参考資料４－２

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３回）

１　日時　令和５年11月27日（月）13時00分～15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、江間構成員、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ァ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

西泉大臣官房審議官、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、ただいまのとおり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３回会合を開催いたします。本日も御多忙のところ当会合に御出席を賜り、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御案内いたします。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３－１から３－３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したら、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、そちらから御覧ください。

なお、本日は、落合構成員が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に移りたいと思います。本日は、構成員のうちから江間構成員、越前構成員、それから奥村構成員から、それぞれ御発表いただきたいと考えております。そこでまず、江間構成員から２０分程度で御発表をお願いいたします。

【江間構成員】　どうもありがとうございます。それでは、今からスライドを共有させていただきます。それでは、話題提供させていただきます、東京大学の江間と申します。本日はよろしくお願いいたします。

私、１回目は欠席してしまったのですが、２回目でほかの構成員の方々のお話もお伺いさせていただき、偽情報・誤情報関係は、私は専門ではないので、もう少しより広い観点で、ガバナンスというところから、現在、今、原則ですとか様々な行動規範だったりガイドラインとかが出てきている中で、どのようにしてそちらのガイドラインをつくっていって、どのような動きがあるのかということを、外観ではございますけれども、紹介させていただいた

後に、後半では具体的にそれらをどのように確認していくのかということで、我々の東京大学の研究チームで行っておりますＡＩシステム、ＡＩサービスの監査に対しての、少し整理したものを御紹介させていただこうと考えております。よろしくお願いいたします。

タイトル、ＡＩガバナンスのＰｒｉｎｃｉｐｌｅｓ　ａｎｄ　Ｐｒａｃｔｉｃｅｓということなんですが、この研究会、総務省さんのこの会の全体的なテーマというのが、デジタル空間における情報流通の健全性ということになってございます。これをＰｒｉｎｃｉｐｌｅｓやＰｒａｃｔｉｃｅｓというところに当てはめたときに、そもそもどういう状態が健全化の議論とある程度の合意というのをしていくのがＰｒｉｎｃｉｐｌｅ、原則の議論に相当するところかなと考えました。

また、健全な状態というところに持っていくというときに、健全性を誰がどのように担保しているか、また、どのように確認するかということの方法論の議論も、同時に必要になってくるのかなと思っております。健全性、偽情報・誤情報に関しても、何をもって偽情報と誰が判断するのかというのは、恐らく非常に難しいグレーゾーンを含んでいるようなものもあるかと思いますが、このような点に関して、広くＡＩガバナンスという観点で、どのような議論が行われているのかということを、この次のスライドで紹介させていただこうと思っておりますが、実際にはこの両方が行き来をしたりするということが重要になってくるのかなと思っております。

こちらは本当に一部だけではございますが、最近ＡＩガバナンスでは動きがいろいろ速くて、いろいろなところで議論が起きてございます。このＰｒｉｎｃｉｐｌｅ、一番大きなレイヤー、ハイレベルなところでありますと、もちろん言うまでもなく、Ｇ７広島ＡＩプロセスというのが走っておりまして、そちらでＰｒｉｎｃｉｐｌｅが出ておりましたり、ＯＥＣＤが最近いろいろと定義も新しく改進したものとかも出ておりましたが、出しているものや、ユネスコですとか、国連がハイレベルアドバイザリーボードというのを開始したというのがございます。

さらに原則をどのように各国、あるいは様々な業界団体で議論していくかとなったときに、ガイドライン、ノンバインディングなものや、あるいは規制があるという意味でのレギュレーション、国による法令というところでバインディングなものというのが、２つ入り交じっているのが現状かと思っております。日本は今、ちょうど総務省さん、経産省さんで新事業者ガイドラインというのが議論をされております。この間出たアメリカの大統領令、エグゼクティブオーダーとかも基本的にはノンバインディングではあるけれども、事業者と

かいろいろなところに、あるいは関係省庁にいろいろ考えてくださいというのは要望を出しているというところでございます。

一方でバインディングで一つ、皆様がとても着目をされていらっしゃるというところが欧州のＡＩ　Ａｃｔでしたり、あまり注目はされてないんですけれども個人的に注目をしている、欧州評議会のＡＩ条約というものがございます。こういうような、ある程度原則を落とし込んでガイドラインというところに持っていくというところが一方である中で、最近は、今日もいろいろニュースも出ておりましたけれども、Ｖｏｌｕｎｔａｒｙ　ｃｏｍｍｉｔｍｅｎｔｓというところで、様々な組織に対して自主的にガイドラインとかをつくって、安全性ですとか公平性とか、そういうところに対応していこうということを促すような層が、新たに結構出てきたのかなというような感じでございます。これは多分、１年前に同じような図をつくると、このＶｏｌｕｎｔａｒｙ　ｃｏｍｍｉｔｍｅｎｔｓというのは恐らく入ってこなかったと思うんです。でも、ただ自主ガイドラインを自らつくっていったり、業界で考えていくということが非常に大事になってきているのかなと思っております。

また、もっとさらに具体的、テクニカルな点になりますと、インターナショナルなスタンダード、国際標準ですとか、あるいは国内の標準機関が定めているものということで、よく取り上げられるのはアメリカのＮＩＳＴのＲｉｓｋ　ｍａｎａｇｅｍｅｎｔ　ｆｒａｍｅｗｏｒｋというのがあったり、ＩＳＯとかＣＥＮ／ＣＥＮＥＬＥＣ、ＩＥＥＥ　ＳＡとかＩＴＵとかが、技術標準をまさに今、いろいろなレイヤーで議論がされているというところがあるかと思います。

それらとはまた、よりプラクティスレベルというところで、事例ベースみたいなところで、ＧＰＡＩをどこに入れるかすごくこの図をつくるときに悩んだのではあるんですけれども、実際にいろいろなワーキンググループが動いて、いろいろなツールですとかアセスメントについて考えていくというのを、プロジェクトベースで動いているというところで、実働に近いというところでこちらに置かせていただいております。また、ＡＩ監査ですとか認証みたいなところは、標準化などと足並みをそろえながら、実際にどのような観点で外部監査、内部監査、あるいは第三者認証みたいなことをやっていくのかということを考えていく試みがあるということで、こちらの具体的な例の近いところに置かせていただいております。

まず、ここの上のほうのレイヤーにございますような、ガイドラインですとかＶｏｌｕｎｔａｒｙ　ｃｏｍｍｉｔｍｅｎｔｓに関して具体的にどのようなお話あるのか、多くの方にとってはもうほとんど復習のようなことになっていると思うのでさらっといきますけれ

ども、実際にはどういう人たちと一緒に、今、何をテーマとして、どういうような領域で議論していくのかみたいなところが、大きくテーマとして考えるべきことかと考えております。特に、Ｆａｉｒｎｅｓｓ、Ａｃｃｏｕｎｔａｂｉｌｉｔｙ、Ｔｒａｎｓｐａｒｅｎｃｙみたいなことは、この図をつくったの自体はもう本当に２～３年前ではあるんですけれども、あまり変わってはいない、当時から重要なテーマになっていると。昨今ではセイフティセキュリティーというのが、生成ＡＩに絡めて重要な議論される項目として立ち上がってきているところもあるかなとは思っております。

このようなＰｒｉｎｃｉｐｌｅを考えていくときにも、よく最近使う、最近でもないですね、もともといろいろなところで言われていますけれども、バーティカルな、垂直的なレギュレーションをつくっていくのか、ホリゾンタルな、いろいろな領域でカバーするようなものをつくっていくのかということで、どちらかというとバーティカルな、今まで医療ですとか交通ですとか金融とかに既に規制があったり、あるいは標準があったり、議論があったりする中で、それぞれと二重三重の規制にならないようにきちんと整理をした上で、かつその領域固有の課題というのはあるわけですので、それを考慮に入れながら、規制だったりガイドライン、あるいはレギュレーションをつくっていくということが重要になってくるだろうということが言われております。

もう１点、重要な観点というのが、ＡＩというのはただ単にＡＩを開発している企業がそのままサービスを消費者に見えやすい形で提供しているというわけではなく、国や組織をまたいでいるということがあります。ＡＩ開発者がいて、ＡＩ提供者がいて、ＡＩ利用者がいるというすごくシンプルな図は、例えばＧｏｏｇｌｅですとかマイクロソフトさんとかは、マイクロソフトさんは開発者ということでオープンＡＩとかとも連携はしていますけれども、比較的、提供されているサービスがどのような人たちに使われているかという関係性が分かりやすいというような海外の企業もある一方で、日本は主にすごくサプライチェーンが長いという特徴があるかと思います。

我々が今、実際受けているアプリケーションですとかは、特にＬＬＭとかは提供しているのは全然違う企業ですけれども、実際に生成ＡＩを使っていて、そのサービス、アプリまでやっていくとなると、もう本当に長いサプライチェーンになってくるというところなので、実際に、ではどこが具体的にアルゴリズムを開発しているのか、誰がデータを提供しているのか、あるいはどういうふうにサービスが提供されているのかということは非常に複雑になってきているので、責任を、何か問題あったときに誰がどうとるのかということが非常に

難しくなってきている、そこの整理をする必要が出てきているというのが、基本的にガバナンスをめぐる一つの大きな論点かなと思います。

そのような中においては、様々、関わっている人たちを特定、ステークホルダーを特定した上で、事業者間での円滑なアジャイルなガバナンスをしていく上というような観点からすると、きちんと契約を取り交わした上で、そこの契約自体に何か寡占ですとか、あるいは社会的なパワーバランスによる不公平とかが起きていないかということは、もちろん市場が監視するということは重要ではあるんですけれども、特に透明性をもって、可能な範囲での透明性ですね、情報提供ということを互いにやっていくということが大事かなと思っております。

この間のＧ７の広島ＡＩに関しても、ＩＧＦの京都でいろいろ事業者の方をお招きして議論とかもありましたけれども、そこにＴｒａｎｓｐａｒｅｎｃｙ、透明性というのがとても非常に重要であるということが、せりふとして出てきました。一方でアカウンタビリティ、説明責任や、あるいは日本語でもう少しきちんと言うならば、説明をする責任ではなくて誰が何か問題があったときに責任をとるのかということを一応考えなければいけないということで、その点に関してはそれぞれの関係者が果たすべき役割ですとか責任というのも、このサプライチェーンの中で考えていくことが重要かなと思っておりますし、個人とか法人とかでは全てが責任を吸収できないというような場合は、公的機関が事件や事故の原因究明とか被害者救済の仕組みなどを、ある種、つくっていくということも重要になってくるでしょうし、前回の会議でもございましたように、ある程度のリテラシー向上、使う側のリテラシー向上でしたり、セーフティーネットをつくっていくというようなことも非常に重要になってくるのかなと思っております。

今、先ほどのＰｒｉｎｃｉｐｌｅからＰｒａｃｔｉｃｅの図の中でも、いろいろなレイヤーで議論が行われていますというような紹介をしましたけれども、法律とかガイドラインだけが規律といいますか、人々や技術やＡＩを提供している組織に要求をするという方法ではないと思っています。国家による強制、エンフォースがあるかないかというところで、法令などはそういうことがあるんですけれども、業界団体がつくっているようなガイドラインやポリシーとかは、ある種、エンフォースはしないんだけれどもそれなりに守るということのインセンティブをうまくつくることによって、影響力があるものになるかもしれないですし、日本なんかは特に事業者ガイドラインとかをつくったとしても、かなり真面目に取り組んでいただける企業の方も多くあるので、エンフォースしない、ある種のソフトロー

でいったとしても、ある程度の規律を守るような仕組みが構築できるのではないかということも、いろいろと議論されているところであります。

一方で、その中で先ほど、国や組織をまたいでＡＩの開発や提供利用が行われているというお話もしました。この中で、それぞれの国ですとか、あるいは国際的な関係性の中で、法令ですとかガイドラインの相互認証ですとか、あるいはどのような透明性もって対応しているのかを確認し合うということは非常に難しいところもありますので、そういうときにはある意味、もう少しスピーディーにいけるような学会の基準ですとか慣習とか、あるいは市場による監視、それから最近、ＥＳＧ投資のようなところで投資家がある程度こういうような製品や商品をつくっていくことを望むというような要望を企業側に出すことによって、ある一定程度のコントロールというか規律をかけていくことでしたり、特に日本は炎上リスクといいますか評判というのを非常に重視するような国民というか、国柄がございますので、そういうところがある種の人々がちゃんとした行いをすると、規律といったようなまさにフーコー的なディシプリンが見られているかもしれないという観点から、対応しなければならないというところでございますけれども、こういうような様々な規律を使いながらＡＩガバナンスをある程度、機敏に、アジャイルに行っていくということが重要なのかなと思っております。

残りあと５～６分程度になってきたのですが、具体的にＡＩ監査というほうの話を少しさせていただければと思っております。なぜ監査の話をさせていただくかといいますと、実際にＡＩシステムがきちんと動いているのか、あるいは生成ＡＩとか様々、問題が出てくるときにどうやって対応していくのか、健全性を確保していくのかということにおいて、一つの方法論として、様々なところでＡＩ　Ａｕｄｉｔ、ＡＩによるＡＩサービスシステムの監査というのが取り上げられているからでございます。

取り上げられているんですけれども、これはそんなに簡単なことじゃないよねということを皆さんと共有したいと思います。こちらのスライドは全て未来ビジョン研究センターのところで、政策提言として出している資料をもとに説明させていただいております。こちらはＡＩ原則、先ほど来から紹介したＰｒｉｎｃｉｐｌｅとかＰｒａｃｔｉｃｅとか様々なものがある中で、じゃあ一体そのサービスの何を監査していくのかといったときに、その対象となる、ここでは立証命題と書かせていただいておりますけれども、をある種ブレークダウンして考えていく必要があるわけです。

例えば公平性ですと、ＡＩシステムの出力に不適切なバイアスがかかっていないかとい

うような形まで落とし込んで、それを、システムを見ていったり、あるいはサービスを提供する組織の運用の仕方とかを見ていくという形になります。ただ、それぞれの立証命題がどのようにしてブレークダウンできるのかといったことの基準というのは様々ですし、あるいは組織かつ各企業ですとか組織においても、何を重視するのか、どういうことを重視するのかというのはかなり多様になっているので、全てのものを一遍に見ていくということはなかなか難しいところもありますし、それぞれがどういう優先順位で見ていくのかということもかなり議論した上で、監査をしていくということが重要になってくるのかなと思っております。

また、ＡＩサービスやシステムを監査するということにおいても、どのレイヤーを監査していくのかということが、もう少し整理する必要があるかなと思っております。一般にアルゴリズム監査とかデータの中身の監査というようなことが、一般的に思いつくことだとは思います。ただ、従来のＡＩシステムがソフトウエアだけではなく、例えば自動運転とかハードの中に組み込まれていくとなったときに、そちらのハードウエア自体のデバイスの性能とこちらのアルゴリズムの関係性ということも考えなければならないでしょうと。また、どのようにしてサービスが提供されているのか、情報が適切に開示されているのかですとかということに関しても、監査というのは入ってくることが考えられます。

一方で、システムやサービスだけではなく、どちらかというとマネジメントのほう、内部統制はどのように効いているのかということを監査するということもあり、どちらかというと技術的にアルゴリズムの監査をしていくというよりは、恐らくはこちらのような統制をどのようにされているかという観点から監査をしていくという、できることからやっていくというのでは、この辺からいくのではないかなと思ってはおります。

ただ、それにしても、どのタイミングで監査するのかというのは非常に難しいなというふうに、いろいろな方々と話をしている中で感じておりまして、ＡＩというのはＰｏＣレベルとか新規開発のところがあって、ある種、それぞれをリリースした後にさらにまた追加開発をしたりとかということで、常にこの時点でもう一旦製品がフィックスされるということはなかなかない可能性もあるわけです。そのときに、監査をした時点と実際に動いているという点でＡＩの出力結果が、精度が異なるみたいなケースも出てくるでしょうし、タイミングがいろいろとある中で、どのように誰がどの段階で、内部監査の人と外部監査の人が場合によっては協力をするみたいなことも必要になってくるのかもしれません。

また、そうなったときに誰が監査をするのかというようなところでも、内部監査、外部監

査、また外部監察、スペシフィックな要件みたいなのもこちらに書いてございますが、いろいろ書いてありますが強調したいのは２点でして、ある種、もう１人が監査するというよりはチームで監査をしていくと、ＡＩの技術の知識もあれば監査の知識もあるみたいなスーパーマンを１人想定するということは少し難しいので、チームで議論していくということが重要になるということと、先ほどアカウンタビリティが大事だというようなお話をしましたけれども、何もかも責任を全て外部監査したでしょうというところで任せ過ぎてしまうと、そういうシステムを監査したくないというような監査人不足になってしまうという可能性もあるということもあるので、この辺、誰が監査するのかどうやって監査するのかということも含めて、ある種の監査人のリテラシーですとか教育というのも一緒に必要になってくるのかなと思っております。

こちらもいろいろな関係者がいますよというところで紹介をしているところでございますが、内部監査のような第一線、第二線と言われているような企業内での監査と、外部の標準化だったりとか、あるいは外部監査実施者による監査みたいなところで、これだけの人たちが関わってくるということで、これにおいてはガイドラインですとか法制度とかということも参照しながら議論していくという形になってくるかと思います。この研究会でも様々なステークホルダーをある種、特定していきましょうという話になってきておりますが、具体的に企業内部の人たちというのも、ある種、重要なファクターかなと思っております。

ＡＩ監査というのは、まとめのページに近いんですけれども、制度的な要因もあれば社会的な要因、そして技術の複雑性という様々な要因がからまっており、非常に難しいと。一方で、ＡＩ監査をやればいいじゃないという言い方もある一方で、具体的に落とし込んでいくなるとかなり課題があるなというのが、現在、我々の研究チームで出しているような政策提言には書かれてございます。

同じように、その政策提言にこちらも書いているんですが、生成ＡＩの監査をめぐる議論というのも、ある意味、生成ＡＩとか対話型のＡＩというのも、今皆さんが御議論されていらっしゃるように、著作権とか偽情報・誤情報、感情操作、いろいろな方が議論の俎上に上っております。でもそれを適切に評価する基準開発というのがまだまだ未熟な段階において、それを監査するというところまで行くのは少しラグがあるかなというふうに思っております。一方で、アルゴリズムそのものというよりは、インターフェースをどのように設計しているのかというようなことも監査対象になり得ますし、生成物に対して適切な表示が

されているのかどうかということも監査の対象となってくるのかなと思います。

こちらで最後のスライドとなります。デジタル空間における情報流通の健全性ということを考えたときに、様々な層においてＡＩガバナンスの議論と国内外の議論の協調というのを進めていくということが、非常に重要になってくるかなと思っています。Ｐｒａｃｔｉｃｅ とＰｒａｃｔｉｃｅ、様々なレイヤーがありますよということをお話ししておりますが、このような動きの速い技術、もともとはＡＩガバナンスですとＰｒｉｎｃｉｐｌｅｓ ｔｏ ｐｒａｃｔｉｃｅｓと言われたんです。まず原則をつくって、これをＰｒａｃｔｉｃｅに落とし込んでいこうと。

しかし、最近では、Ｇ７広島ＡＩプロセスがもうＣｏｄｅ ｏｆ ｃｏｎｄｕｃｔとＧｕｉｄｉｎｇ ｐｒｉｎｃｉｐｌｅを同時に出したということからもお分かりになるとおり、Ｐｒｉｎｃｉｐｌｅｓ ａｎｄ Ｐｒａｃｔｉｃｅｓになってきているわけです。なので、ほぼほぼその両輪を同時に議論していくということが重要になってきているのかなというふうに考えております。非常に抽象的な話が多かったかと思うのですが、私からの情報提供、以上となります。ありがとうございました。

【宍戸座長】　江間先生、ありがとうございました。

それでは、１０分程度でございますけれども、今の御報告に対して御質問・コメントのある構成員の方は、お願いをいたします。チャット欄で私にお知らせいただきたいと思いますが、いかがでございましょうか。

後藤先生、お願いいたします。

【後藤構成員】　後藤でございます。江間先生、どうもありがとうございます。ＡＩガバナンスの難しさが理解できました。と言いつつ、まだ分からないところがあるので教えてください。スライド３、ＰｒｉｎｃｉｐｌｅからＰｒａｃｔｉｃｅまでレイヤーがありましたが、今、マップしていただいている取組の相互の参照関係、しっかり相互参照できているのか、それともまだ独立しているのか。例えば広島G7関係のＰｒｉｎｃｉｐｌｅとＶｏｌｕｎｔａｒｙ ｃｏｍｍｉｔｍｅｎｔｓは、連携しているのは分りますが、それ以外のものは相互に連携、または尊重し合っているのか、それともまだばらばらになっている状況なのか、その辺り、教えていただければと思いました。

【江間構成員】　どうもありがとうございます。私も全て追い切れているところではないので、見聞きしている程度というところではございますが、まず、出てきた順番みたいなところもあるかと思いますが、ＯＥＣＤのＡＩプリンシパルというのはかなり先に、

２０１７年だったかな、２０１８年？　出てきているというところがありますので、様々な議論の、ある種の前提というか、定義もＯＥＣＤのものを使いましょうみたいな形になっているところがございます。

ＯＥＣＤの基本スタンスといたしましては、とにかくＯＥＣＤが新たなガイドラインをつくるとか、ＰｒｉｎｃｉｐｌｅとかＣｏｄｅ　ｏｆ　ｃｏｎｄｕｃｔをつくるというよりは、様々出ているいろいろな原則ですとかツールみたいなものを整理して、ある意味で相互運用可能なものにしていこうというような考え方に基づいているところがあります。ですので、ＯＥＣＤが、いろいろなところが出しているこういう原則とか議論とかを、例えば一覧表にしてどういうふうに関連し合っているのかみたいな分析をしていたりもするというのが、我々、ある種、後ろから見ている、マップを見ているところとかからは非常に助かるなというふうに思っているところではあります。

例えばこのＰｒｉｎｃｉｐｌｅの図というかこの縦線も、大体Ｖｏｌｕｎｔａｒｙ　ｃｏｍｍｉｔｍｅｎｔｓは後で私のほうで加えたものですけれども、こういうようないろいろなレイヤーでありますよねといって整理をされていたりとかするのは、大体ＯＥＣＤだったりします。そういうのが一方である中で、ハードローとして出てきているＥＵａｃｔ、ＡＩ法案の影響も非常に大きいと考えておりまして、これとＣｏｕｎｃｉｌ　ｏｆ　Ｅｕｒｏｐｅ、欧州評議会が今まさに起草交渉をしておりまして、来年の５月には妥結をしようと考えているＡＩ条約に関しても、同じヨーロッパということで、ＡＩ法案はもうできているので、ＡＩ　Ｅｕｒｏｐｅのほうはそれと反しないというか、Ｃｏｎｆｌｉｃｔが起きないような形にはしようということは、議論がされているという話はあると聞いております。もちろん、困るのはヨーロッパの人たちになってしまうので、そういう形で相互にお互いに見ているというような形があると思います。

そう考えたときに、ある種、上のレイヤーのほうのＰｒｉｎｃｉｐｌｅとか原則、Ｇｕｉｄｉｎｇであろうがノンバインディングの議論をある程度具体に落とし込んでいくというときには、この辺のレイヤー、技術レイヤーの標準化をどういうふうに連携させていくかということが非常に重要になってきますので、ＮＩＳＴの議論が参照されていたり、ＣＥＮ／ＣＥＮＥＬＥＣに関しても、この辺、実は非常にもうちょっと細かく言ってしまうというか、裏を少し言ってしまうと、この辺の活動に関わっている人はこちらにも関わっていたりとかするので、人がある種共有されている。なので、この辺の表向きの連携というよりは、中に入って関わっている人たちがある種、連携がされているという形なのかなというふうに

思っております。

もちろん、いろいろなところが国単位で例えば出してくるようなものに関しても、例えば日本の新事業者ガイドラインにおきましても様々なところでの議論というのを、もちろん構成員の方々も入っていらっしゃる方はそちらを参照されていらっしゃったりもしますし、事務局サイドでもその辺を全く無視してつくっているということはないわけでありますので、そちらに関しては今ある、今出てきているもの、それからこれから出てくるだろうものに関しても、お互いにウオッチしたりあるいは人同士のインフォームな情報共有みたいなものもありながら、ネゴシエーションしながらつくっているというところが実態かなと思っております。

というような状況ではあるんですけれども、一点、その意味で個人的に気になっているのは、ここに私が見聞きしている範囲内だけで出しているのが、本当に欧米中心の議論なんです。例えば中国はどういうふうな話で動いているのかというのはまた違うロジックでいろいろ出している、中国はむしろ規制のほうが強かったりするようなところが出てきたりとか、グローバルサウスでももちろんこのような議論が起きていたりして、各国レベルのガイドラインとか法令規制みたいなところではもう既にいろいろなところが出していたりします。

なので、欧米がかなりこの辺のガバナンスの議論を先導しているとはいえ、ほかの国々との関連性というのを全く無視していいというわけではないですし、むしろそことのハーモニーをいかに考えていくかということが重要になってくるので、恐らくこのＵＮのハイレベルのところはその辺も考慮に入れながら、その意味で国際的な多様性というところも考慮して、誰もが誰もというか、いろいろな国の人たちが入ってこられるような枠組みとは一体何なのかということを議論する場所、アドバイザリーではありますけれども、になるのかなというふうに考えております。

【後藤構成員】　ありがとうございました。大分、理解が進みました。ありがとうございました。

【江間構成員】　ありがとうございます。

【宍戸座長】　ありがとうございます。

このラウンド、森先生までとさせてください。森先生、お願いいたします。

【森構成員】　ありがとうございます。江間先生、御説明ありがとうございました。大変勉強になりました。

この次のスライドだったかもしれませんけれども、ハードローとソフトローを左右に対峙させるようなものをお書きいただいていたか、これですね、すみません、大変失礼しました、様々な規律の在り方、このお話でいろいろな議論があって、様々な考え方があるということを御紹介いただきましたが、特に日本の場合、事業者が割と真面目に事業者ガイドラインを守るとか、あとレピュテーションリスクを恐れるというお話をしていただきました。

ただ、この分野は結構ベンダーが外国企業なんじゃないかなということもありまして、果たしてそういう日本人だからという国民性にどの程度依存していいのかなというところが、どうだろうと思うところがございましたのが１点と、あとＡＩももう全産業横断的にといいますか、物理プロダクトの製造からモビリティのような物理的なサービスから、本件のようなコンテンツモデレーションみたいな割とサイバーなことから、もう何から何まで使われるようになっていると思うんですけれども、そんな中で一つは国民性とか遵法精神みたいな問題があると思うんですが、他方で産業別の考え方として、国家による規律なのか、どちらかというと自主規制、共同規制なのかという、そういう分け方もあるのかなと思ったのですが、そこについてはいかがでございしょうか。

【江間構成員】　御質問どうもありがとうございます。おっしゃるとおりで、あくまでこれはさくっとしたまとめというところで書いておりますが、具体的に実際にこれを強制力があるかないかというところで議論していくとなったときに、おっしゃっていただいたとおりバーティカルなレギュレーションなのかホリゾンタルなレギュレーションなのかみたいな議論もあるかと思っておりますが、いろいろな分野にＡＩが導入されていくとなったときに、基本となるのは既に存在している産業別での議論がもとになってくるのかなと思っておりまして、安全性に関して非常に重要度が高いようなところに関しましては、もちろん既に法令があるところはありますので、それに基づいて議論をしていくという形になるかなと思っております。

この図を出したのは、ある意味、そういうのは前提とした上で、速い進展がある技術に対してどのような形で議論していくのかというようなときに、様々な規律がありますよねという、バリエーションを出していくというところではあるのですが、昨今のＡＩの安全性に対する要求の議論が高まっている中においては、ある程度の法令とか、ノンバインディング、バインディングの話も含めて、議論をしていくことが大事なのかなとは思っておりまして、この図をつくった後にあえて入れていなかったというか入れなかったところがあるんですけれども、こういうモラルとか市場みたいなところがある中で、技術である種、方向性とい

いますか悪さができないようにするですとか、技術の問題は技術で対策するということがある意味、誤情報・偽情報のところでも議論されていますが、レッシグが出している『ＣＯＤＥ』みたいな話とかも含めて、いろいろな規律を使いながら議論していくということが、一つ重要なのかなという話をしたかったということで、出しております。

【森構成員】　ありがとうございました。理解が進みました。

【江間構成員】　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。

まさにいろいろな国内外のＡＩをめぐる議論で、共有されている江間先生にこちらにも入っていただいて、我々で共有させていただいてということで、引き続きよろしくお願いをいたします。

それでは、続きまして越前構成員から御発表、お願いいたします。

【越前構成員】　それでは、技術動向について少し話題提供させていただければと思います。どうぞよろしくお願いいたします。私、国立情報学研究所の越前でございます。略歴はもう何度も最初に御説明したので、ここは割愛させていただきます。

このようなアウトラインで本日、話題提供させていただきたいと思います。前半はイントロダクションということで、中段は顔を対象とするのが非常に分かりやすいので、そういったものの生成手法と、さらにその対策について述べた後に、私が関わっている国のプロジェクトについて少し御説明したいと思っております。さらに、社会実装についても御説明させていただきます。

それでは、最初に導入ということで、３つの顔画像、３枚の顔画像がございますが、このうちのどれかが実はこの世に存在しない、生成ＡＩがつくり出した顔なんですが、どれがフェイクかリアルか、お分かりになりますでしょうか。ぱっと見ても分からないかなと思うんですけれども、実はこれ全部、フェイクでございまして、顔の生成に特化した生成ＡＩがつくり出した顔画像です。よく見ていただくと、髪の毛の一本一本、生え際も虹彩も、本物と見まがうようなレベルでつくられているということが分かります。これは、私のような研究者が見ても非常によくできているなというふうに感じます。要は、こういったものがソーシャルメディアに出てきたときに、一般の方に真贋の見分けがつくかと、ＡＩ製なのか否かというのも極めて厳しいんじゃないかなというふうに感じております。

これは別の技術でございまして、Ｇｏｏｇｌｅが公開したデータセットです。右のほうが違う顔に置き換えた、Ｆａｃｅ　ｓｗａｐという技術によるものです。通常、動画ですと時

間軸で粗が見えやすいのですが、非常になめらかに、これは４年前に公開されたデータセットですけれども、ほかの顔に置き換えるということができてしまっているということです。次に、フェイクメディアの生成ということで、これはもう御存じかと思いますが、顔・音声・身体・自然言語など大量の高品質な、人間由来の情報をＡＩが学習しまして、本物と見まがうようなフェイクメディアの生成が可能になってきたということです。

ＤｅｅｐｆａｋｅやＧＲＯＶＥＲ、もともとＤｅｅｐｆａｋｅは顔をｓｗａｐするというところで出てきた技術でございますが、そういったものが２０１８年ぐらいから出てきて、ＧＲＯＶＥＲと呼ばれるフェイクニュースの生成モデルというのも出てきたという状況でございます。実はもう２０１９年頃から海外ではこういったフェイク音声で現金を搾取した事例とか、架空の人物なりすますというような事例が起きておりました。例えばこのビデオは、イーロン・マスクの一枚の顔写真を使った表情操作と呼ばれる技術でして、こういったものがビデオコミュニケーションに使われて参加者はびっくりしているという状況のビデオです。これは事故ではないんですが、こういったことができるというテクニカルなデモになります。

あとは御存じのように、ウクライナ大統領のＤｅｅｐｆａｋｅによるロシアへの降伏呼びかけとか、あとは最近問題となっている拡散モデルを用いた偽・誤情報の拡散ということでございます。これはプロンプト一つで非常に高品質な画像ができてしまうということで、大きな社会問題になっているかと思います。あと、御存じのようにＣｈａｔＧＰＴを用いたマルウエア作成ですね、フィッシングメールの作成もプロンプトをうまく工夫するとできてしまうという問題が顕在化してきているということです。

こういった背景もありまして、ここで葉、顔を対象としたフェイクメディアの生成手法について概観していきたいと思います。大きく分けて５つのタイプがございます。一つずつ、端的に御説明できればと思っております。

まず、最初が顔全体の合成ということで、これは先ほどの導入でお示したものです。潜在変数というのを用いて、実世界に存在しない顔画像全体を生成するというものです。特にＳｔｙｌｅＧＡＮというのが有名ですけれども、顔の生成に特化したモデルというのが出てきているということです。

それ以外にもこの２番目、顔の属性操作というものがございます。例えばこのインプットはリアルな顔ですが、属性、髪の色だとか性別を自然な形でＡＩによってスタイル変換するという手法も出てきております。スライドの真ん中がなぜ空いているかといいますと、実は

顔全体の生成において、今年になって、Ｓｔａｂｌｅ　Ｄｉｆｆｕｓｉｏｎなどの拡散モデルを対象とした効率的なファインチューニングの方法が出てきております。要は、今までの拡散モデルというのは、例えば洪水とか災害とか一般的な事象を表すような自然画像などはつくるのは簡単でしたが、特定の人物を対象としたチューニングを行うということも、今年になって非常に簡単にできるようになってきております。

これは、そういったものの一例ですけれども、チューニングした結果を使って、プロンプトを入れることで簡単に特定の人物に特化した、そういったフェイク映像というのが簡単にできてしまいます。ここにダウンロードありますが、ダウンロードしたパラメータ群と、例えばこの場合はＳｔａｂｌｅ　Ｄｉｆｆｕｓｉｏｎを組み合わせて、特定の人物に特化した、拡散モデルというのができてしまうということが、現在、大きな問題になっております。これによって、プロンプトのみで、つくれる画像の可能性が極めて高まってしまっているというのが大きな問題かと思います。

続きまして、顔映像、画像の表情操作というものでございます。これは、攻撃者の表情とターゲットの顔画像、映像を合成して、攻撃者の表情と同期したターゲットの顔映像を生成するというものでございます。最初に出てきたのはコンピュータービジョンのトップカンファで出てきたＦａｃｅ２Ｆａｃｅと呼ばれるものです。こちらがターゲット、これはリアルな顔映像です。これがソース、攻撃者の表情ですけれども、これにリアルタイムで攻撃者の表情をターゲットに転写してビデオを生成しようというものが、２０１６年に提案されました。

Ｆａｃｅ２Ｆａｃｅは、ソースでビデオを使っていたんですけれども、一枚の写真を使って表情を転写するという手法もその後、出てきております。これは現状、極めて大きな問題となっていると我々、考えておりまして、例えば今、国内で非常に盛んである、スマホを使ったインターネットバンキングのアカウント開設に使う顔認証、ｅＫＹＣと呼ばれるものですけれども、そういったものに対しての脅威となっております。要するに、免許証の写真一つで表情を与えることができてしまうというところで、そういったところが今、ｅＫＹＣにおける顔認証の脅威として顕在化しつつあるということでございます。

こちらは、顔映像の話し方操作ということでございます。よく知られている言葉で言うとリップシンクと呼ばれるものです。これはリファレンスの映像で、オバマ元大統領がしゃべっている映像ですけれども、そこに違う音声を、異なるタイミングで録音したオバマさんの音声を入力すると、その音声に合わせてオバマさんの口元が自然な形で同期する方法、そう

いった手法を提案しております。これはもう、デモを見ていただければすぐ分かるんですが、この映像の左側がリアルなオバマさんの顔映像と音声でして、この音声に合わせて、右側のオバマさんの顔映像の口元部分が自然な形で同期しています。リップシンクは現在、国内で大きな問題となっております。口元のみを生成してフェイクをつくるということです。右側のオバマさんの映像を見ても、これはフェイクなんですけれども、自然な形で口元が合成されているということがお分かりになるかと思います。

最後、これが先ほどの最初のイントロダクションの２番目でお見せした顔の入替え、Ｆａｃｅ　ｓｗａｐでございます。これもよく知られた手法でございますけれども、近年は非常に効率化して、一枚の写真を使って元の映像の顔を置き換えるという方法も提案されております。

以上が、顔を対象としたフェイクメディアの生成におけるものでございますが、続いては防御手法について少し御説明したいと思います。実は、我々のグループは世界で最初に先ほどの顔の入替えとか表情操作において、ＡＩを用いてリアルかフェイクか判断可能な手法を、世界で最初に提案しております。我々、研究者の評価としては、論文がどれだけリファレンスされたかというのは重要な尺度ですが、この論文は２０１８年１２月にフォレンジックス系のワークショップで発表したものですが、千回を超えているということで、この分野において最初に出された論文いうことは、このコミュニティーの中では認知されております。

原理はそれほど難しくなくて、シンプルな４層のＣＮＮ、ニューラルネットワークを使いまして、実は当時は学習するデータがほとんどなかったので、手作業でデータを集めて来て、リアルとフェイクのラベルをつけてモデルを学習させたということが記憶に残っております。このモデルに関しては、中間層を見てみますと、大体口元とか目の辺りのアーティファクトと呼ばれるようなノイズを見て、ＡＩが与えられた顔映像に対してリアルかフェイクか判断しているというのが推察されます。

さらに翌年、これを改善した方法を提案しました。これも結構高い引用数を誇っておりますが、これは先ほどより少しインテリジェントな方法になっておりまして、先ほどの手法は４層のＣＮＮだけでリアル、フェイクを判定していましたが、複数の目利きというかネットワークを使って、この３つの映像は、同じフェイク映像なんですが、カプセルと呼ばれるニューラルネットワークを用いて、異なる３つのカプセルにより、リアル、フェイクを判断しています。映像の右列の３つの映像は、違うところがアクティベートしているのが分かるか

と思いますが、要するに複数の目利きが参加することで、いろいろな観点から真贋判定してもらって、最後、それを上位のカプセルで真贋判定するということです。これは何を意識しているかというと、新たな攻撃、新たなフェイク手法が入ってきたときも、いろいろな目利きがいるので、未知の攻撃に対してもそれなりに強いということを意識してつくった方法です。

これは、この方法のデモ映像でございまして、先ほどのＦａｃｅ　ｓｗａｐも非常に高い精度でリアルとフェイクが判断できているかと思います。これは先ほどの表情操作、Ｆａｃｅ２Ｆａｃｅと呼ばれるものですが、同じ人物ですが口元が違うというものでございます。これも精度高く真贋判定ができているということがお分かりになるかと思います。

さらに、我々のグループは、真贋判定と同時に、どの部分が改ざんされたかを推定する手法というのを世界で最初に提案致しました。当時から顔のフェイクの生成手法というのは１通りじゃなくて、複数出てきておりまして、そうするとこの改ざんされた領域を見ることで、どの手法でつくられたか推測したいというようなニーズに基づいて、この手法を提案したものでございます。これはこの方法のデモ映像ですが、フェイクと判定して、さらにこういった領域が改ざんされているということを同時に推定しています。このフェイクは、Ｆａｃｅ　ｓｗａｐのスムージングによって生成されたものだなということが、この改ざん領域を見て分かるということでございます。

それでは最後に、我々のプロジェクトと社会実装のお話を少ししたいと思います。その後、技術課題について少し御説明いたします。

我々、国から、ＪＳＴから予算をいただきまして、ＣＲＥＳＴというプロジェクトを２０２０年１２月から開始しております。大阪大学の馬場口先生と東京工業大学の笹原先生が主たる共同研究者になります。この申請を提案した時はまさにCOVID-19の真っただ中でございまして、インフォデミックと呼ばれる社会に恐怖や混乱を引き起こす不確かな情報が氾濫していた状況でございました。例えば、科学的根拠のないフェイクニュースや、この図は、ソーシャルディスタンスを守って並んでいる画像ですが、望遠カメラによって特定の方向から撮影することで、意図的に密集状態を演出するような、こういったものも広義にはフェイクじゃないかと考えたわけです。素材自体はリアルなんですけれども、意図的なディレクションによって大衆の世論を誘導するとかいうものも、広義にはフェイクメディアではないかと、我々は考えたわけでございます。

このプロジェクトは、２０２０年初頭に申請しましたが、当時から今後そういったＡＩを

使った愉快犯や攻撃者が、多様なフェイクメディア、これはＦＭと呼んでいますけれども、多様なＦＭを駆使してインフォデミックを意図的に発生させる可能性があるんじゃないかということを考えたわけです。具体的には、３つのタイプのＦＭを考えておりまして、１つ目のメディアクローン型ＦＭというのは、ディープフェイクのような、本物に限りなく近いが本物でない、生成ＡＩがつくったようなフェイクであり、２つ目のプロパガンダ型ＦＭというのは、世論操作のために、この図に示した例ですね、メディアを意図的に加工するようにつくられたフェイクメディアであること。３つ目が、人間ではなくＡＩを誤動作、誤判定させるものに特化した敵対的サンプル型ＦＭというもの、これも広義の意味ではフェイクメディアだと我々は考えたわけです。

これらに対処するために、最終的には人間中心の健全なサイバー社会の実現を目指している訳ですが、このような多様なＦＭへの対処と意思決定支援が重要だということで、このプロジェクトでは、幾つか技術開発の項目を挙げております。例えば高度なＦＭ検出技術として、リアル、フェイクだけではなく、ある程度説明可能な形式でユーザーに情報提供する方法や、無毒化というのは後で少しお話ししますが、思考誘導や誤動作、誤判定が生じないように、例えば敵対的サンプルといった、ＡＩを誤判断させるようなノイズが画像に重畳されたときに、ノイズをきれいに除去してから学習データに使うとか、通常のメディアとしての視聴を可能にするという方法を考えたわけです。こういったこともこのプロジェクトでやりたいということを挙げさせていただきました。

最終的にはこういったツールを使って、情報の信頼性を高める社会システムの原理と技術を確立したいというのが、このプロジェクトの大きな目的でございます。少しデモをお見せいたします. これはこれまでの研究成果でして、顔を対象とした真贋判定の手法のデモです。さらに、これは、真贋判定と同時に改ざんされた領域についても推定を可能とした手法です。次にこのプロジェクトでやりたいことですが、画像や音声がどのフェイク生成手法によってつくられたか特定する説明可能な手法や、無毒化手法の究極の目的として、フェイクからリアルを復元するような手法を確立できないかというモックアップのデモになります。こんなことをプロジェクトの申請時に提案させていただきました。

そして目的ですが、フェイクメディアがもたらす潜在的な脅威に適切に対処すると同時に、多様なコミュニケーションと意思決定を支援するソーシャル情報基盤技術を確立するということで、人間を中心に配置すると、このような悪玉菌がなりすましや思考誘導、誤動作、誤判定をしかけるわけですが、こういったものを検出し、防御するということと、無毒

化というのは悪玉菌を善玉菌に変えて視聴やモデル学習に使うということ。さらに意思決定支援や合意形成支援をすることで抗体が生成され、健全なサイバー社会を実現するということになります。

このプロジェクトは、３つの領域がありまして、私は全体を統括すると同時にフェイクメディアの検出と無毒化を行うということと、大阪大学の馬場口先生はフェイクメディアの生成、無毒化というところで、笹原先生は思考誘導の排除とか偽情報拡散の抑制ということで、取り組んでおります。次のスライドですが、おかげさまで多くのジャーナル論文を始めとした、学術成果を上げております。メディア掲載も多くて、いろいろな成果を上げておりますが、以降では、幾つかの研究についてスポットを当てた後に、社会実装のお話もしたいと思っております。次のスライドですが、また、総務省の情報通信白書やＣＲＳの報告書に、このプロジェクトが取り上げられているような状況でございます。

ここからは少し特定の研究手法、このプロジェクトの成果としてお見せいたしますが、先ほどのモックアップのデモの最後でお見せした、フェイクからリアルを復元する手法というのを最近、考案いたしまして、この手法をＣｙｂｅｒ　Ｖａｃｃｉｎｅと呼んでいます。具体的にどのような手法かと申しますと、これはリアルの顔なんですが、この顔の中央部分の特徴を顔の周辺に分かりにくいように埋め込むことをワクチン接種と呼んでいます。これはワクチン接種済みの顔画像ですけれども、見た目はオリジナルと分かりません。この画像が公開の後、Ｆａｃｅ　ｓｗａｐという顔の置き換えという攻撃を仕かけられたときに、この画像に復元モデルを用いると、この画像のみからオリジナルのものを復元できるということです。

これは我々、いろいろなユーザー企業さんと真贋判定について話し合っておりますが、どのようにフェイクが作られたかといった来歴情報が必要だと。経緯がある程度分かることが重要だというような御指摘もありまして、これだと例えばこのオリジナルのＡさんの顔画像がどのような経緯を経てＢさんに置き換えられたのかということが分かるということで、このようなニーズがあって開発したという手法になります。名前もＣｙｂｅｒ　Ｖａｃｃｉｎｅということで、接種することで復元が可能だということを実現した手法です。

２つ目、これはＭａｓｔｅｒ　Ｆａｃｅと呼ばれる手法でございまして、これは顔識別システムに登録された複数の顔特徴と類似するような顔を、ＡＩによって生成しようという試みでございます。攻撃者側の手段でございます。生体情報のデータセットは非常に多く公開されているので、そういった特徴を読み取って、複数の顔識別に登録されたものと類似す

るような顔を生成するということです。これが登録された複数の顔で、これがこれらの顔と類似した特徴を持つＭａｓｔｅｒ　Ｆａｃｅと呼ばれている顔ですけれども、こういったものを作成可能ということで、脅威として挙げさせていただきました。Ｍａｓｔｅｒ　Ｆａｃｅは私たちの開発したフェイク顔映像の検出手法では検出可能ということでございます。次のスライドですが、今こういった成果は我々のプロジェクトのウェブサイトで積極的に公開しているということです。

次のスライドですが、国内では生成ＡＩの脅威が２０２１年頃から深刻化しておりまして、我々はそれに対処するために、次のスライドですが、このプロジェクトでＳＹＮＴＨＥＴＩＱ　ＶＩＳＩＯＮと呼ばれるフェイク顔映像の自動検出プログラムというものを開発いたしました。これはどういうプログラムかというと、ユーザーから見ると問い合わせたい動画の投稿と結果を取得するだけでいいんですけれども、バックではＧＰＵサーバーがうまくキューイングしながら推論を仕かけて、真贋判定の結果をちゃんと保存していくような仕組みになっています。

ユーザーから見て非常に簡便に使えるということでこのプログラムを開発いたしました。現状、多くの企業からリクエストをいただきまして、次のスライドですが、サイバーエージェントさんで事業利用開始のためのライセンスを開始しております。それ以外に複数の企業さんからお問合せいただきまして、現在複数の企業に有償のソフトウェアライセンスを実施しているところでございます。

次のスライドですが、これは実際のＳＹＮＴＨＥＴＩＱ　ＶＩＳＩＯＮの真贋判定の例でございますけれども、問い合わせたい画像をサーバーにアップロードするだけで結果をすぐ返してくれるというものでございます。次のスライドですが、これはサイバーエージェントさんのニュースリリース、国内最初の実用例ということでニュースリリースしております。次のスライドですが、さらに用途が非常に広がっておりますので、大学の研究室だけでは非常に難しいということで、ＡＩに対して知見があるようなパートナー企業さんを募集しまして、ライセンス事業者として募集したところ、複数から手が挙がってきたので、現在、契約の締結をして、準備をしているような状況でございます。

次のスライドですが、ＳＹＮＴＨＥＴＩＱ　ＶＩＳＩＯＮの広がる、潜在的なユースケースということで、これまで多くのユーザー企業さんとのやり取りから読み取った、想定ユーザー企業とユースケースをお示ししています。幾つかありますけれども、エンターテインメント業界だと著作権侵害のために真贋判定を使ったフェイクメディアの削除とか、報道機

関においては報道前の真贋判定による誤報道の防止、ソーシャルメディアでは、真贋判定結果（確信度）のユーザーへの提示による偽・誤情報の拡散防止や、金融機関は先ほど申し上げたｅＫＹＣのなりすまし検知ですね。法執行機関では真贋判定による証拠などの真正性確認、コミュニケーションプラットフォームだと例えばオンライン面接とか試験の不正があるので、真贋判定よるなりすまし検知や不正監視というのがあるかと思います。

次のスライドですが、ここは割愛させていただきます。私が所属する国立情報学研究所でシンセティックメディア国際研究センターをつくったということでございます。

次のスライドですが、偽・誤情報の拡散に対する技術的対策について、少しコメントしたいと思います。幾つかございまして、まず、最初の回で申し上げましたが、ＡＩを用いたコンテンツモデレーションの必要性でございます。これは、私は効果と効率性の観点から必須だと思っております。しかしながら、透明性、アカウンタビリティの確保が極めて重要だという認識は、その通りと思いますが、課題もあるかと思います。ＡＩによる推論というのは原則、ブラックボックスでございます。ここでセキュリティ的にはＡＩの学習データやベンチマークを公開しますと、それを逆手にとってＡＩの自動検知を迂回するような偽・誤情報の生成手法が出現する可能性があるので、こういうところに注意しながら透明性、アカウンタビリティを確保するのが肝要かと思っております。

さらに、多種多様な偽・誤情報の生成手法が出現しつつあります。もちろん我々が開発したＳＹＮＴＨＥＴＩＱ　ＶＩＳＩＯＮも定期的なデータセット更新やモデルの追加学習をしておりますが、既知の手法で生成された偽・誤情報の検知精度を確保しながらの追加学習はかなり難しい、時間がかかります。さらに、今後、極めて多種多様の生成手法を安定的に自動検知できるかというのは大きな問題もございます。ＡＩ製を示すような情報をコンテンツに不可分に埋め込む電子透かしの活用にも期待をしております。

さらに、自動検知モデルやデータセット、ベンチマーク自体における課題というのもございます。研究レベルでは様々な提案がされているんですが、ほとんどが現実の環境を反映しておりません、残念ながら。課題を解決するために産学連携、可能であれば国に主導いただいて、産学連携による開発・実証が極めて重要だと考えております。さらに、プラットフォーム事業者から、ＡＩ関連事業者からでもいいんですが、研究者に対してデータを提供いただけると非常にありがたいと思っています。

最後に自動ファクトチェックの課題ですけれども、これは自動検出と相補的な活用が期待されるんですが、重要なのは信頼できる情報源、誰がどのように収集してメンテナンスし

ていくのかというのも、非常に今後大きな課題となってくると思います。こういった課題を解決することで、こういった技術的対策というのは大きく進展するかと思います。すみません、伸びましたが、以上でございます。

【宍戸座長】　越前先生、ありがとうございました。

それでは、御質問・御意見がある方はチャット欄で私にお知らせいただければと思いますが、いかがでしょうか。生貝構成員、お願いします。

【生貝構成員】　ありがとうございます。スライドの４０ページ目で、ＡＩ事業者、プラットフォーム事業者から研究者にデータ提供が重要という御指摘について、具体的にどういうデータがアクセス可能ならよいかについて、少し教えていただければ幸いです。よろしくお願いします。

【越前構成員】　生貝先生、貴重な御意見ありがとうございます。御質問ありがとうございます。社会実装しているんですけれども、研究者が使っているデータセットというのはかなり特殊な環境でございまして、例えばテレビ番組のような背景がずっと同じような人物映像とか、人が１人しかいないとか、また、かなり理想的な環境で顔が照らされたような画像でございます。これを実際の用途に使おうとすると、例えばスマホのカメラのようなものでも光の環境は多様でございますし、顔の向きも全然、斜めからとか違う向きのものとか、自然な環境とは大きく、我々がやっている研究で使っているデータセットは違うんです。

実際に事業者さんから生のこういったデータをいただくことで、例えば真贋判定における精度というのは確実に向上しますし、さらにそういったものを想定してアーキテクチャ自身も開発できるというところで、実際にそういったデータ提供があるとこの分野というのは極めて大きく進むと思います。現状は、研究者は若干たこつぼ的で、与えられた研究用のデータセットに対して一番を取るというところに非常に熱心になるので、技術開発、本当に社会実装を目指すのであれば、事業者さんとかから実際のデータをいただきながら開発するというのが非常に重要かと思います。

【生貝構成員】　どうもありがとうございます。

【宍戸座長】　ありがとうございます。

石井先生、お願いします。

【石井構成員】　私の声も自分に聞こえているような状況ではありますが、質問させていただきます。大変勉強になるご発表ありがとうございました。２点ありまして、1点目は、先生の研究は基本的には顔のコンテンツを検知していくということでしょうかという質問

です。信頼できる情報源を誰が収集してメンテナンスしていくかという問題提起をされていましたが、ほかの情報、テキストや評価などが正しいかどうかはなかなか判断できない。そのあたりは、研究として技術的に難しいと思っておられるのでしょうか。２点目は、先生の御立場から見たときに、プラットフォーム事業など様々な関係事業者がいる中で、例えばコンテンツモデレーションなどのルール形成をすることなどへのご意見がありましたらお聞かせいただければと思います。よろしくお願いします。

【越前構成員】　まず、最初の御質問でございますが、我々のドメインはまず顔を対象としております。あまりにもドメインを広げ過ぎると学習が追いつかないということがありまして、顔ということをターゲットにしておりますが、方法論としては、データさえある程度、大量に集めてくれば、ほかのモダリティにも使えます。もちろん、私が一緒にやっている共同研究者は音声をやっている方とかおりますけれども、そういったことにも活用が期待されます。

２点目のファクトチェックでございます。これも、研究的にはテキストを対象にして、そのテキストが言っている文言が本当にファクトなのか否かという研究がございます。ただし、ドメインがかなり限定されておりまして、学術論文のアブストラクトのデータセットがあったときに、それと比較してクエリとなるテキストの文言が科学的に妥当かどうかという検証する問題がございます。

手法とかアルゴリズム的にはある程度、研究論文が出ておりますが、現実の環境で、これらを実装するときに問題となるのが、信頼できる情報源として我々が研究で使っているのが、たかだか論文のアブストラクトとかそういうレベルなんですよね。現実で使うとすると、大規模な信頼できる情報をどうやって収集してメンテナンスしていくかということも重要でして、逆に言うと、この課題をある程度解決できれば、研究者としてはそこのドメインにどんどん入っていくので、研究開発が進むのではないかと思っております。

すみません、最後の質問が聞き取れなかったのですが、もう１回よろしいでしょうか。

【石井構成員】　ありがとうございます。コンテンツモデレーションなど、ルール形成に対して、技術者の観点から合意形成が社会的に必要かという点について、お考えがあればお聞かせいただければと思います。

【越前構成員】　ありがとうございます。ＡＩを用いた自動検知なんですが、ＡＩというのは１００％真贋判定できるわけではないんですね、御存じのように。そうしたときの、我々は社会実装をいろいろしているんですけれども、場合によっては間違って誤判定した

結果、それの責任もアカウンタビリティもあるかと思うんですが、そういったところをいろいろな技術者以外の、例えば法律家とか、プラットフォーマーなどとの連携が必要になってくると思います。

技術的には、真贋判定というのはある程度の精度で出るんですが、エラーが出てきたときにどう対処するかというのが、我々自身だともう技術的には対処できないという中で、いろいろな、法律家とか、場合によっては弁護士の先生方とかに御相談しながら、こういったコンテンツモデレーションを進めていければなと思います。確率的には提示できても必ずエラーが存在するという中で、技術だけでは解決できないところを他の分野の専門家と連携しながらモデレーションできれば、技術としては非常に進展があるかと考えております。以上でございます。

【石井構成員】　ありがとうございます。

【宍戸座長】　このラウンドはここまでとさせていただければと思います。越前先生、本当にどうもありがとうございました。

【越前構成員】　ありがとうございました。

【宍戸座長】　それでは、今日３番目の御報告をいただきたいと思います。奥村構成員、よろしくお願いいたします。

【奥村構成員】　それでは、始めさせていただきます。奥村でございます。よろしくお願いします。私の専門領域のニュースとかジャーナリズムの世界で、情報の安全とか安心とか正確さとか信頼を守る仕組みがどのように作用しているかということについて御紹介をいたします。スライドのほかに一つ、『ジャーナリズムの原則』という有名な本がありますが、そちらの本のエッセンスを御紹介するために、一枚ペラのＰＤＦを御用意しております。お聞きいただいている中で、参照していただければと思います。ほかのお二人の発表とは違いまして、私の分野は釈迦に説法の部分も、この皆さんの間では非常にございますけれども、ただこういう当たり前と思っていることを言語化してみることもかなり重要だと思いますので、お付き合いいただければと思います。

ジャーナリズム、ジャーナリストはＵｎｌｉｃｅｎｓｅｄ　Ｊｏｂと言われています。これさえ覚えていればオーケーというような、コンクリートな基準があるものではありません。アートやサービスの側面もございます。では、その中で最も重要な技術は何かというふうに考えていきますと、『ジャーナリズムの原則』という本があります。『ＴＨＥ　ＥＬＥＭＥＮＴＳ　ｏｆ　ＪＯＵＲＮＡＬＩＳＭ』といいます。２００１年に刊行された本で、４版

になっております。

１０の原則のうちの３番目に、実はこのようなことが書いてございます。情報を検証する能力であると、ｄｉｓｃｉｐｌｉｎｅという言葉が使われています。かなり厳しく訓練をした上で情報を検証するという能力を身につけるというような意味合いと思われます。ジャーナリズムの原則の１番目はｔｒｕｃｅ、真実ということを言っていますけれども、真実に迫るためのものとして検証が必要であるというような建てつけになっています。一定の手続は共通のもので論理的なものです。しかし、経験の積み重ねとか、かんとか、そのようなことに依存する領域もございまして、これが結構複雑でございます。

皆さんも多分、お読みになったことがある２つの本を御紹介します。『クライマーズ・ハイ』というのは日航ジャンボ機事故を取材している北関東新聞という架空の新聞社のお話ですけれども、この中で事故調査委員会が尻餅事故による後部の圧力隔壁の損傷ということをスクープしたいというときに、事故調査委員会の委員長に一番、エース記者を何とか当てて、締切りを１時間半延ばしまして、その人に当てて、反応を見て、スクープを出すかどうか決めようというシーンがあるんです。

佐山という記者がそれに対応するわけですけれども、上司に伝えてくるのは「サツカンならイエスです」という言葉なんです。サツカンというのは警察官のことです。毎日警察官と接していて、あまり表情も変えずに、時にうそを言われたりもするわけです。明白な答えではないけれども、日常の警察官と触れているような態度であれば、多分口ぶりとか表情とかそういうことで絶対にイエスだと思うけれども、事故調査委員会の委員長は学術研究者でした。というわけで、主人公の悠木は、これはスクープにできないと言って、踏みとどまるというシーンがございます。

それから右側、『ＳＨＥ　ＳＡＩＤ』といいまして、「＃Ｍｅ　Ｔｏｏ」の原動力になったワインスタインのハラスメントを、実名告発記事を出したニューヨーク・タイムズの記者のスキルですけれども、被害者の怒りを抑えつつ、そしてそれを理解しつつ、何がかんどころか、家族にさえ秘密にしておきたいというような葛藤や迷いを酌み取りながら、怒らせず、失望させず、タイミングを見て、粘り強く実名を説得していくという過程は、非常にすばらしいものがあります。それから、ワインスタイン側はかなりニューヨーク・タイムズに巧妙な圧力を仕かけてくるわけですけれども、毅然と戦略的に組織的に跳ねつけるという、記者と編集者の連携というようなものも含めて、多分、検証のスキルと言えるのではないでしょうか。

こういうのは、こういうことも含めて恐らく情報の正確性、安心を考えるためのモデルになり得る部分があるのではないかということです。しかし、ニュースや読者が視聴者に信頼されるのに、とにかくすごいところがあるんだからということでは通用しません。その手続が確実であること、そして誠実であること、それが保障されなければなりません。そのためにメディアにはＣｏｄｅ　ｏｆ　ｅｔｈｉｃｓ、倫理規範というものが存在します。先ほどの『ジャーナリズムの原則』というのは、その基盤を明確に、実は表現をしております。

実はこれが多分、欧米のメディアの中では非常に重要な側面です。なぜかというと、『ジャーナリズムの原則』がどうして優れているかといいますと、民主主義とニュースの関係というのを非常に鮮やかに論理立てて説明をしているからです。ジャーナリズムの目的は自由と自治なんですけれども、これは、この本ができたアメリカでもみんな分かっているだろうといって、長らく言語化されてこなかったことです。この本は２００１年に初版が出ましたけれども、１９９７年頃から、実はプロジェクトが始まっていました。

元ニューヨーク・タイムズのビル・コヴァッチら２人が、ジャーナリストとメディア研究者などを招集し、若いジャーナリストたちが、自分たちが何のために仕事をしているのかという目的意識を失っているので、それを何とかしたいと言ってディスカッションを始めたところから始まっています。３００人以上のジャーナリストにヒアリングをいたしました。しかし質問は、あなたがジャーナリストとして一番大切にしているものは何か、ジャーナリストとしての責任は何だと思うかというような姿勢とか価値を問うような質問をたくさん投げかけました。そして、彼らの回答を集めて編み上げたのがこれらの言葉になっています。シンプルですが、十分です。

例えば、日本の議論と結構違うのは、中立という言葉が実はこの原則の中で出てきません。中立は理念としてはありますけれども、現実としては実践できないという考え方だからです。それで終わりではなくて、彼らはニュースの消費者である一般市民とのフォーラムを十数回開きました。そして、これがあなた方が必要としているものですか、ニュースメディアに求めるものですかということを何度も何度も問いかけたわけです。というわけで、この本にはこのように書いてあります。一般市民が当然期待していいもの、これらの原則はそうやって精緻に編み出され、２００７年、２０１４年、２０２１年に版を改めて出版されていますが、著者たちはそのたびにこのデジタル、ミスインフォメーションの世界で、この原則や表現がまだ通用するのかということについて、厳しく問いかけを行って、残ってきたのがこの１０の表現です。

２０１４年にはソーシャルメディアの発達に伴って１０項目め、こちらが追加されました。ソーシャルメディアでメディアにフラットに働きかけられるようになったニュースの消費者も一定の責任があるだろうという考え方です。少なくともアメリカやイギリスなど欧米の先進国では、明文化はされていなくても、このような大原則が社会で承認されていて、ニュースメディアが存立しているという構造だと理解していいと思います。そうすると、メディアの倫理規範というようなものは、どうやってその価値を守るかということに焦点が移ります。要するにＨｏｗの議論です。そして、それを公開して、読者、消費者と共有して品質保証をしていくという考え方です。程度の差こそあれ、クオリティメディアはほとんどやっていると言っても過言ではないと思います。

それをどのような形でやっているかということについて、すごく分かりやすい例で考えてみようと思います。ジャーナリストは株の取引をしない。不祥事も、新製品の開発も、会社の市場的な評価を上下させる情報をいち早くゲットできるということですから、当たり前といえば当たり前のことです。ただし、例えばＡＰという会社がありますけれども、こちらには明確にビジネス関連の記者とエディターは株を持つなとか、それから自分の取材領域の関連企業の株は持つなとかというようなことが、がっちりと書いてございます。

こちらはＢＢＣです。ＢＢＣは、少なくとも自分の金融商品がどれぐらいの利益を上げたかということを上司に説明しなければならないと書いてあります。例外は、親の資産を相続したとか、親戚の会社の株を引き受けなければならなかった場合としか書かれていなくて、あとは上司に全て報告する、反対に言うとすごく面倒くさいですから、ＢＢＣの人はほとんど持っていないということです。これは、ＢＢＣの方から直接聞いております。

こちらはニューヨーク・タイムズです。もっと細かく書いてあります。持っていい金融商品は自分でコントロールできないものに限ると書いてあります。そして、ニュースに関わるスタッフは自分で管理するタイプの金融商品は持たないと書いてあります。そして、やむを得ず保有している金融資産が現在担当のニュースで利益相反を起こす可能性があるとその記者が判断すれば、必ず上司に報告して判断をあおいで、そして場合によっては担当替えをさせられるということを受入れろと書いてあります。反対に言うと、記者が上司への報告義務を怠ったり、上司が判断をしなかったりということで、責任の所在がかなり明確になります。ニューヨーク・タイムズはさらに責任の所在を明確にするために、スタンダードエディターとかオピニオンエディターとかマネージングエディターとかというようなポジションを説明して、この人たちが一体どんなところで何の判断をするかというようなことを全て

書いてございます。

別の枠組みのやり方を少し御紹介しておきましょう。国際ファクトチェッキングネットワークというのがあります。ポインター研究所というジャーナリズムの研究所が、セントピータースバーグというフロリダ州にあるんですけれども、そちらに附属している機関です。ヨーロッパよりはアメリカ大陸と、それからイギリス、それからアジアを中心にしてファクトチェックの振興などを行っている機関です。こちらにシグナトリーという認証ファクトチェック団体がありますけれども、その資格を与えるプロセスについて少し御説明をします。

ファクトチェックはとにかく信用されなければ成立いたしませんので、認証は定評のある大手メディアでも時間をかけて行われます。しかし、世界各国の国内事情も違いますので、大きな５つの原則というのを提示しています。これらに合致した活動をしているかどうかということを、外部の人が判断するということです。どの党派にも偏らないというのが１番目、それからどのように情報収集を行っているかという、ファクトチェックにどういうソースを使ったのかということが必ず明示されているかということを、記事１本１本審査します。それから、Ｃｏｎｆｌｉｃｔ　ｏｆ　ｉｎｔｅｒｅｓｔを生じないために、資金減をきっちりと明示をして、どこから幾らもらってどう使っているかということをちゃんと公開しているかどうかということ。

それから４番目は、どのような手順をとったかということについて、２番と似ているんですけれども、その後どのように処理をしたか、ファクトチェックにはレーティングというのがございます。白か黒かというふうにすっぱり判断できるというより、グレーのものが非常に多いわけですけれども、どういうような形でそのグレーという判定を出したのかというような、根拠みたいなものを問われてくるわけです。それから、真実とかファクトというようなものは日々変わっていきます。新しい技術がどんどん明らかになって、情勢が変わってくるわけです。そうすると、ファクトチェッカーはそれにスピーディーに対応して、ここはこのように間違っていたのでこのように訂正しますというようなことを言わなければならないという原則です。

これらの基準に合致しているかということを外部のアセサーが判断いたします。一応、私もアセサーです。ただ、なかなかチャンスがないので、実はまだ実績がないアセサーなんですけれども、一応、アセサーのページを持っております。このようなことが載っています。シグナトリーに申請があるとアセサーに一斉に連絡が来ます。私は日本語と英語でアセサ

ーができると登録をしています。担当になりたい人は手を挙げると４万円ぐらいの、多分謝礼がもらえるという感じですけれども、多分そのメディアと数か月にわたって、多分１０回ぐらいのやり取りをしなければいけないので、かなり割に合わない仕事でもあります。アセサーはファクトチェック団体がやっている日々の記事を調べて、それで基準に合致しているかというようなことを、説明やレポートを受けながらやり取りをしていきます。彼らの承認のプロセスはアセサーの名前とともに記録が公開されています。日本でも今年、３つのシグナトリーが誕生しております。定着してくれることを願ってやみません。

ただし、こちらの審査方法にも弱点がございます。外部のアセサーの実力が一致しないということで、属人的な問題があります。それから、ＩＦＣＮは、実は慢性的な人手不足と資金不足に悩んでいますので、審査方法の検証が必ずしも十分とは言えない、特にインドやなんかではかなりたくさんの申請があるために、一旦取ったシグナトリーが翌年認められないというようなことがあったりもします。それから、言語の問題があります。日本は特に日本語というかなりユニークな言語を使っているために、アセサーの数が非常に限られるというような、アジアの国々特有の問題も実はあります。

このような形で、なるべく情報を出す人たちがちゃんとした行動をとっているということを保障して、ニュースやファクトチェックの安心というようなものを社会的に担保しているわけですけれども、日本のメディアというのはどうなっているかということを少し御説明しておきましょう。実は、株を保有してはいけないということを明記して公開しているのは、私の知る限りテレビ東京しかございません。テレビ東京のこれはガイドラインですけれども、見てください。マイナーチェンジでいろいろな部分は変えられているんですけれども、ガイドラインが２００２年にできてから大きな変更があまりなされていません。要するに、公開しているけれども、その後、どこまでそれをアップデートしているかということについても怪しいということなわけです。

ほかのメディアの方の偉い方に、何で株を持っちゃいけないと書いて公開しないんですかというのを、まともに質問したことがあるんですけれども、そのような方々が何人か同じようなことを言いました。言うまでもないことだからとおっしゃるんです。本当に言うまでもないことでしょうかというのは考えてみる価値があると思います。これは、日本のニュースメディアが倫理規定として公開しているものの抜粋を持ってきました。項目は結構、網羅しているものがあります。ただし、「注意する」、「期する」、「細心の注意を払う」、「努める」、要するに「How」がないということです。目標の列挙になってしまっているわけです。

これにはからくりがありまして、各メディアは内部文書としてもっと詳細なルールを設けています。ほとんどの社が持っています。記者ハンドブックとか、イントラネットでスタッフが検索できるようなシステムとかというのは必ずあります。私も全部見せてもらったことがあります。ただし、公開しないんですかとか、研究で出してもいいですかというと、それは勘弁してくださいと。いや、でもかなり精緻なものなんですよというので、とてもそういうやり取りがたくさんあって、残念なことでもあります。

そうすると、確かに内部文書には自社とか他社がやらかした事例とかがたくさん書いてありまして、そういう地雷をもう踏まないようにしましょうというようなことも書いてあるんです。ただ、そのまま公開しなくてもいいので、自分たちがどのようにしてニュースの安心や安全や正確さを保つのかという、自分たちの行動原則というようなものはもう少し明らかにすると、もしかすると社会全体がこういう手続が真っ当なものだというような認識も広がっていって、いいのではないかと思いますけれども、実はそこまで日本は行っていないというのが現状です。

これは、欧米ではジャーナリズムスクールという大学院の教育システムがあって、それを経て一定のリテラシーと能力を得たものがメディアで仕事をするというようなことになっているわけですけれども、日本のメインストリームのメディアは長らくインハウス・トレーニングをしてきました。かえって、メディア学があまりハッピーな発展状況じゃなかったというのもあるんですけれども、そちらを専攻した人をあまりとらないというような傾向もありました。

日本はメディアがメディアであるだけで信頼されていた時期がすごく長く続いていた、幸せな国だったのかもしれないです。ただ、今はもうそうではありませんので、信頼をつくり直す社会的な営みに、ぜひとも首を突っ込んでいってもらいたいと、ぜひとも思うわけです。そういう意味で、オブザーバーに民放連が加わっていらっしゃらないことは非常に残念なことでもあって、そちらのことは一言、申し上げておきたいと思います。

もう一つはコラボレーションの話です。現在、メディア同士はコラボレーション、協力しなければならないという風潮が世界的に高まっております。正確な言い方をすれば、協力できる分野、一定のスキルなど共有できる部分を探し出して、合意して、実行に移すということです。一つの出来事でも複数のメディアの評価があったほうがいい場合もありますので、それは分野をちゃんと特定をしてやらなければならない。現在、世界的に恐らく合意があるものは、大体こちらの分野だと思われます。

特にファクトチェックの世界では国単位、あるいは南米とかアフリカなどでは共通の言語を話す国、全ての中でのメディア間の連携が進んでいます。そういう意味では、東アジアとかアジアは言語がすごく違いますので、非常に苦しい環境でもあるわけです。これはミスインフォメーションがもう大量に発生していて、ファクトチェッカーの数が足りないからです。なので、世界のファクトチェッカーの中ではとにかくＤｕｐｌｉｃａｔｉｏｎは避けましょうというのが合い言葉になっています。要するに、同じ問題を重複して、複数の者がファクトチェックしなくていいじゃないかという考え方です。ファクトチェッカーの集団をセクターとして捉えて、それ全体で仕事をしようというような認識です。

幾つかの例をざっくりと御説明します。こちらはフィリピンの＃ＦａｃｔＦｉｒｓｔPHという試みです。２０２２年の大統領選挙などのミス／ディスインフォメーションの対策のための連携です。１４０以上のニュースメディア、市民団体、ビジネス団体、企業、弁護士など法律家の団体や研究者でグループになりました。中心となったのはノーベル平和賞を取ったラップラーというところのマリア・レッサさん、そういう人です。彼らの仕事は４つのレイヤーに分かれていました。４つのレイヤーというのはファクトチェックをして、それを社会のどのようなネットワークで広めていくかという、メッシュというプロセスがあって、それを研究者が記録して、分析して、アカウンタビリティ、それを記録して、分析をして、論文としても発表するということで、ファクトチェック記事を９００本近く出して、そして学術論文を２０本近く出しているという大きな成果を上げました。

ノルウェーでは２０１７年から、主要の６メディアが協力してファクトチェックをシェアする団体を運営していたんですけれども、２０２２年２月からウクライナ戦争が始まりましたので、主にロシアのミスインフォメーションを分析するという、ＯＳＩＮＴに関するプロジェクトをスタートさせました。こちらは、読めないですね、Ｖｅｒｉｆｉｓｅｒｖａｒとかいうんでしょうか、これは要するに空いているバーという意味だそうです。要するに、コロナで閉まってしまったバーを拠点にして、１２のメディア、３２人の記者が分担して仕事をしようと。こちらで面白いと思いましたのは、ファクトチェックをする労力の分担が細かく決められていることです。

これは、メディアの実力とか人員とか収益とか、そのようなものでポジションが決められていまして、０．４とかというポジションもあるわけです。ただ、この０．４のポジションの記者はどうしているかというと、別に何か２週間に１回来るとかそういう話ではなくて、記事を最終的に出すというような責任までは負えないけれども、ただファクトチェックには

加わるとか、このようなプラットフォームがあると、自分のメディアにそういうスキルを学んで帰れるとかというようなメリットもあるので、そのようにして合意をしましょうと。特定の分野について競争を保留して、情報のボリュームを社会的に厚くしていこうという共通認識ができているということです。

ブラジルの大統領選挙にも、コンフィルマ２０２２というプロジェクトがございました。５つの主要メディアが参加して、そしてＭｅｔａがやっているＷｈａｔｓＡｐｐという世界最大のソーシャルメディアですけれども、そちらも参加しました。特筆すべきは、ミダンというアメリカとイギリス、南米、インドなどにネットワークをつくる非営利のエンジニア集団があるんですが、そちらのＴｉｐｌｉｎｅというアプリを提供してもらって、そちらを重点的に使ったということです。

ユーザーが質問をチャットボットに投稿しますと、既にファクトチェックの結果があれば、５社のファクトチェックの中からそれをピックアップしてその人に答えを返してくれるというチャットボットなんですが、このチャットボットが優れていますのは、その答えがなかった場合、その５社にこの右側に書いてあるようなＴｉｐｌｉｎｅ　Ｉｎｂｏｘといいますけれども、こんな感じでこんな質問が来ているんだけれども、どこかの社がファクトチェックしないんですかというふうに、各社にそれを送るということです。それを何日かかかって各社がファクトチェックをすると、ちゃんとチャットボットでフィードバックをしてくれるというようなことで、３３万件を超える質問を処理したというのが、こちらの成果です。

これまで見てきたファクトチェックのコラボレーションの特徴といいますのは、ちゃんとリーダーシップがあってスタートしていること、それから自分たちが損だ得だというようなことを一定度外視しなければならない分野をちゃんと決めて、そこの部分は腹をくくるということ、それからエンジニアとか研究者とかビジネスとかの連携をして活動を社会的なものに広げていくことです。日本のメディアはこういうことに関しては、実はコラボレーションがあまり得意ではないです。そもそも、引用する文化が非常に希薄です。他社のスクープを引用して速報するというような例はあまり見たことがありません。数日後とかに非常に小さな扱いで、小さく伝える事例のほうが非常に多いです。

これから紹介するのは、２０１７年に私たちが行ったヒアリングのやり取りなんですけれども、今でもそんなに大きな変化はないと思われます。このヒアリングは、当時東日本大震災、福島第一原発の報道について、主要メディア１０数社に次々とインタビューしていま

した。振り返って、どのような教訓を得て、そして、ニュースメディアの組織内でどのように共有して、これからあるかもしれない南海トラフなどにどうやって対処していくつもりなのかということを聞いたわけです。当時、テレビも新聞もどこを見ても同じというような、欲しい情報がないというような批判がありました。自分に必要な情報がいつ伝えられるか分からないし、どこにあるか分からないから待っていられない。当時のツイッターなどでは、例えば電力とかガスとか水道とか鉄道などについて分野に分けて、このニュースメディアを見ればそれについては全部分かるというようなことはしてもらえないのかと。テレビは１日に数時間でもいいからそういう放送をしないのかとか、というようなこともあったわけです。

そういう分担とか、それからコラボするアイデアはどうですかということも質問の中に入れたわけですけれども、そうするとほとんど各社があり得ないと、今この段階ですることは考えていないというような反応でした。これを精神論で嘆くのは簡単ですけれども、もう少し現実的に考えておきたいと。多分、消極的な理由はもう少し深いところにあるかもしれません。それは先ほど言った、『クライマーズ・ハイ』やなんかで見たような、多少自分たちの数値化できないようなスキルに契機をしているのかもしれません。例えば、電力とか水道とかインフラなんかに分けたときに、あのときは一番の焦点が東京電力だったわけですけれども、例えば、ある一定の期間の取材を放棄して、一つの社に電力の取材を任せてしまった後、遅れをとってしまうのではないかとか。それから、ある一定の分野を自分たちが専ら担当するということになると、取材のノウハウが流出してしまうのではないかというようなことです。

ただ、共有できる取材のノウハウ等、絶対秘密を守り通したいという人脈とか情報源とかというのは区別できるはずで、実際、報道実務家フォーラムというのがございます。早稲田大学のジャーナリズム大学院などが中心になりまして、報道の優れた実例をその当事者が自分で発表して、そして日本全国からメディアの人が集まってきて、聞くような取組です。ですから、多分そういう手のうちを明かすということができないわけではなさそうです。

それから、コラボレーションを調整する労力の負担というのも考えておかなければなりません。ですから、こちらは誰がどのような形でやるのかということをちゃんと考えておいたほうがいい。プラットフォーム上でやり取りされる情報の多くはニュースであるということを考えますと、こういうコラボレーションのモデルとか枠組みとか、基準を示す責任も期待も多分、ニュースメディアにあると思っています。なので、このような動きが少し、こ

のような場で加速することを期待しております。

以上です。ありがとうございました。

【宍戸座長】　奥村先生、ありがとうございました。

それでは、御質問・コメントのある方は、私にチャット欄でお知らせいただければと思いますが、いかがでございましょうか。いかがでしょうか。

森先生、お願いします。

【森構成員】　奥村先生、御説明をいただきまして、ありがとうございました。大変勉強になりました。お話を伺っていまして、特に株のことなんかについては、確かに欧米と日本で全然違うんだなということが分かりましたし、何となくその違いが、駄目だって考えているのは両方考えているんでしょうけれども、統制の方法が全然違うというのは非常に興味深いなと思って伺っておりました。

そういうことは、このメディア全体の信頼性みたいなことに大きな影響を与えていると思うんですけれども、他方で今日のお話でありました、メディアの中立性の確保ですとか役割分担ですとか、そういったこととはまた別に日本の大手メディアが信頼を、昔は信頼されていたけれども今はそうじゃないというお話がありましたけれども、それは誠に申し上げにくいことながら、そうだなと私も思っておりまして、そこで、しかし大きな問題になったのは、確かに新聞社の人がインサイダーで検挙されて、みたいなこともありましたけれども、そういうことではなくてむしろ誤報をめぐるものであったと思いますし、誤報そのものとその後の誤報についての対応ですね、従軍慰安婦の問題が筆頭だと思いますけれども、そういうことによって大手メディアが大きく信頼を損なったのではないかと思っておりますが、それについて果たしてそうなのかということと、もしそうだとしたらどのようなアプローチがあり得るかということを、今日のお話の文脈で教えていただければと思います。よろしくお願いします。

【奥村構成員】　御質問ありがとうございました。２０分しか時間がなかったので、実は信頼を取材の過程でどういうふうに得るかというシステムについては、全然御説明、スコープになかったのですけれども、先生が御指摘になるように、そういう取材手法のところではもっと細かい規定が山ほどございます。多分、慰安婦やなんかのこと、御指摘がありましたけれども、多分これは匿名の情報源をどうするかという扱いに非常に影響すると思うんですけれども、これについても扱いは様々です。匿名の情報源は一つだけでは絶対にニュースにしないというところまで明記している社もございます。

ただ、日本の政治ニュースを見ますと、そういうようなものはほとんどなく、例えば日本の政治欄を見ますと、筋もののシングルソースの記事がどんどん乱発されているというような、もう普段のＰｒａｃｔｉｃｅからして匿名な情報やなんかに関してかなりカジュアルな面があります。これは取材元と取材先とも日々渡り合いながら、それをここまでは明かせるでしょうというような形で、多分せめぎあいながら少しずつできることを増やしていかなければならないというようなことを怠けてきた結果なのではないかなと。

それから、そういう手順も決められていないまま、何となくそういうＰｒａｃｔｉｃｅだけがありまして、ぼわっとした、それも明文化されていなくて、多分、先輩から見よう見まねで取材をするというようなことになってきますと、参照できる基準がなく、それで自分たちがやっていたとおりにやるので、悪いことはしちゃったんだけれども、どう是正するかということについての言及は甘いということになってしまいます。

ただ、程度の差こそあれ、メディアには欧米でもそういう問題というのは多数発生していますけれども、参照する基準というものが非常に弱いということになりますと、そういうことは起きざるを得ないと。実は参照する基準は、先ほど申し上げたように、彼らは持っていたりするわけですので、それをなぜ公開して、そういう議論にしないのかなというのは、理解に苦しむ部分でもあったりするわけです。お答えになっていますでしょうか。

【森構成員】　はい。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した。

ほかに、奥村先生の発表について、御質問・コメントございますでしょうか。

落合先生、お願いします。

【落合構成員】　すみません。ありがとうございます。御説明いただきまして、私も今日、大変重要なお話を伺ったように思っております。私は放送関係の検討会にも幾つか参加させていただいておりまして、その中でインターネット配信における放送局の在り方であったりですとか、その中における情報空間の健全性確保というのにどうメディアが関わっていかれるのかというのを、議論で参加させていただいておりました。

もちろん、もともとは電波の独占というものがありましたので、放送免許というものが一つの信頼性担保になっていたというところもありつつも、ただ新聞に関してはそういったものもなくといいますか、ほかのメディアにおいてもそういう中ではありますが、メディアというふうに称されている方々に対する信頼感というのは一定程度ありましたし、その中で独自の手法をそれぞれとられて、信頼性のあるスクリーニングをかけられていたのだと

いうふうに思っております。

ただ、結局インターネットの社会になってくる中で、いろいろな情報源であったりですとか、どういう過程でこの情報がつくり出されているのかということが相対化されてきたり、個人であっても信頼性のある情報を発信できるような方というのも出てきたりするであろうしという中では、一つこういった形で信頼性担保に関する取組を行っていて、それを我々としてはこういう検証をしてより向上させているんだという、多分何かそういうある種のディスクロージャーとそれに対する説明みたいなものが大変重要になってくるといいますか、そういうものをしっかりやられているものは比較的信頼性が高いといいますか、そういう受け手側の評価もされやすくなるであろうし、何かの機会に、信頼性の高いものは何なのかといったときに、適切に情報公開を行って一定の手続を定めている人たちが、そういうより信頼性の高い方々なんだというふうに見ていくのがよりいいのかとも、先生のお話をお伺いしていて大変、思いましたので、これはこの検討会でもそうではありますが、今後、地域情報などをぜひ、民放の方々にもしっかり継続して発信していっていただく中で、そういういろいろな情報開示だとか何とかというのをお願いしたりしていた立場でもありましたので、そういう方向って重要なのかなというのを改めて、私のほうで感じさせていただきました。すみません、御質問というよりかは、大変共感させていただいたので、コメントみたいな形になりましたが、本日はどうもありがとうございます。

【奥村構成員】　ありがとうございます。民放も、それからＮＨＫも、それから新聞社の方も、個人的にお話をしますと非常に問題意識が深い方はいっぱいいらっしゃるんですけれども、組織として動けないと、マネジメントがそこまでちゃんと決断するかという問題と、それから例えば彼らの倫理規程の中にも、例えばこういうときには複数の情報源を取れみたいなことがちゃんと書いてあるような社もあるわけなんですけれども、そこの部分だけでも公開してはどうですかと言っても、それをより分ける作業って膨大で、精緻にやらなければいけないとなるとそこまでのヒューマンリソースを割けないから、それは後回しになってしまうというような社内的な事情というようなものも非常に抱えていたりするというようなことも理解しておりまして、ぜひ落合先生なんかがそういうときに、ぜひ働きかけていただくことが彼らを行動に起こさせる、非常に大きな原動力になると思いますので、今後ともよろしくお願いします。

【落合構成員】　はい。ありがとうございます。なかなか内容規制というのはできないとは思いますけれども、そういう適切な取組をしている方がなるべく報われるような形とい

うのは、比較的考えやすいのかなというふうに思います。今後も御意見を踏まえて検討していきたいと思います。どうもありがとうございます。

【奥村構成員】　ありがとうございます。

【宍戸座長】　ありがとうございます。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。奥村先生、ありがとうございました。

【奥村構成員】　ありがとうございました。

【宍戸座長】　本日、お三方、江間先生、越前先生、奥村先生から、それぞれ御発表いただきましたが、ここからは残り１０分ちょっとでございますけれども、自由な意見交換とさせていただきたいと思います。何か御意見等ございましたら、チャット欄で私にお知らせをいただきたいと思います。発言の御希望があればお知らせいただきたいと思いますが、いかがでございましょうか。

奥村先生、お願いします。

【奥村構成員】　すみません、全く別件で皆さんに問題提起をしておきたいことがございます。それは、フェイクという言葉についてです。欧米でミスインフォメーション、ディスインフォメーションという言い方をしてフェイクニュースという言葉を使わないのは、かなりフェイクという言葉に政治的な意味合いが帯びているからです。ただし、日本語ってそういう英語やなんかからは遠い言語なものですから、なかなかそういうニュアンスが共有されていないのですけれども、それをそのまま英語にしてしまうとかなり問題が出てくる言語にもなっています。

これはトランプ元米大統領の存在が大きくて、２０１９年にニューヨーク・タイムズが記事にしているんですけれども、彼がひと月に４０回ぐらいメディアをフェイクニュース、フェイクニュースと言って自分の気に入らないメディアを罵倒するものだから、実は世界で四十か国ぐらいの、特に独裁者の国々が、政府の高官とか政治家とかがメディアをフェイクニュースと呼びだして、政治的な攻撃を行っていると。

実は、そういうことなので、ファクトチェックコミュニティーでは、実はほとんどフェイクという言葉はうそぐらいにしか使われなくて、ほかは全部、かなり正確にミスインフォメーション、ディスインフォメーションという言葉を使います。ただし、この言葉、すごく長いんです。それで、日本語ではすごく使いにくい言葉だったりもしますので、いろいろな頭脳が集まっているこういう場ですから、もし先生方におかれましては何かほかのいい言い方なんかを編み出していただけないかなというのは切なるお願いでございまして、一応申

し上げておこうということでございます。

【宍戸座長】　ありがとうございます。

ほかに御発言、いかがでしょうか。

少し、時間稼ぎで申し上げますと、プラットフォームサービス研究会で議論し始めたときに、一番最初はフェイクニュースと、２０１８年当時は言っていたんですけれども、奥村先生がおっしゃるようにまずいなということで、ディスインフォメーション、ミスインフォメーションの訳語として偽情報、誤情報を充てたんですけれども、ディスインフォメーションに偽という言葉を使っているのが、このままいいのかどうかは確かに問題のような気がいたしますね。ありがとうございました。

ほかに、この機に御意見・御発言等ございましたら、いかがでございましょうか。

今後の当検討会の進め方等についてでも結構でございますけれども、いかがでございましょうか。

少しまた、時間稼ぎ的に申し上げますと、本日、江間先生、越前先生、奥村先生からそれぞれ御発表をいただきましたけれども、越前先生からＡＩによる、まさにミスインフォメーション、ディスインフォメーション、あるいはＤｅｅｐｆａｋｅの問題、そしてそれに対抗するＡＩの取組、またＡＩの研究者の現状と、また連携の必要性をお話しいただきました。江間先生からは、広く言えば規制ということになるのかもしれませんけれども、ＡＩをめぐるＰｒｉｎｃｉｐｌｅと、それから現実に対応するための施策の問題、Ｐｒａｃｔｉｃｅの問題ということで、お話がありました。ＰｒｉｎｃｉｐｌｅとＰｒａｃｔｉｃｅというお話は、奥村先生のお話の中にも実は通じるところがあったかなと思いますけれども、特にＡＩガバナンスという観点から、江間先生からお話をいただきました。

また、奥村先生からは、デジタル空間における情報流通の健全性という場合に、まさに流通する情報そのものをこれまで生成し、加工し、担ってこられたジャーナリズムが現状抱えている課題、あるいはファクトチェックをめぐる議論の在り方について御議論いただいて、今までは例えばプラットフォーマーを中心とする議論だったところから、だんだん外円を広げて、問題状況の全体像が浮き彫りになってきていると思います。

クロサカ先生、お願いします。

【クロサカ構成員】　クロサカです。ありがとうございます。前回や今回の先生方の発表を伺い、大きな課題が見えてきたと感じました。具体的には、外形的な情報に対する評価と内部の規律、この２つをどのように接合させることによって、我々は信頼性や真正性を評

価・検証し得る状態に至れるのかということ、ここが実はミッシングリンクになっているんじゃないかということに気がつきました。

というのは、例えばエンジニアリング観点で言うと、例えば先ほどのＤｅｅｐｆａｋｅは、明らかにオリジナルから改変されているという評価が、もちろんかなり巧妙にはなってきているものの、一定程度可能だと思います。また、先ほどの奥村先生のお話にもあったとおり、内部の規律として健全なジャーナリズムを目指す方々の自主規制、あるいは場合によっては、欧州では法制化も進んでいますので、共同規制的なアプローチも含め、取組をしている。これはこれでまた分かるというところだと思います。

この２つがリンクしている状態であれば、情報を受け取る側は一定程度信頼することができる、ないしは何かエラーがあったとしてもきっと訂正してくれるというような、ぎりぎりの信頼感の担保が可能になるわけですが、この２つがばらばらな状態だと、誰が何を信じればいいのか、どのような理由で信じればいいのかということが、いま一つはっきりしなくなってしまうのではないか。

ミッシングリンクと申し上げましたが、これをただユーザーリテラシーだけに帰するのではなく、ほかの方法も含めて、何らかここに結びつきを持つということが、誰がどのようにそれを担えばいいのだろうかというようなこと、あと、どのような結びつきがあるのだろうかというようなことを検討してみるということが必要なのではないかと思いました。

ここで今、誰が結びつければいいのかということについて、拙速に政府だとか行政というふうに言わなかったのは、本当にそれでいいのかという吟味が必要ということが一つ。ただ、もしかするとそこに何らか制度的な担保を持つことによって、結果的に多くの利益が得られるのだとすれば、何らか役割があるのかもしれない。非常に挑戦的な物言いになってしまいますけれども、そのどちらかだけをやっていると、もしかするとあまり意味がないということかもしれないなというふうに思いましたので、その辺りがもしかすると論点の一つに今後なり得るのかなということを今日、感じた次第でございます。すみません、意見にまだ終始しておりますが、私からは以上です。

【宍戸座長】　ありがとうございました。非常に重要な御指摘だと思います。

森先生、お願いします。

【森構成員】　意見というわけではないんですけれども、今のクロサカさんのお話で私、一瞬聞き取れなかったんですが、何と何とを結びつけるところがミッシングリンクだというお話だったんでしょうか。すみません、お願いします。

【クロサカ構成員】　外形的な情報の検証と、例えば情報発信者の内部の規律、この２つです。

【森構成員】　なるほど。情報そのものの性質とその情報を生み出した主体における、中でどうやって生み出したかという、そういう話のつながりということですよね。

【クロサカ構成員】　はい。

【森構成員】　分かりました。ありがとうございました。本当に重要な御指摘だと思います。どうも、すみませんでした。

【宍戸座長】　ありがとうございます。

私も一言、今の関係で申し上げると、恐らく情報をつくる、あるいは取り扱う主体と、そしてその内部にはいろいろな規律が当然にある。その規律に服する主体が情報を発信することが結びついていた時代は、一定のメディア環境といいますか、まさに発信した情報の評価がすなわちその主体の経済的な利益も含めた評価に関わっていて、なればこそ情報を発信する、あるいは取り扱う主体が内部規律をしっかりすることによって自動的に情報のレベルを確保することと、自動的な一致があった時代ですね。

それに対してデジタル空間、今現在の情報技術の現状がそうなのか、そうでないのだとすればどういうことをすればいいのかというお話なのかなと受け止めたところです。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。私も大変、クロサカさんの御意見、ごもっともだと思いましたので、座長のお話にほとんど同じことを言っているかもしれませんけれども、分かりやすく言うと、もちろん統制のお話、内部において発信者側でどう統制するかということは非常に重要で、そこがまた今日の奥村先生のお話だったと思うんですけれども、今やプレーヤーは、座長がおっしゃるように、そんな統制をどうするかなんて真っ当な人ばかりではなくなっていて、欲望のままに発信をしているわけです。

ここに、一方では真面目な人たちが統制をどうするかという話があり、他方であまり真面目じゃない人たちが欲望に基づいて発信する、しかしながら欲望というのは一つの大きな形というか大きな流れを持っていて、それがアテンション・エコノミーなんだと思うんです。ですので、一方で真面目発信のときの統制の問題があり、他方でアテンション・エコノミーが個々の発信者と場合によってはプラットフォームさえも駆り立てるというところを、これからさらにこの検討会で詳細に見ていくのかなというふうに思いました。以上です。

【宍戸座長】　ありがとうございます。

もう１点だけ、あまり座長が余計なことを言ってはいけませんが申し上げると、まさにアテンション・エコノミーのメカニズムの中でマスコミに対する批判のドライブがかかってしまうというか、まさにアテンション・エコノミーの中での議論ではマスコミが攻撃の対象になりやすいんですね。そこが今の情報空間の健全性を変な形でゆがめている部分があるので、そこをどうしようかという話でもありますね。ありがとうございます。

落合先生、お願いします。

【落合構成員】　ありがとうございます。基本的にクロサカ先生、森先生がおっしゃられたことと似たような方向ではあると思いますが、基本的に信頼できる情報の主体に対してどう評価していくかについては、ある程度視点が見えてきたかと思っております。

一方で、媒介者の役割と、個人も含めたいろいろなタイプの発信者がいる中で、どうするとより問題が少ない可能性が高いような情報が広く拡散されていく可能性が高まるか、そこをどう追求していくかが論点かと思いました。そういう意味ではプラットフォーマー等の拡散について重要な役割を持っている方々についてどうしてもらうといいのかという点と、個人に対してもし誤った情報を拡散する可能性がある場面で、なるべく思いとどまってもらうような対策としてはどういうことが考えられるのかを全部総合して考えていく、ことがあると思います。最終的にはいろいろな方策を組み合わせて、今後の情報格差の状況を、変化を見ながら対策を少しずつチューニングしていくことを今後、議論していく方向かと思いました。私のほうからは、以上です。

【宍戸座長】　ありがとうございました。

活発な御議論をいただきましたが、本日、時間もございますので、ここまでとさせていただければと思います。

最後に事務局から何か連絡事項等、ございますでしょうか。

【内藤補佐】　事務局でございます。本日、一部音声がエコーして聞きづらい時間帯がございました。御迷惑おかけして、大変失礼いたしました。

次回会合の詳細につきましては、別途事務局から御連絡を差し上げるとともに、総務省ホームページに開催案内を掲載いたします。

以上でございます。

【宍戸座長】　ありがとうございました。

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３回会合を閉会とさせていただきます。

本日もお忙しいところ御参集いただき、ありがとうございました。これにて閉会といたします。