2026年度 横浜市立大学 データサイエンス学部

特別選抜入学試験

【海外帰国生／国際バカロレア／科学オリンピック／外国人留学生／社会人】

総 合 問 題

【注意事項】

１．試験時間は９０分である。

２．試験開始の合図まで，この問題冊子を開いてはいけない。ただし，表紙はあらかじめよく読んでおくこと。

３．問題の印刷は１ページから１２ページまである。

４．解答用紙は３枚である。

５．試験開始後，すべての解答用紙に受験番号と氏名を所定の欄に記入すること。

６．問題冊子に落丁，乱丁，印刷不鮮明な箇所等があった場合および解答用紙が不足している場合は，手をあげて監督者に申し出ること。

７．解答は必ず解答用紙の指定された箇所に記入すること。解答用紙の裏面に記入してはいけない。

８．問題番号に対応した解答用紙に解答していない場合は，採点されない場合もあるので注意すること。

９．問題冊子の中の白紙部分は下書き等に使用してよい。

10．解答用紙を切り離したり，持ち帰ってはいけない。

11．試験終了まで退室を認めない。試験中の気分不快やトイレ等，やむを得ない場合には，手をあげて監督者を呼び，指示に従うこと。

12．試験終了後は問題冊子を持ち帰ること（面接時に使用するため保管しておくこと）。

〔問題〕　次の文章を読んで，以下の設問に答えなさい。

YouTube、TikTok、X（旧 Twitter）、Facebook、Yahoo!ニュース……。

多くの読者は、私たちの生活の一部にもなったこれらのプラットフォームを「無料」で利用していると思う。では、何も払っていないのだろうか？

答えは “No” だ。確かにお金は払っていない。でも、お金と同じぐらい大切なものを払っている。アテンション（関心）や時間である。人間の脳は、自分を取り囲む情報のすべてを処理できない。ある瞬間にアテンションを向けられる対象は限定的なのだ。その意味で、YouTube を視聴するのに、金銭と同じく有限なアテンションや時間を払っているわけである。もしこれらが、金銭と異なり、増やすことも取り戻すこともできないものだとすれば、私たちは、プラットフォームの「無料」サービスを受けるのに、金銭以上に貴重な——かけがえのない——価値を犠牲にしているともいえる。

冒頭に挙げたいくつかのプラットフォーム企業は、ページビュー（閲覧数）やインプレッション（表示数）、滞在時間などのかたちで具体的に示される私たちの「アテンション」を広告主に売って、ビジネスを成立させている。このようなプラットフォーム企業のビジネスモデルは、貨幣経済（monetary economy）と区別され、「アテンション・エコノミー（関心経済）」などと呼ばれる。インターネットの普及による情報過剰時代には、供給される膨大な情報量に対し、私たちが払えるアテンションや時間が圧倒的に希少になるため、交換財として経済的価値をもって取引される。プラットフォーム企業は、刺激的で魅惑的なコンテンツを提供することでユーザーのアテンションや時間を奪い、これを交換財として広告主に「販売」しているというわけである。

（中略）

「アナログな能力」がある人とない人の格差が生じる？

山本　生成 AI の現状をざっとうかがったところで、アテンション・エコノミーへの影響についてうかがいたいと思います。

アテンション・エコノミーの課題として、例えばフィルターバブルですとか、エコーチェンバー、さらに、これらとも関連しますが、偽情報・誤情報の増幅・拡散といった問題が挙げられます。こうした課題の先には、社会的・政治的分断や民主主義の(A)スイタイがある。

以前、栗原さんと雑談していた際に、生成 AI はアテンション・エコノミーの課題を悪化させることもあれば、それを克服する “救世主” になることもある、というご指摘をされていました。

栗原　まずアテンション・エコノミーは過度な格差を生み出すわけですが、これまでもデジタル機器を使える人・使えない人の格差の話があったと思います。そして、今後は人工知能を使える人・使えない人、人工知能で利益を貰える人・貰えない人との格差が、残念ですが、確実に出てくると思います。た

だ、使える人・使えない人の差がどこで出てくるのかという点は面白い点で、それは「自分の考えを文章として書くというアナログな能力」の有無なのです。

生成 AI を使いこなすためには、人工知能にどう入力するか——プロンプトという言い方をします——が重要です。普通の人は大体ひと言・ふた言ぐらいしか書き込みませんが、それでは実は生成 AI の良さをほとんど引き出せないのです。入力するところには数千字、GPT-4 に至っては数万字、書けます。より適切に細かく書けば、その分だけ(B)コウミョウに答えてくれるわけです。つまり、生成 AI を使うためには、ちゃんとした文章を書かないといけない。「アナログは古い、これからはデジタルだ」という時代が久しく続いていましたが、生成 AI が登場してきたいま、なんと蓋を開けたら、「最新型の人工知能を使うには超アナログな能力が必要だった」ということです。

山本　なるほど。おもしろいですね。

栗原　僕も先日印象的な話を目にしました。Twitter（現 X）で、プロンプトエンジニアリング[1]をしているエンジニアが、「必要なのは、テクニカルな能力ではなくて、いかにして自分が伝えたいことを文章としてきれいに的確に書けるかだ」といったことを言っていたのです。昔ながらの「状況を察知して、書く」能力が必要だと、若い人が言うわけです。これはなかなか示唆的だなと思いました。

山本　もしかすると若い世代で分断が広がっていくかもしれませんね。若い人のなかには Google 検索すら使わない人たちがいる。彼らは、知りたいことがあっても Twitter、TikTok、Instagram などで、ハッシュタグを使って検索するわけです。要するに、Google の検索すら得意ではない人たちが出てきている。そこにきて、より高度な検索技術が必要な対話型 AI が登場すると、そうした人たちは「自分たちとは関係ない技術だ」となっていく可能性があります。

文章が書ける人、自分の考えていることを文章として表現できる人は AI をうまく使いこなしていく。一方で、そうでない人は 1 日中、ショート系動画を見続け、アテンション・エコノミーの中にどっぷり浸かってしまう。そんな分断のある世界が生じ得るということでしょうか。

栗原　そうですね。そして、生成 AI が出てくるタイミングでよく言われる負の面が、「思考力が低下する」ということです。僕としても、AI を使いこなせない側・AI に使われる側の人々の思考力は低下するんだとは思いますが、その原因を 100 %生成 AI のせいにしてはいけないと思うのです。それ以前に、インターネットが登場し、皆が当たり前のように SNS を利用する状況がそもそも思考力の低下の主たる原因です。そして SNS の台頭がアテンション・エコノミーを顕著にしたわけです。

その延長線上に対話型 AI が出てきたら、さらに人は考えなくなってしまう。だからその意味においては、生成 AI が今の状況をさらにひどくする効果というのは、やはり一面としては持っているわけですよね。

山本　なるほど。考える人は考えるけど、考えない人はアテンション・エコノミーの世界に没入し続け

[1] 生成 AI を効率的に使用するためのプロンプトを開発すること。

て、さらに熟慮しないような感じになるということですね……。

私も大規模言語モデルが出てきたときに、同じような印象をもちました。これを使いこなすには、プロンプトでそれなりの文章を作らなければならないし、ChatGPT の回答もそれなりに長く、読むのに時間がかかります。ショート動画 15 秒の世界にいる人、タイムパフォーマンス至上主義的な価値観にはまっている人が、そのような時間的コストや思考のコストに耐えられるのだろうか。すごく単純化して言えば、TikTok のヘビー・ユーザーと ChatGPT のヘビー・ユーザーは属性が違うように思うのですね。もちろん、もっと実証的に見ていかなければいけませんが。

栗原　事実 ChatGPT が生成する文章は、ChatGPT が学習に利用しているデータがそれなりの知識層からのデータに偏っていることから、当然そのような層でのやりとりを反映したものになっているという指摘があります。

山本　そうなると、反射的でファストなアテンション・エコノミーの世界は、「思考する世界」からますます切り離されていく可能性があります。その (ア)分断が問題だ、とまずは言えるのかもしれません。

熟慮する時間をつくる生成 AI は「救世主」になり得るか

山本　生成 AI を効果的に使えるかどうかはプロンプトのつくり方によるとのことでしたが、今後、それを補助する AI は生まれてくるでしょうか。例えば、こちら側の感情やニーズをうまく汲み取って、自動的に文章化してくれるような機能が考えられます。

栗原　僕らもそのような研究をやっていますが、徐々にそうなっていくと思います。誰でも文章を入れるだけで “使えてしまう” ので、「人工知能の民主化だ」という言われ方もしていて、その認識は間違ってはいないものの、現段階の生成 AI は、実はそんなに使い勝手がいい代物ではありません。

今後はおそらく「御用聞き AI」のようなイメージのものができていくのだと思います。ユーザーが不十分な入力をすると、「それじゃわからないから、もうちょっと違う情報はない？」のように答えるか、きちんと整ったプロンプトをつくって入力してくれる。そういう形になっていくのだろうとは思います。

山本　そこに、生成 AI がアテンション・エコノミーの救世主になる可能性があるのですかね。例えば、生成 AI の長い回答を 15 秒ほどの短尺動画にしてくれるような AI が出てくると、アテンション・エコノミーと “整合的” ではあるけれど、(イ)もう少しまともな世界が訪れるかもしれない。

栗原　アテンション・エコノミーにおいては、僕らが「考えない」「自分で考える時間ないし、調べるのが大変」という状況が問題の一つとしてありました。しかし、一瞬のうちに何かしらを調べてくれる生成 AI が出てくることで、それを使いこなして、きちんとものを考えるようになる層が厚みを増す可能性はあるのだと思います。さらに、そういう人たちのなかには、皆に注目されることよりも、自分なりの意見をしっかり考えたいと思う人も、少ないと思いますが出てくると思います。

単に注目されればいいということではなくて、思慮深くすることが AI を通してフィードバックされ、それが創造性につながり、あるいは付加価値が付いて戻ってくるという仕組みが出てくれば、それこそ

があるべき報酬系なのだと思います。

山本　ルーティンのような情報整理とか、“前捌き” をうまく生成 AI にやらせれば、そこで浮いた時間を重要なことに使える。熟慮ができたり、議論が深まったりすることで、創造性が生まれることもあるというわけですね。

栗原　はい。無駄な時間がなくなることで、「考える時間」ができてきますし、自分のクリエイティビティをアンプリファイする（増幅させる）ために使う人も出てきます。そして、それを見た周りの人も、いいなと思えば真似するでしょう。いずれにせよ今は皆、考えなくなっているというよりは、考える時間がなくなって、考えられなくなっているのですよね。

山本　ただ、生成 AI の利用で浮いた時間を熟慮に充てるという発想は、どこかまだ人間を信じている部分があるような気がします。仮に人間も反射的な生き物だとすると、余暇ができても、それをショート動画の閲覧やスマホ・ゲームに使うとか、結局アテンション・エコノミーの世界の中で時間を消費していくだけだという見方もできるかもしれません。

栗原　海外と比較すると、その傾向は日本において顕著のように思えるのですが、日本人に特有の何かがあったりするのでしょうか。

山本　特に EU では、人間が自律的・主体的な決定主体である、またはそうあるべきだという「人間中心主義」が重視されているように思います。ヨーロッパでは、ルネッサンス期にアリストテレス的な人間理性が再評価され、「神に従う世界」から、自らの理性をもって人間自身が決める世界へと変わっていきます。社会システムが神ないし教会中心から人間中心へと変化していくわけですね。さらにプロテスタントの台頭により、「教会」から「個人」へ、という流れも出てきて、宗教改革や宗教戦争を経て「個人」が決めることが重視される。こうした、自らの自由意志による自己決定という考えが、近代立憲主義の基礎になっていきます。EU には、こうした精神的な革命の歴史がある。だから、AI の利活用が広がって、個人の自律的で主体的な意思決定が浸食され、人間が自ら考えなくなること、人間が再び非中心的な存在になることへの強い抵抗感があるのではないかと思います。

現在、世界の AI 倫理原則などで強調されている「人間中心」は、こうした歴史的文脈を無視できないのではないか。日本の AI 原則でも「人間中心」が謳われているのですが、どうも本質的な理解が足りていないように感じます。

アテンション・エコノミーの課題と生成 AI

山本　生成 AI によってアテンション・エコノミーを悪化させる可能性についてもお聞きできればと思います。例えば「フェイクニュース」や、「エコーチェンバー」。この辺りと生成 AI はどう関係してくるでしょうか。

栗原　技術的な観点で言うと、いま注目されている生成 AI はしっかり人間がアライメントしていて、ある程度 “保証付き” なので、そうそう変なことは起きないですし、技術も進んで間違いも減っていくで

しょうから問題はありません。

では何が問題かというと、生成 AI はオープンな技術を使っているので、小さな規模のものであれば誰でもつくれてしまうということです。

例えば「ハッキングするための文章をつくるためだけ」の生成 AI や、「誰かを攻撃するためだけ」や「個人情報を抜き取るためだけ」の AI だって、つくろうと思えばつくれるわけです。もちろんそういうものは、アライメントはまず施されてはいないのです。

そうすると何が起きるかというと、社会を混乱に陥れようとする人や、プロパガンダをつくろうという人たちも出てくるでしょう。それにフェイクニュースですよね。今でも Bot[2]が使われていますが、生成 AI を使ってより “秀逸” な Bot をつくり、意図的にフェイクニュースをつくって垂れ流すということが起こり得ます。そうしてネット上に散乱してくる情報をまた AI が学習することで、負のループが完成してしまうのです。

山本　生成 AI を使って情報の “津波” を起こせるということですね。生成 AI によって、人間がつくったかのようなプロパガンダやフェイクニュースがつくれる。またそれを一気に情報空間に送り込める。その情報汚染は、次第に善意の生成 AI にも影響を与え、我々のコミュニケーションをミスリードしていく。そうして情報の “津波” を起こし、情報空間を大混乱に陥れることができるということですよね。

栗原　そうなるともう、民主主義はみごとに壊れてしまいます。

山本　そうですね。人間によるアライメントの部分でも、「透明性」や「説明責任」が乏しいと、怖い面が出てきます。どういう目的で調整したのかがわからないと、「きれいな回答に見えるが、実は特定の価値観がかなり反映された回答になっている」ということを認知できない。

栗原　そうですね。仰るようにそこは、データの透明性やしっかりしたアライメントなど、確実に起きたことをトレースできるようにしておくといったガイドラインはあっていいのではないでしょうか。

山本　今後、個々のユーザーが生成 AI をチューニングしていく方向性もあると思いますが、その場合、生成 AI がフィルターバブルを助長するようにもなるでしょうか。

栗原　AI が現在の “道具” のタイプである限り助長することになる可能性が高いと思います。生成 AI が個人適応すれば、ユーザーが「自分が好きな情報だけをレコメンデーションしてもらう」ように AI に問いかければ当然好きな情報しか集まらず、フィルターバブル度は高まります。結局は人間が「だらけて」しまうことになりかねません。

先日『ドラえもん』を見ていて面白いストーリーがありました。そのお話では、言うことを聞いてくれないドラえもんに不満たらたらで「なんとかして」と言うのび太に対して、ドラえもんが、「よし！最新型の人工知能を出す」と言って今風の AI ロボットを出しました。何でも言うことを聞いてくれるので、のび太が「ドラえもんなんかより、こっちのほうがいいや」とか言うと、AI ロボットはドラえも

[2]インターネット上における一定のタスクや処理を自動化するためのアプリケーションやプログラム。

んを家から締め出してしまいます。それで何が起きるかというと、のび太はだらけてしまって、最後には定番のしっぺ返しを食らうわけです。

今の人工知能は「道具」つまり、人が命じるままに動くものです。そういう意味で、もしも個人適応が実装されたとしたら、フィルターバブルがどんどん進みます。それはさすがにまずいので、やはり必要なのは、上述のように、のび太の言うことをたまに聞かないこともある(ウ)「ドラえもん型」なのだと思います。

「ドラえもん型」と「道具」の違いは、自律性・能動性があるかどうかです。そして、自律性を持つAIには自ら能動的に動作するための動機・目的が埋め込まれていることが必要です。目的というか、いわば「考えるためのタネ」です。もちろん僕らがそのタネを与えます。例えば、「多様性を育(はぐく)むような情報を提供して」とか、「相手のことをきちんと考えて行動して」といったことを、それこそ憲法のようにAIの目的として組み込んでおくのです。そうすれば、その“ドラえもん”のような自律型AIは、膨大な情報の中から不要な情報を除去したうえで、今はどんな情報を与えればいいかを考えてくれます。それは必ずしもその人が好む情報ではないかもしれません。

これは、親が子どもに向き合うときと同じ態度・目線です。例えば「暑いときにジュースばかり飲んだら体に良くないから、今は我慢させよう」と考えるのと同じように、現時点の道具である生成AIにおいてもそのような機能を拡張したいと僕は思っています。

山本　これはまさに「情報的健康」の話ですよね。

栗原　まったくその通りです。ただここでの課題は、ではそれをどう解決するか、というところで議論が止まってしまうことです。今の道具（生成AI）だと正直解決は難しい。もちろん情報に様々なメタデータを付けてもいいのですが、今度はそれを誰がどう使って情報のフィルタリングをするのかという問題が出てくる。

山本　そうですね。

栗原　それに、イヤな情報、つまり個人にとって好ましいものではない情報が次々に提示されるようになったら、もうそのような生成AIは使われなくなってしまいますし。ではなぜドラえもんはのび太にとってイヤな情報を提示できるのでしょう？ それは、のび太がドラえもんを心から信頼しているからです。

「道具」であれば、100％言われたとおり動かなければ僕らは信頼しませんが、この信頼は制御可能性に対する信頼です。これに対して子どもは、「イヤなこと」も言うけれど自分の親を信頼しています。それは「親が決して自分から搾取しているのではなく、回り回って自分のためになることをしているんだ」ということを理解しているからです。そういう意味で、僕らが本当に「信頼」できる人工知能が登場すれば、「その信頼できる人工知能が言ってくるんだから、その情報も我慢しよう。ちょっとイヤだけど聞こうか」となると思います。そしてそれは「道具型」の人工知能ではできないことなのです。

山本　情報的健康というプロジェクトの一つの具体的実装として、そのような AI をつくっていくことが考えられますが、今のビジネス構造の中だとなかなか難しいですよね。

自分が好む情報、要するにアイスクリームが食べたければアイスクリームをレコメンドしてくれるような仕組みのほうが結局エンゲージメントは高まるし、アテンションを得られます。現状のビジネス、つまりアテンション・エコノミーのビジネス構造の中だと、どうしてもそちらに流れていってしまうのではないか。「短期的には違うけど、中長期的に見れば、あなたにとって利益になる」という情報提供のあり方をビジネスモデルとしてどう成り立たせるのか。

私はそのためには、社会規範を醸成していくことが重要だと考えています。昔は「肉ばかり食べていてもいい」と思われていたのが、「野菜もバランス良く食べよう」と考えられるようになってきた。しかし、本人の “欲望” としては肉を食べたいわけですから、それは短期的には自己利益に反するわけですね。でも我々の多くは、食育などのリテラシー教育もあって、その “欲望” を抑えてバランスよく食べよう、となってきた。こうした社会規範を情報摂取行動についても醸成していかなければならないと思うんです。そうすれば、ビジネス構造も変わるかもしれない。ユーザーの欲を刺激して情報を “偏食” させるアルゴリズムをつくっている企業は信頼を失い、市場でも批判されていく。逆に、「ドラえもん型」で、ユーザーの情報的健康を考えて多様な情報を提供する企業は信頼を得て、市場で優位に立つようになる。

栗原　食べ物で偏食すれば、ほんとに体を壊すじゃないですか。肉体的な苦痛を伴うからこそ、僕らは是正するわけですよね。しかし、情報の場合は違います。例えばいわゆるエコーチェンバーが進んだときに、回り回って色々な悪影響が表出してくるわけですが、それらが「情報的に偏ったものを摂取していたからだ」と短絡的に結びつくとは言えないわけです。

山本　そこは重要ですよね。ですので、まずはその「結びつき」をしっかり実証し、可視化する必要があると思います。先日、栗原さんの研究室が開催している勉強会で、若い方が「自分の情報摂取傾向を可視化できるアバターをつくったらどうでしょうか」と発言されているのを聞いて、とてもおもしろいと感じました。例えば、特定の情報ばかりを摂取していると、自分のアバターがどんどん太っていく。で、最終的にパタリと倒れてしまう……。

栗原　そうですね。情報の偏りというネガティブなものが具体的に可視化などできるといいですよね。

山本　はい。ただ注意しなければならないのは、個々人に情報的健康を強制できないという点です。憲法論からすると、究極的には、自分が好きな情報ばかりを摂取する自由は否定できません。「健康」をスローガンに特定の情報を押し付けたり、特定の情報を遮断したりするのは憲法上の権利の侵害になるし、民主主義にとっても大変危険です。食べ物だって、あれを食べろとか、これを食べるな、と “強制” することはできない。最後は個人の自律的で主体的な判断を尊重すべきです。第三者ができるのは、健

康になりたいと願う人に適切な情報を提供し、健康に向けた取り組みをナッジ[3]していくことなのだと思います。

まだ「道具」としての生成 AI と、ナッジされた民主主義

栗原　前述のように生成 AI はオープンな技術を使っているので、「ハッキング用の生成 AI」のような“野良 GPT” が多数出てくる懸念があります。ただ、コンピューターウイルスに対してそれを除去するプログラムが出てくるのと同じで、野良 GPT を見つけては通報するようなものも出てくるとは思います。私たちも、フィッシング詐欺など引っ掛からないように気を付けているのと同じように、生成 AI についても、きちんとアライメントが施されていて、それなりに透明性が確保されているものを使おうという、ある種の判断力・認識力が最後は問われるのだと思います。

山本　なるほど。

栗原　僕は現時点での道具型の AI に対しては、特にアライメントが重要だと思っています。ハサミのように「切る」ことしかできない道具は用途限定型ですが、ChatGPT のような AI はキャンプで使う多目的ナイフのように汎用(はんよう)性はあるので、汎用 AI（Artificial General Intelligence, AGI）と呼ばれてはいます。銀行のオンライン窓口用の AI だと銀行業務に関する知識しか持っていませんが、ChatGPT はどんなジャンルのことにも答えてくれるからです。

ただ、AGI は多様な能力を持つものの、その時々においてどの能力を使うかは、人間が決めるのです。なので、汎用 AI はあいかわらず道具型 AI なのです。

山本　イニシアティヴは人間にある。

栗原　ところが、その人工知能がさらに進化し、人工知能自らが「自分は、今、どの能力を使うべきか」を選べる能力を身につけたとなると、それはもう “人間（生物）” と一緒です。一番わかりやすい例がドラえもんで、「自律型 AI」と言われるものです。汎用性に加えて自ら考える「自律性」を持つ AI です。

そのような自律型の人工知能が出てくると、何が起きるかというと、のび太とドラえもんの関係を見ればわかります。のび太はもともと何も考えないし、何かといえば楽をしたい人の典型ですが、それだとダメだというので未来から送られたのがドラえもんです。ドラえもんは、「のび太君、しっかりしなさいよ！」というふうに諭しますよね。自律型の人工知能は僕らに対して文句も言うわけです。

山本　おもしろいですね。例えば、民主主義や基本的人権といった憲法の基本的価値を、AI に教え込んでいく。それをよく学んだ自律型 AI が出てくると、民主主義に反するようなこととか、基本的人権の理念からズレたことを我々が行おうとすると、叱ってくれる……。ただ、こうした「ナッジされた民主主義」が (エ)本当に民主主義なのかどうかは、さらに突き詰めた議論が必要ですよね。

“ドラえもん” については私も様々なところで議論するのですが、結局、「のび太は良くなったのか」という問題があると思っています。最終話ではのび太が人間として成長したような見せ方になっています

[3]行動科学の知見から「自分自身にとってより良い選択」を自発的に取れるように手助けするアプローチ。

が、やはり全編を通してのび太は「ドラえもんに甘えてしまう」という問題があります。結局、自律型のAIが出てきたときに、人間はAIに甘えてしまって、民主主義や基本的人権を守るということを「自分で」しないのではないか。

栗原　なるほど。それについては、人とAIとの関係を考える上で示唆的な話があります。『西遊記』です。『西遊記』で孫悟空(そんごくう)が觔斗雲(きんとうん)に乗って、「俺は自由だ！」などと言いながら飛びまわるじゃないですか。それで、「俺は世界の果てまで来たぞ」と思ったら、すごく大きな柱が出てきて、それがお釈迦(しゃか)さまの指だったという話です。

山本　思い出しました。

栗原　「あ！ 俺は自由だと思ったが、お釈迦さまの手の上にいたんだ」というわけです。これはネガティブに言えば、「結局、お釈迦さまの手の平で踊らされていた」という言い方になりますが、ポジティブには「気がつかないうちにお釈迦さまに守られている」わけですよね。

お釈迦さまは何をしているかというと、人間社会というのを見守っている。たまには良かれと思い苦しめるようなこともあるかもしれませんが……。

山本　ある程度、自由にやらせるけど、限界は設定する。

栗原　ええ。僕が考える一つのAIの最終形というのは、そのお釈迦さまに相当するものなのです。重要なことは、人間はAIに守られているとか、操られているということにすら気づかないことです。AIがいい形で人間世界に対して間接的に介入するわけです。場合によっては、ちょっとした揉(も)め事を起こすかもしれませんし、何かしら人間にとって不幸なことをするかもしれません。でも、結果的にそれにより民主主義といったものが維持され、陰ながら「個の主体性に気づかせる」というようなことをする。

ドラえもんはそれを直接的にやるわけですが、AIの最終形では恐らくそれが見えなくなる。そんな風に思ったりすることがあります。

山本　ドラえもんより、もうちょっと遠くから見守るような感じですよね。

SNSで誹謗(ひぼう)(C)チュウショウ的な投稿をしようとすると、アラート（警告）メッセージが出てきて、「もう1回、考え直したほうがいいですよ」と言ってくる。こうした機能は既に実装されていますが、それに近いイメージでしょうか。つまり人権とか民主主義の矩(のり)を人間が踰(こ)えようとすると、AIが「考え直したほうがいいですよ」と注意してくれる。

栗原　そのイメージに近いですが、そのアラートをどういうふうに与えるかというところで、お釈迦さまAIは直接的なアラートは出さないのだと思います。そこでのキーワードが「ナッジ」なんでしょうね。つまりは直接「ダメだよ」と言って止めるのではなく、「自由意志で好きでやっているという図式ではあるが、実は方向を定められている」というやり方ですね。

山本　以前栗原さんと、とても興味深い議論をした記憶があります。それは、分断され、鋭く対立している人たちが対話しようと思っても、言葉が「武器化」してぐさぐさ刺さるので、なかなかうまくいか

ない。AIがその橋渡しをできないか、ということでした。

例えば天皇制や、憲法９条について冷静に議論したいと思っても、一方のグループが他方のグループに攻撃的な言葉をぶつけるので対立が深くなる。しかし、生成AIなどが、その言葉を、本質を変更せず、相手の立場を慮（おもんぱか）ってマイルドな「言い方」にしたりすることで、コミュニケーションが成立するかもしれない。要するに、AIが、分断する人々の「媒介」になるわけですが、本人たちは意識しませんから、自然と対話がナッジされて、本質的な議論が進んでいく――そのような形で、ある種の「翻訳」を生成AIにやらせると、民主主義が実現される可能性があるなという気はします。

栗原　あと、僕ら人間は根本的に合理的に判断しない生き物です。プライドがあったり、意地を張ったり、素直に謝れなかったりと。そういったことをきちんと理解したうえで、仲介する技術というのが必要なのでしょうね。

山本　コミュニケーションが滑らかにいくようにナッジしてくれるということですよね。しかし、そうすると、これまで自由意志と考えられてきたものに、しれっとAIが入り込んで、ハイブリッド化したかたちになってくる。

栗原　それをもって、「AIが僕らの思考を変えている」と言えば、そのとおりかもしれません。ただ、少なくともそれにより民主主義を維持するというゴールがあるレベルで維持されるということになるのだと思います。

山本　ただ、AIにナッジされているということがわからないというのは、やはりマズい気がします。その意味でも透明性は重要ですよね。透明性が確保されることで、AIに媒介されていることに気づく。あ、これは自分の言葉で直接伝えなければ、と思えば、AIの“スイッチ”を切れる。その自己決定を可能にするような仕組みが必要だと思います。

最終的にAIを使うか、使わないか、どのように使うかという決定が本人に(D)リュウホされていれば、今までの世界とそんなに変わらない気もします。そもそもこれまでも自由意志はフィクションだといわれてきましたし、民主主義だってナッジされてきたところがある。例えば、日本国憲法は二院制を採用していて、熟議をナッジしている。また、民主的に成立した法律でも、最高裁判所の違憲審査を受けることになっている。要するに、憲法は「民意」を100％信用していない。冷静に熟議するようナッジするための仕掛けを随所に組み込んでいるわけですね。その意味では、私たちの民主主義は、もともと裸の民主主義ではなく、憲法に理性的熟議をナッジされた「立憲」民主主義だったわけです。こうした「立憲」的な仕掛けのなかに、AIやテクノロジーが新たに加わったと考えれば、ではそれをどう使っていこうか、という話になる。

問題は、憲法は主権者である国民が制定したという建前になっていますから、憲法上のナッジは、あくまで我々人間が、自らの意思で組み込んだもの、ということになります。ホメーロスの叙事詩に出てくる英雄オデュッセウスは、美しい歌声をもつ海の怪物セイレーンの誘惑に負けないため、船のマスト

に自らの身を括(くく)り付けた。憲法も、人間は感情的になり、暴走するだろうことを見越して、最初からこれを(E)ヨクセイする仕掛けを用意した。しかし、重要なのは、その仕掛けを我々の意思で、透明性をもって作為的にデザインしているということです。AI によるナッジは、そのように我々自身がコントロールできるものなのか。もしそうならないとすると、AI と人間の関係は逆転することになります。AI によるナッジはとても魅力的だけれども、ここに気をつけなければならない。

今回の対談で、ChatGPT などの生成 AI によって、当面はアテンション・エコノミーが悪化する可能性があることがわかりました。ただ、利用の仕方によっては、アテンション・エコノミーの課題が克服され、より良い民主主義が実現する可能性があることもわかった。だとすれば、人間と AI との主従関係が逆転しないための “ガードレール” をしっかり設けて、後者について積極的に議論していくことが必要であると感じました。

出典：山本龍彦 (2024) アテンション・エコノミーのジレンマ—＜関心＞を奪い合う世界に未来はあるか　株式会社 KADOKAWA　を一部改変

（1） 下線部（A）～（E）のカタカナを漢字に直しなさい。

（2） 二重下線部（ア）に関して，生成 AI によってどのような人々の間での分断が，なぜ生じるのか。100 字以内で答えなさい。

（3） 二重下線部（イ）に関して，まともな世界とはどのような世界か。また AI によってなぜまともな世界が訪れるのか。さらに「訪れるかもしれない」という不確実性を含んだ表現となっているのはなぜか。合わせて 125 字以内で答えなさい。

（4） 二重下線部（ウ）に関して，「道具」ではなく「ドラえもん型」が有効に機能するためには何が必要とされているか，二人の対話に即し，箇条書きの形で 4 点を挙げなさい。

（5） 二重下線部（エ）に関して，生成 AI によって本当の民主主義が実現するのか，それともしないのか，本文で触れられているそれぞれの理由に言及した上で，あなたの考えを理由とともに 400 字以内で説明しなさい。

（6） 山本氏は最後の段落で，生成 AI によってアテンション・エコノミーが悪化する可能性についても，アテンション・エコノミーの課題が克服される可能性についても触れている。生成 AI によってアテンション・エコノミーは悪化するのか，それとも課題が克服されるのか，具体例を挙げながらあなたの考えを 600 字以内で答えなさい。