】 ありがとうございます。また例によって私の進行が悪くて時間押してしまいましたが、生貝先生、ありがとうございます。

また、非常に活発な御議論を本日、田中先生、水谷先生、生貝先生から話題提供いただき、構成員の皆様から御議論いただいたことに御礼申し上げたいと思います。自由討論を事実上、兼ねていたということで御容赦をいただければと思います。

それで議事3でございますが、全体を通じて構成員の皆様から何か御意見等ありましたらいかがでしょうか。あるいは構成員の皆様で本日の御発表等についても、何か聞き逃したということがあれば、別途、事務局までお寄せをいただければと思っております。本日のところは、これでよろしゅうございますか。ありがとうございます。

それでは、事務局より連絡事項があればお願いをいたします。

【内藤補佐】 ありがとうございます。次回の第5回会合につきましては12月25日（月）13時から15時の開催を予定してございます。議事等の詳細につきましては、別途事務局より御案内を差し上げます。

以上でございます。

【宍戸座長】 ありがとうございます。それでは、以上をもちまして、デジタル空間における情報流通の健全性確保の在り方に関する検討会の第4回会合を閉会とさせていただきます。本日も活発な御議論いただき、ありがとうございました。