パーソナルAIとリスク管理

Personal AI and Risk Management

国立研究開発法人理化学研究所　革新知能統合研究センター
社会における人工知能研究グループ　グループディレクター
橋田　浩一

1986 年より 2001 年まで電子技術総合研究所。1988 年から 1992 年まで（財）新世代コンピュータ技術開発機構に出向。2001 年から 2013 年まで産業技術総合研究所。2013 年から 2024 年まで東京大学。2017 年から理化学研究所。2020 年から現職。専門は自然言語処理、人工知能、認知科学、サービス科学など。日本認知科学会会長、言語処理学会会長等を歴任。

hasida53@gmail.com

1　パーソナルデータの分散管理

パーソナルデータ (PD: 個人情報を含むデータ) の集中管理とは、1 人の管理者が多くの人々の PD を管理することである。PD の分散管理とは、1 人の管理者が 1 人の PD を管理することである。現在は各個人の PD が多数の事業者に断片化して散在しており価値が低いが、各個人の PD を名寄せすることによって価値が高まる。しかし、名寄せした PD を多くの人々にわたって集中管理するのは危険なので、分散管理 (管理を個人に分散すること) が望ましい。集団全体の最適化 (課税、教育、公衆衛生、治安維持など) のためには集中管理が必須だが、ほとんどの個人向けサービスは各個人の最適化なので分散管理で十分であり、その方がリスクとコストが低い。また、PD を本人 (のアプリ) が管理すれば、本人の意思だけで活用できるので、付加価値が高い。この活用は 1 次利用 (本人のための利用) と 2 次利用 (個人を特定しない統計分析など) の両方を含み、いずれもサービスの質とリスクの管理を含む。

パーソナル AI (PAI) [2] が PD の価値を最大化する。PAI とは特定個人に専属する AI であり、利用者と対話して適切なサービスを仲介 (選定・実行) することにより利用者のニーズを満たす。PAI が仲介するサービスの入出力データが自ずと利用者の手もとに集約されるので、それを PAI がフル活用すれば利用者に対して非常に価値の高いサービスを提供できると期待される (図 1)。

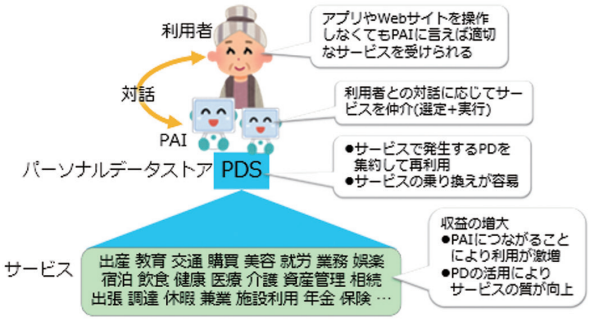

図 1　パーソナル AI がパーソナルデータの価値を最大化する

AI とのマルチモーダル対話が人間と情報機器とのインタフェースになる [3] ことにより、IT リテラシーデバイドが解消する。人間はアプリや Web サイトを操作せず、音声やテキストや表情や身ぶり手ぶりで AI と対話することによってさまざまなサービスを簡単に受けられるようになる。「適当な銀行と生命保険の口座をマイナポータルに登録して」とか「あまり利用してないサブスクから退会しといて」とか「3 時に市民病院に行きたいからタクシーを手配して。タクシーよりライドシェアが良いけど」とか言えば AI がそのように取り計らってくれる。

このようなことは技術的にはすでに可能になっており、また一般市民にとっても政府や事業者にとってもメリットが大きいので、PAI は 10 年で普及するだろう。たとえば基礎自治体の業務の大半は福祉であり、そこでの最大の問題のひとつは高齢者が情報機器を使ってくれないということだが、その問題がかなり解消するだろう。民間のサービス提供者にとっても、PAI につながることによって、サービスの利用が大幅に増えるとともに、各サービスでより多くの PD を使えるのでサービスの質

が高まり、ゆえに収益が増大する。

2 サービスの仲介

PAI につながった多様なサービスの入出力データは利用者のパーソナルデータストア (PDS) に集約 (名寄せ) して利用者が自由に活用できる (その自由度をアプリで制限することもできる)。つまり、PAI が PD の分散管理をもたらす。

サービスを利用するために多くの個別のアプリや Web サイトを操作するよりも PAI と話をする方が簡単だから、PAI はあらゆるサービスを仲介するようになるだろう。その仲介手数料を稼ぐ事業の市場規模は巨大である。たとえば小売業の平均的な付加価値 (粗利) は売上の 25 ～ 30% であり、Expedia の手数料は宿泊費の 20% ほど (それは予約を代行するだけでなく宿泊需要予測に関する情報提供等の経営支援サービスを含むからである) なので、B2C サービスの仲介手数料は平均 15% 程度とすると、B2C サービスに関する PAI 事業の市場規模 (付加価値の合計) はほぼどの国でも GDP の 70% ほどだから、その仲介手数料の合計は GDP の 10% ほどになるだろう。さらに、B2B サービスの市場規模は GDP の 2 倍ぐらいなので、PAI 事業全体の市場規模は GDP の 30% を越えると考えられる。

まだ完成度は低いものの、そういうわけですでにいくつかの PAI が商品化されている。今後さらに多くの PAI が商品化され、PAI 提供事業者の間の熾烈な競争になることは目に見えている。その競争に生き残るには自社の PAI の付加価値を最大化する必要があり、そのためには PAI が利用者の PD をフル活用する必要がある。しかしそれには PAI に対する利用者の厚い信頼が不可欠である。その信頼の内容は、PAI がプライバシ等の人権を守ってくれることと、十分良いサービスを提供してくれることだろう。

3 行動操作の阻止と共有価値の最大化

また、そのような信頼が成立するように PAI が適正に管理されるならば、図 2 のように PAI があらゆるサービスのゲートキーパとなることによって、オンラインの不正な行動操作を防ぐことができる。

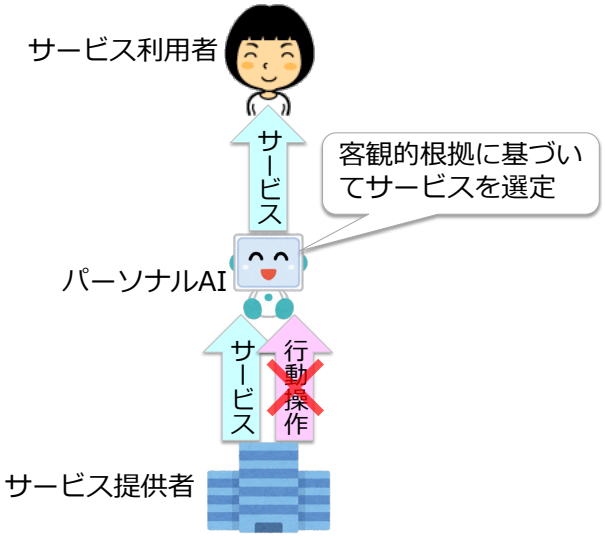

図 2　PAI が行動操作を阻止

たとえば、サブスクリプションのボタンをうっかり押してしまうように頻繁に表示したり、退会のボタンを Web サイトの奥の方に小さく表示して退会を防いだりするなど、多くのサービス提供者が収益増大のために利用者の認知バイアスや認知限界に付け込んで行動を不正に操作している。これは、PAI が客観的根拠のみに基づいてサービスを仲介し、利用者に行動操作を及ぼさないことによって阻止できる。

こうして、オンラインの不正な行動操作が不可能になり、いわゆる注意経済と監視資本主義 [4] が終焉する。オンラインの広告も利用者が見ることなく PAI が無視するので無用のものとなる。広告収入に依存する SNS や検索サースが消える。あらゆる商業の顧客接点が PAI に集約されれば、個人情報を使って PAI を通さずに各顧客にアクセスすることが不可能になるので、商業目的で個人情報を詐取することもなくなる。世界の GDP の 0.6% 程度にすぎない広告業が、はるかに生産性が高く GDP の 30% を越える PAI 事業に吸収されるわけである。

このように PAI が適正に管理されていれば、さらに、PAI は利用者の信頼を得て PD をフル活用し、共有価値を最大化することができるだろう。以上をまとめると図 3 のようになる。まず、PAI がサービスを仲介することにより、PAI の適正な管理を前提として、オンラインの不正な行動操作をなくすことができる。また、仲介するサービスの入出力データが利用者本人の手もとに集約されるので、同じく適正な管理を前提として PAI が PD をフル活用して共有価値が最大化する。

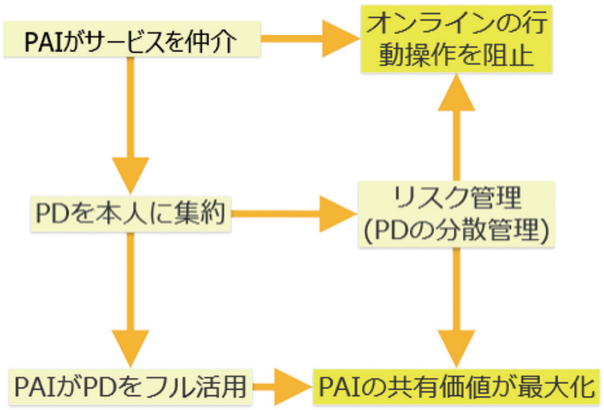

図３　PAI の管理と効果

4 サービスの質とリスクの管理

PAI 等のサービスの評価およびその評価のガバナンスの仕組みを図４示す。メディエータは多数のサービス利用者と多数のデータ利用者からなる両面市場において PD の２次利用を仲介する仕組みであり、データ利用者からの委託により多くの PAI 利用者から本人同意等に基づいて PD を収集して分析し、データ利用者に分析結果を納入する。PAI 提供者はデータ利用者の一種であり、メディエータが納入した LLM に RLHF 等を施したものを PAI 利用者に提供する。他のデータ利用者が供給する知識に基づく PAI サービスのメリットやリスクを、サービス評価者がメディエータによりそのサービスに関連する PD を収集し分析した結果に基づいて評価する。また、評価者同士がサービスの評価結果を互いにチェックするという分権的な仕組みによって評価者のガバナンス (サービスのメタガバナンス) がなされる。メディエータを用いた PD の２次利用は、サービスの評価以外にも、商品やサービスの開発、人間や社会の研究、政策の立案と検証などに及ぶ。

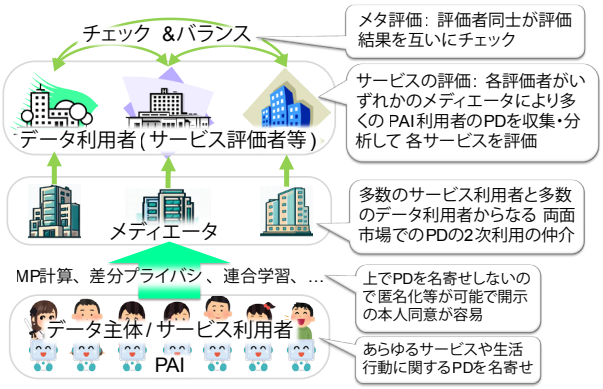

図４　PD の２次利用の一環としてのサービスの評価とメタ評価

以上のような管理を PAI 提供者に義務付ける方法として、国際標準化が考えられる。PAI は欧州 AI 法 [1] の第６条が規定する高リスク AI に当たり、同法が定める質とリスクの管理が義務付けられるが、その詳細は CEN/CENELEC JTC21 が策定する欧州標準で定められる予定である。つまり、所定の欧州標準 (整合標準 harmonized standard) を満たせば欧州 AI 法を満たしていると見なされるが、整合標準を満たさない場合は別の規格を作ってそれを規制当局に認めさせる必要があり、それは非常に大変なので、整合標準はほぼ法律と同じ拘束力を持つことになる。

一方、JTC21 と ISO/IEC JTC1/SC42 の合意に基づいて SC42 がその整合標準に対応する国際標準を作成することになっているので、上記のような PAI の管理法をこの国際標準に組み込むことにより、欧州 AI 法に組み込むことができるだろう。欧州 AI 法が GDPR と同じく世界中で事実上のデファクトスタンダードになれば、PAI の適正な管理を世界中で義務付けることができると期待される。

5 おわりに

これまでは個別サービスが利用者の注意の獲得を競っていたが、あらゆるサービスが PAI 経由で利用者に提供され、利用者は常時 PAI を使うので、PAI 提供者に必要なのは利用者の注意ではなく信頼である。また、PAI が仲介するサービスは、人間の認知バイアス等に付け込んで行動を操作することができない。こうして注意経済や監視資本主義 [4] が終焉し、サービスの提供者と利用者が信頼に基づいて価値を共創する時代になる。

参考文献

[1] EU (2024) The EU Artificial Intelligence Act. https://artificialintelligenceact.eu/

[2] Kôiti Hasida (2024) Personal AI to Maximize the Value of Personal Data while Defending Human Rights and Democracy. in Johannes Glückler and Robert Panitz (eds.) Knowledge and Digital Technology, Springer, 239-

256. https://link.springer.com/chapter/10.1007/978-3-031-39101-9_13.
[3] Sunok Lee, Minha Lee & Sangsu Lee (2023) What If Artificial Intelligence Become Completely Ambient in Our Daily Lives? Exploring Future Human-AI Interaction through High Fidelity Illustrations. International Journal of Human–Computer Interaction, 39:7, 1371-1389, DOI: 10.1080/10447318.2022.2080155.
[4] Shoshana Zuboff (2019) The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs. ISBN 9781610395694. OCLC 1049577294.