

AI を利用した意思表示と錯誤取消に関する一考察

——ドイツ法を素材として——

中 谷 崇*

目 次

- I. 序 論
 - 1. 問題設定
 - 2. 先行研究と課題
- II. ドイツ法の議論の整理
 - 1. 道具説
 - 2. 独自の信頼責任説
 - 3. 代理人説
- III. むすびに代えて
 - 1. 各学説のまとめ
 - 2. 日本法への示唆

I. 序 論

1. 問題設定

近年、人工知能（AI）の高度化に伴い、契約締結などの法律行為に AI が直接関与する場面が急増している。たとえば、オンラインマーケットのエージェント（Agent）や金融市場でのロボットアドバイザー（Robo-Advisoren）が用いられている¹⁾。こうした取引における問題の1つは、AI が生成した意思表示に「錯誤」が介在した場合、どのように扱うべきか、という点で

* なかや・たかし 立命館大学法学部教授

1) Dornis, Künstliche Intelligenz und Vertragsschluss, AcP223 (2023), 717f（以下、断りがなければ「Dornis」として引用する）。

ある。従来の民法体系は、人間の意思表示を前提に構築されている。しかし、AI による表示は、必ずしも法律効果に結びつく人間の具体的な意思を直接反映しているとは言えない。そのため、錯誤取消の制度をそのまま適用できるかは明らかではない。

2. 先行研究と課題

この問題に対して、ドイツ法ではすでに活発な議論が展開されている。たとえば、本文で言及する A. ツインマーマンは、AI による表示をどのように法的に構成すべきかについて、少なくとも三つの学説を提示している。すなわち、AI を単なる「道具」とみなす立場（道具説）、AI の自律性を前提に新たな「信頼責任」の構成要素を導入する立場（独自の信頼責任説）、そして AI を代理人に準じて扱い代理規定を類推適用する立場（代理人説）である。

すでにわが国においても、ドイツ法を参照して AI を用いた契約に関する論考が公表されており²⁾、スマート冷蔵庫による自律的注文を素材に、意思表示の帰属や錯誤取消の可否が論じられている。

本稿では、ドイツ法における AI と錯誤取消をめぐる議論を参照しつつ、その整理を通じて日本法における議論の出発点を明らかにし、今後の検討に資する基礎的な問題提起を行うことを目的とする。

Ⅱ. ドイツ法の議論の整理

A. ツインマーマンによれば³⁾、AI を用いた契約における AI の役割の位置づけについて、学説上のいくつかの立場が提示されている。以下詳細に

2) 白井豊『『コンピュータ・AI による契約締結』に関わる法律行為法上の論点整理と解決手法の覚書』立命392号（2021年）138頁。

3) A. Zimmermann, Irren ist menschlich? – Irrtumsanfechtung und künstliche Intelligenz, ZfPW 2024, 490（以下、断りがなければ「A.Zimmermann」として引用する）。なお、A Zimmermann 自身は代理人説を支持している。

紹介する。

1. 道 具 説

この立場は、AI を自立していない道具（*unselbstständiges Werkzeug*）として理解し、自動化された意思表示に関する原則を AI にも拡張する。ここでは、コルネリウスの見解を取り上げて紹介する⁴⁾。

① 意 思 表 示

まず契約締結の基礎には意思表示が必要であること、そして意思表示には、客観的第三者の観点から法律効果を惹起する意思を認識させる明白な行為ないし推断の行為が必要であり、また主観的な最低条件として、表示を企図する人間の意思が存在する必要がある（S.354）。

② AI の法人格

もっとも、意思表示は人間の行為を前提としているが、AI⁵⁾ は社会的行為をする能力を持つものの、いまだに自己の意思形成の基礎である自己の存在の認識ができるほどに技術は進歩しておらず、また AI には法人格も与えられていない（S.354）。そこで、AI の表示が利用者に効果帰属するかが問題となるが、これはコンピューター表示（*Computererklärung*）と同じと考えることができる（S.354）。

③ コンピューター表示⁶⁾

コンピューター表示とは、意思表示の一つで、表示が事前のプログラム

4) Cornelius, *Vertragsabschluss durch autonome elektronische Agenten*, MMR 2002, 353。本節で引用する際は本文でページのみ挙げる。

5) Cornelius 論文では *elektronische Agenten* の語が用いられているが、本稿では便宜上 AI と表記する。本稿では、ドイツ語での表現如何にかかわらず、AI を用いる主体を「利用者」と統一表記する。

6) 電子的意思表示（データが電子的媒体を介して伝達される表示全体の総称）の一種とさ

に基づいてコンピュータープログラムによって自動的に作成され、電子的に伝達されるものであり、人間はそれに関与しない (S.354)⁷⁾。利用者は、プログラムされた指示 (Vorgabe) に従って表示が作成されるという意思を持つだけで、いつ、誰に、どういう具体的内容で表示がなされるかを知っているわけではないが、通説によれば、これは意思表示だと認められている (S.354)。自動販売機とそこで用いられる不特定の者への申し込み (Angebot ad incertas personas) という法形象と同じように考えることができ⁸⁾、コンピューター表示でも契約締結に至るそれ以上の意思活動 (Willensbetätigung) や表示は必要ない (S.354)。

④ 代理人説への言及

別の見解は、代理規定 (BGB164条 [代理人の表示の効力] 以下) の適用を指

れる。下位分類として、①電子的に伝達された意思表示、②コンピューター表示、③自動化された表示、④エージェント表示がある。①は、自然人がコンピューターを用いて構築し、流通させるもので、コンピューターはその伝達を引き受けるに過ぎない。②は①とは区別される。完全にコンピューターシステムによって生成される意思表示である。ただし、自律的なものではなく、人間があらかじめ設定した条件に基づいている。③は、①と②の中間だと位置づけられる。人間の意思表示をコンピューターが修正するという見解と、②と同じだが、コンピューターが独立していない点が相違するという見解がある。いずれにせよ、通説によればこの概念は②の類語である。④は、コンピューターシステムの自律的な行為を対象とし、具体的に規定されたアルゴリズムに基づいて、コンピューターシステムにより自立的に意思表示が生成される (Herold, Vertragsschlüsse unter Einbeziehung automatisiert und autonom agierender Systeme, 2020, S.61ff)

7) Mehrings, Vertragsabschluss im Internet, in: Hoeren/Sieber, Handbuch Multimedia Recht, 2000, Rdnr. 24 ff.

8) Fritsche / Malzer, Ausgewählte zivilrechtliche Probleme elektronisch signierter Willenserklärungen, DNotZ 1995, 7f によれば、次のように説明される。自動販売機を設置することによって、商人は、——通説によれば——不特定の者への申し込み (Angebot ad incertas personas) の意思を表示して、顧客は硬貨を投入し商品取り出し口を開けることで、承諾している。この申込は、自動販売機が作動し、顧客が望んだ商品の在庫がまだあるという条件を含んでいる。商人による具体的な契約締結を目的とした意思表示や意思活動 (Willensbetätigung) は、もはや必要ない。オンライン通販 (コンピューター表示) も本質的にはこれと同じである。

摘する。コンピューターシステム自体には法人格がないのだから、せいぜい代理規定の類推適用が考慮に値するに過ぎない。最終的に、表示の作成と表示をなすことがコンピューターシステムによって行われることがこの見解を裏付ける。コンピューターシステムにおいては、プログラム及びデータ記憶に基づいて、一定の自己による動き（Eigendynamik）を発展させることがある。しかし、類推適用が正当化されるのは、匹敵する利益状況がある場合だけである。BGB165条〔制限行為能力者の代理人〕によれば、少なくとも代理人の制限行為能力が要求される。これは現代のコンピューターシステムには当てはまらない。さらに、無権代理人の責任に対する取引保護のために設定されたルールが骨抜きになってしまうことも代理規定の適用を裏付けない。というのは、コンピューターシステムは権利の担い手でも責任主体（Haftungsperson）でもないから、瑕疵ある表示に対して自ら責任を負うこともできない。つまり、コンピューターシステムは自立性（Selbstständigkeit）が欠如するゆえに、代理人と同視できない。また、このシステムは、意思表示の内容規定を行うから、使者とみなすこともできない（S.354-355）。

⑤ 一般原則

クーン⁹⁾とメーリングス¹⁰⁾によれば、意思表示としてのコンピューター表示の問題は一般的な考え方に基づいて解答できる。つまり、ここで決定的なのは、コンピューター表示を人間の表示とみなすために、どの程度に人間の影響が必要かということである（S.355）。「必要なことは、表示行為が人間の意思に還元できることだけである。意思表示は、『延長された』——人間と機械とで分業的に完遂される——プロセスにおいても作ることができる。その場合、すでに自動的な意思表示の作成を目的にした意思による準備行為において必要な人間の協働行為（Mitwirkungsakt）が観察され

9) Kuhn, Rechtshandlungen mittels EDV und Telekommunikation, 1991, S. 65.

10) Mehrings, (o. Fh. 7) Rdnr. 24 ff.

るべきである。さらに、受け手にとって認識可能なように、当該表示を自己の意思表示として帰属させる意思が加わらねばならない。この要件は、コンピューター表示の事例では通常は問題ではない。なぜならば、コンピューター利用は、自動化された注文や予約システムに対して、取引締結を迅速化する意思でもって行われるからである。そこでは、表示の受け手は、コンピューターシステムの利用者が同人の包括的な (generelle) 意思によって把握されたコンピューター表示を同人の意思表示として効果帰属させることを望んでいないということを信じるきっかけがない。それによれば、コンピューターシステムによって完成される表示行為は利用者の意思表示として整序されるべきである。そこでは、法的な拘束は、コンピューターを自身の意思表示完成の道具として用い、特別な統制をあきらめているという利用者の法律効果意思に基づいている (S.355)。」

⑥ AI の表示との比較

AI による表示は、コンピューター表示のように、表示の具体的内容は AI 自身が確定するので、これはソフトが作成し表示したコンピューター表示と同視できる (S.355)。もちろん、AI の意思表示は、たとえば利用者が価格について特定の価格ではなく上限と下限だけ設定するように、具体化の程度はより少ない。これは AI が最終的に交渉を行うためである。また、AI は自律的な判断ができる (S.355)。しかし、これらの相違があっても、コンピューター表示と異なる整理をすべきではない。AI の自律的判断は、利用者が AI に対して定めた目的システムの範囲内でのみ可能であり、つまり、利用者は、表示がなされた場合に対して、正常に経過すれば、自らが設定した枠内にその表示が収まることを知っている。つまり、AI によってなされた表示は、利用者たる人間の表示に還元される。また、AI による表示は利用者に効果帰属すべきである。表意者は、AI が利用者を代表して交渉していることで、表示に自身が拘束されるという正当な信頼を受け手に生じさせている (S.355)。

⑦ 錯誤について

A. ツインマーマンによれば、道具説に立つ場合、人間の入力が直接に表示に取り入れられる場合に限り、取消可能な錯誤が生じる。つまり、自律的な AI の表示においては、定義上こんなことは起こりえず、事実上、取消は排除されることになる¹¹⁾。

2. 独自の信頼責任説

ドルニス は、信頼責任の一般原則から契約締結の際の AI 利用に対する新たな信頼構成要素を導くことを提案している¹²⁾。ここでは事業者の AI 利用が念頭に置かれている。同人は以下のように主張する。

① AI の技術的自律性

AI の自律性は、AI の契約締結を支持するあらゆる立場の出発点だが、その分析と描写ははっきりしていない。AI システムはもちろん、その利用者から全面的に解放されてはいないので、人格的にも自律的ではない。しかし、アルゴリズムによる行為が機能的に指示された行為枠組み内での人間の統御から切り離されていることを示すことができる。これは学術的構成にとって決定的である (S.725)。

② 他 律 性

トイブナーは、AI というデジタルのエージェントは自律的に行為するが、プリンシパルに従属するというデジタルのプリンシパル＝エージェント関係 (Prinzipal-Agenten-Beziehung) をローマの奴隷の立場と比較して、AI はデジタル奴隷であるが、超人的能力を持つ奴隷であると指摘している¹³⁾。し

11) A. Zimmermann, S.496.

12) Dornis, 717. 本節で引用する際は本文でページのみ挙げる。

13) Teubner, Digitale Rechtssubjekte?, AcP 218 (2018), 155, 162 (以下、断りがなければ「Teubner」として引用する)。プリンシパル＝エージェント関係については、伊藤秀史・

かし、両者の従属性 (Abhängigkeit) を同視することは誤解を招く。人間が原則として自由に自らの行為を判断できるのに対して、AI システムにとって自由の余地は限られている。つまり、AI システムは常に他律的に行為する (S.725-726)。

哲学やロボット工学では、本来の自律と非本来の自律、また技術的自律と人格的自律そして強い自律と弱い自律という対概念によって AI の自律性が特徴づけられる。本来の自律・人格的自律・強い自律という考え方は、「自ら選択したルールや法律の基準に従って、態度決定し、行為をなす」人の能力を表しており、そうした人格的自律の表現方式は、無数の行為の選択肢から自発的に選び出すことができる自由として表現される。奴隷ですら人格的に自律している。しかし、これと AI の自立性は区別されるべきである。確かに、AI の行為は人間の統御には服していないが、それは技術的なものに過ぎない。AI の自由の制限は確実に設定され、また克服もできない (S.726)。

③ 内的自律性

アルゴリズムによる自律性という現象は、法学的な整序にとっては、その帰結、特に AI の行為の説明不可能性と不透明さとして記述され、単にブラックボックス効果と呼ばれることも多い。AI の独立性は機械学習ないしディープ・ラーニングプロセスに基づく開発に特徴的である。AI システムは、部分的に自発的な進化プロセスを経て、最初の構想では詳細に予見されていなかった能力を開発する。外部的な枠組み内では、AI 技術に備わっているものとして、際限ないほどの自由が存在する。これはアルゴリズムの意思性や自己認識の内省的側面と混同されてはならず、この技術的な自立性は、システムに際限のない行動の選択肢と発展の選択肢を開いている技術設計 (Veranlagung) にのみ還元されるべきである。たとえば、ス

小林剛・宮原康之『組織の経済学』（有斐閣、2019年）136頁以下参照。

マート冷蔵庫のアルゴリズムも特定の食品が少なくなると補充することだけを目的にはしていない。それは自動化されたシステムで担保される。スマート冷蔵庫の自律性はそれに尽きず、日々消費される食糧を、可能な限り低価格で、持続的にあるいは環境に配慮して、ひょっとすると、想定を超えて（überraschend）、整えることが重要である。たとえば、季節的あるいは個人的な特性を考慮したり（クリスマス、誕生日、家族のお祝いなど）、システムの独自のインターネット検索によって、最新の「食べ物のトレンド」を考慮することすらあるかもしれない。このように行われた取引は、個々に見ても、具体的にあらかじめ定められていないし、人間が統御することもできない。AI システムが商品だけでなく、提供者・条件（価格、取得時点など）も自由に選択できるならば、契約の在り方は無限である。利用者が想定した内容の行動枠内で AI システムは包括的かつ無限に技術的に自律している（S.727-729）。

④ AI の行為から利用者の意思を切り離す

本来の意思形成と意思活動は、人格的な自律が欠如する AI では語ることができないが、同時に、技術的な自律性により、AI の行為は利用者の意思から切り離されることが確認されるべきである。この AI の人間からの解放（Emanzipation）は、自動化と自律の違いも説明している。古典的決定論的ソフトによって動き出す自動化された注文補助と違い、技術的に自律的な AI は、原則として結果が決定されていない進化を経る。AI の行為の発展と具体化（Ausprägung）は、前もって決定されていない上に、それは個々の点においても予測できず、あるいは後から説明することもできない。結果として生じる偶然性（Aleatorik）により、システムの構想と利用者による利用開始は、人工知能による行為の創発的な結果から包括的に分離される（S.729）。

⑤ 法体系的な整序

純粹に道具として用いても技術的な自律性のある AI の行為から人間の意思が切り離される点で、AI の表示を利用者自身の意思表示として法律行為において効果帰属させることはうまくいかない。つまりそれだけでは契約は有効に成立しないが、利用者は、特定の前提のもとで、AI によって基礎づけられた信頼に基づいて責任を負う (S.729)。

⑥ 意 思 表 示

意思表示は法律行為論の前提である。意思表示の効力発生仕組みにとって必須であり、かつ現代の学説でも確立している意思と表示の原則的な結びつきは、技術的に自律性あるシステムを用いる場合には、もはや存在しない (S.730)。

⑦ コンピューター表示

いわゆるコンピューター表示では、意思表示の主観的構成要素(行為意思、表示意識、効果意思)も客観的構成要素(外部への表明)も存在する。利用者が表示の時点では影響を持たないにもかかわらず、自動化されたシステムは人間の統御から解放された自由な裁量を持っていない。それは技術的には自律性がなく、人間の意思の定式化と表明の時点が先送りされているに過ぎない。プログラムの進行中に出力された意思表示は、何であれ、利用者によって前もって定められているがゆえに、予見可能である。学説では、いみじくも、意思表示の貯蔵庫 (Reservoir)¹⁴⁾ や意思表示のプール¹⁵⁾ と呼ばれる。自動的なシステムの行為の意思表示は潜在的に行われたあらゆる表示と直接結びついている。つまり、これらの表示のいずれも、AI 起動の時点で、利用者の行為意思、表示意識及び効果意思が把握されている (S.730-

14) Spindler in Hilgendorf (Hrsg.), Robotik im Kontext von Recht und Moral, 2014, 63,64.

15) Grapentin, Vertragsschluss und vertragliches Verschulden beim Einsatz von Künstlicher Intelligenz und Softwareagenten, 2018, 90.

731)。

⑧ 行為意思も表示意思も欠如し、意思表示の発出がない

利用者に一般的または包括的 (generelle) 行為意思があると観念することは、AI の行為を直接統御する可能性の欠如という意味を見誤っている。AI はシステムに設定された限界内で他律的に作動するが、この行為の枠の内部領域で利用者の意思から切り離されている。利用者の行為意思は、システムの起動を含むことがあるに過ぎない。この時点後には利用者はもはや AI の行為に干渉して統御することはできない。これは BGB105 条 [意思表示の無効] 2 項の構図に類似している。意識のない状態での行為や一時的な精神活動の障害の場合に、支配可能性と包括的行為能力 (Handlungsfähigkeit¹⁶⁾) の欠如ゆえに、表示とその表示の作成者との意思によって担われた結びつきが欠如している。AI システムの起動は、人間の意思による統御とアルゴリズムによる行為とを切り離している。起動の時点から、利用者は実際には、包括的行為無能力である。従って、行為意思は欠如しており、表示意識と効果意思にも同じことが当てはまる (S.732-733)。取引保護や信頼保護のために、効果帰属の原因として過失が意思の代わりになると主張されている¹⁷⁾。このアプローチは基本的に疑わしい。やはり実際の意思が欠如するので、法律行為上の構成ではなく、行為に対する責任の割り当てのみが重要となる。この点を置くとしても、AI 利用という特別な場合には、潜在的に構成された表示意識を介した効果帰属はうまくいかないだろう。真の意思形成と意思による定式化が欠如するのみならず、特に意思表示の発出 (Begebung) も欠如する。法的取引において意思により表示をなすことは、

16) 行為能力 (Geschäftsfähigkeit) と不法行為能力 (Deliktsfähigkeit) の総称である (山田晟『ドイツ法律用語辞典〔改訂増補版〕』(大学書林、1983年))。

17) Hübner, Festschrift Nipperdey, Bd. 1, 1965, 373, 388; Schapp, Grundfragen der Rechtsgeschäftslehre, 1986, 42; Musielak Zum Verhältnis von Wille und Erklärung. Eine Auseinandersetzung mit dem Tatbestand der Willenserklärung AcP 211 (2011), 769, 794.

受領が必要な表示では有効要件である (S.733)。AI 利用の場合には、利用者は、あらかじめ指示した行動枠組みに基づいて AI の取引を一般的には意識していると言えるが、利用者は AI を稼働することでもって、個々の取引への意思的形成の影響力を放棄している。その後に AI が生成したいかなる表示にも利用者による意思形成及び定式化が欠如しているだけでなく、法的取引において意思によって表示をなすことも欠如している。単に表示意識が欠如している場合とは違い、意思によって担われた外部に対する利用者の意思による行為が欠如している。結果として、直接的な法律行為による効果帰属も失敗する (S.734)。

⑨ 個人による委任の限界

仮に、AI システムの起動だけで効果帰属を可能にするなら、逆に、代理権付与の際に権利発生 (Geschehen) の統御を放棄した時点で、なぜその後に続く代理人の行為に法律行為上の拘束力が発生しないのかが問われなければならない。ドイツ民法の代理人行為説 (Repräsentationstheorie)¹⁸⁾によれば、法律行為の構成要素は、本人に効果が帰属するにもかわらず、常に代理人の人格 (Person) において実現しなければならない。代理人は単独で法律行為を行う。従って、客観的な表示の構成要素は、代理人の有効な意思表示に基づいてその権限内で生じた場合にのみ、代理権授与者を法律行為において拘束する。フォン・トゥールによれば¹⁹⁾ 行為無能力者にも代理権を与えることができる。本人を拘束するのは、行為の時点で無能力者ではなくなっている場合に限られる。無能力の状態で行った行為は本人

18) MüKoBGB/Schubert, 10. Aufl. 2025, BGB § 164 Rn. 11によれば、代理人行為説とはドイツの代理制度の基礎になっている理論で、およそ次のようなものである。代理においては、法律行為するのは代理人自身であり (BGB166条1項)、その意思表示が本人に効力を及ぼす。代理とは無関係に法律行為は成立するが、本人への効果帰属は代理権に依存し、有効性や取消可能性の判断基準も原則として代理人である。もっとも、本人が実質的に自ら行為している場合には例外が認められる (BGB166条2項)。

19) V, Tuh, Allg. Teil des Deutschen Bürgerlichen Rechts, 1918, Bd. II/2, 388.

を拘束せず、信頼責任の形で効果帰属するに過ぎない（S.734-735）。技術的に自律した AI 利用も、本人に法律行為によって効果が帰属しない。AI の創発は、行為無能力者の行為と異ならない。委任（Delegation）に続く「媒介者」の自己の意思決定による有効な権利形成という行為が欠如している。ドイツ民法の法律行為論において、法律行為のやり方のための組織的秩序は、閉鎖的に規範化されており、かつ私的自治による拡張も許されず、それは白紙（Blankett）の事例を除けば、「意思による代理」（他人や他の物への委任）のための代理以外の法形象を含んでいない（S.735）。

⑩ 白紙表示

代理のルール の 直接適用も準用も うまくいかない場合、白紙表示（Blanketterklärung）の原則²⁰⁾ に立ち返ることは最後の手段として驚くには値しない。AI 利用のプロセスは、白紙を他人にゆだねることと似ている。AI の利用者は、法律行為上の表示を世に送り出すことを意図してはいるが、その内容を自ら確定することはない。両者とも、表示は白紙の発行者あるいは AI 利用者の包括的なコミュニケーション意思及び参加意思（genereller Kommunikations- und Partizipationswille）に基づいているように見える。一見

20) 白紙と錯誤の問題は、およそ次のようなものである（MüKoBGB/Armbrüster, 10. Aufl. 2025, BGB § 119, Rn.59）。ある者が、白紙署名の証書を作り、それを他人に渡す場合、同人は、それが合意に反して補充されることを考慮しなければならない。つまり、同人は権利外観責任の一般的な原則に照らして濫用のリスクを負担する。たとえば、K が署名した用紙は B 銀行への消費貸借の申込であり、V は 5000 ユーロと補充しなければならないところ約定に反して 7000 ユーロと補充し、B がその申込を承諾する場合である（Brox/Walker, AT des BGB, 46. Aufl. 2022, § 18, Rn.17）。BGB119 条による取消は認められず、BGB172 条により、書面による代理権証書に基づいてその代理権の存立を信頼した者が保護される（判例・通説）。実際に、白紙に含まれた意思表示の存立を信頼する者は、約定に反して補充され流通した白紙であることがその者に明らかではないがゆえに、少なくとも、書面による代理証書の観点で代理権の存立を信頼している者と同様に保護に値する。ここから区別されるべきは、白紙の所持者が自らが補充する際に錯誤に陥っている場合、白紙の発行者に錯誤取り消しの権利が認められるべきかどうかという問題である。この場合、BGB166 条 1 項の準用（MüKoBGB/Schubert, 10. Aufl. 2025, BGB § 172 Rn. 3 は類推適用）により取消が肯定されるべきである。

すると、白紙による構成は説得的であるが、決定的違いがある。白紙は、白紙の発行者の不完全な意思表示であり、それを補充者が完全化する（カナリスによれば仕上げる（*perfiziert*）²¹⁾）に過ぎない。AI 利用は、利用者の意思表示が欠如しているためにこれとは区別される。AI 利用者は、一度も意思を定式化していないか、表明もしていないので、文書による白紙に匹敵するような最初の段階では不完全な表示の入れ物（*Erklärungshülse*）も欠如している。従って、契約締結のための AI 利用は直接本人に効果は帰属し得ず、BGB164条〔代理人の表示の効力〕以下においても位置づけられず、あるいは白紙の表示の構造にはめ込むこともできない（S.736-737）。

⑪ 意思表示の人間中心主義

ドイツ民法の法律行為論の意味での意思を作ることができるのは人間のみである。契約締結のための AI 利用にとっては、客観的な表示の構成要素の惹起ないし働きかけだけでは、利用者を法律行為により拘束するには十分ではない。法的安定性の必要を指摘してこれを批判することができるかもしれないが、学術的な構成が実用主義の下に置かれてはならない。むしろ、契約を主張する者には、法律行為による拘束の根拠づけをする義務がある。そこでは過度に契約の尊重（*favor contractus*）を強調する実務的な契機はない。法実務では、AI 利用から帰結される有効な契約の拘束力は、すでに解釈論として（*de lege lata*）担保される。例外的な場合にのみ、AI 利用者の信頼責任を問えばよい（S.737-738）。

⑫ 権利外観責任と白紙との違い

AI が生成した取引において契約の拘束力が必要だという実用主義の異論は、学術的な断絶を乗り越えていない。もちろん、付随的な約定（たとえばオンラインプラットフォームの利用者の利用規約）によって契約の拘束力を確実に

21) Canaris, Die Vertrauenshaftung im deutschen Privatrecht, 1971,56.

することはできる。そういう合意がない場合、代替的な構成、特に権利外観の責任が問題になる。白紙の発行者の責任に立ち返ることが学術的な構成に対して最初の手がかりを与えている。たとえば、白紙の責任のように、相手方の処分行為（Disposition）は、AI への信頼においても、行為者の背後にいる責任者としての利用者による契約の履行の期待から帰結される。AI 利用者も、白紙の発行者と同じく、信頼相当の原則（Grundsätze der Vertrauensentsprechung²²⁾）に従って、契約に類似して履行責任を負う。もっとも、白紙責任の原則を直接適用することは、具体的な権利外観の結びつきを考慮すれば、考慮に値しない。むしろ、AI の信頼責任の特別な構成要素を形成すべきである（S.738）。AI の利用では、白紙とは異なり、完全化され、伝達されるべき利用者の意思表示が欠如している。しかし、利用者は、諸事情のもとで相手方に対して、保護に値する信頼を生じさせている。もっとも、利用者の独自の表示行為は権利外観の結節点としては役立たない。決定的なのは、AI の具体的な行為表現である。たとえば、Eメールのメッセージ、ウェブサイトあるいはチャットボットのコミュニケーションの枠組みでの行為表現である（S.738-739）。

⑬ 信頼の構成要素と効果帰属

技術的に自律したシステムの利用にもカナリスの信頼責任の一般的メルクマール²³⁾は当てはまる。ここではBGB172条〔代理権証書〕に立ち返る必要がある。同人は、権利外観の最低条件として、関係する取引集団における「通常の慣行（übliche Gepflogenheiten）」に則ることを提案している。たとえば、自書ではなく印刷されファックス送信された署名も、その集団で受け入れられているなら、そこからも権利外観が生じるという²⁴⁾。判例も

22) いわゆる積極的信頼保護のことで、信頼者が想定した状況が実際に存在するとしたならば発生したはずの法律効果が生じるというものである（Canaris, (o. Fn. 21), 521 ff.）。

23) Canaris, (o. Fn. 21), 49 ff.

24) Canaris, (o. Fn. 21), 59 f.

同様である²⁵⁾ (S.739)。取引における慣行に依拠している AI への信頼の客観的要素として、問われるべきは、契約の準備や締結の際に（たとえば E メールメッセージの発信やチャットボットの利用の際に）、アルゴリズムが「通常」用いられているかどうか、である。その際、相手方が AI と交渉していることを認識しているかどうかは区別されない。説明の指示がなければ、AI の利用は、通常は人間の行為だと理解されてよい。そこでは、相手方の信頼は、人間の商事代理人及び代理人 (Handlungsbevollmächtigte und Vertretungspersonen) に対する責任負担 (Einstehen) という期待に基づいている。これに対して、利用者が、AI 利用を明らかにする場合（これについては法律上義務を負うこともある）、相手方は、権利能力も行為能力もない AI が「権限をもって」行為すること——端的に言えば利用者が AI によって行為し拘束される意思があることを出立点とするだろうことは正当である。市場での AI 取引の信頼保護は、HGB56条に類似して正当化される (S.739-740)。

効果帰属の問題については、最終的に、利用者が AI の起動でもって法的取引に対して、白紙の表示の濫用の危険に匹敵するリスクを生じさせていることが顧慮されるべきである。つまり、利用者は、白紙の発行者と同じやり方で、権利発生 (Geschehen) に関する支配を与えており、両者において、その後の権利発生を支配できていない。従って信頼責任の主観的構成要素も満たされている (S.740)。

⑭ 拘束力と AI の機能性

AI の行為に拘束される範囲も信頼責任の一般的ルールに則るべきである。特に利用者が知らずに意欲もせずに起動した場合、権利外観を意識して作出したものとはみなされず、効果帰属も問題にならない。第三者によって AI が操作された場合も同様である。もっとも、意思によって起動したなら、

25) BGHZ 40, 297 (304 f).

AI 利用者は、主観的に意欲していない契約内容だとして、BGB119条〔錯誤に基づく取消可能性〕に基づいて取消することはできない (S.741)。AI の誤作動の場合、特に設定枠組みを超える場合、どうなるか解明が必要である。たとえばスマート冷蔵庫が利用者の望んでいない商品を購入する場合、不具合が瑕疵あるデータ資料の使用に基づいている場合は、意思形成の欠陥であり、錯誤取り消しは認められないとされる²⁶⁾。また、プログラムの誤りやシステムエラーが機械学習に影響した場合、動機錯誤に匹敵し、取消はできないとされる²⁷⁾。こうした AI の行為の法律行為上の効果帰属を支持する者によって挙げられた取消の限界は、信頼責任にも妥当せねばならない。つまり、利用者は、システムに特殊なリスク（不具合など）に対しても責任を負う必要がある。利用者は、そのコストを公衆（Allgemeinheit）に転嫁してはならない。その場合のみ、AI の機能を改善するための技術革新の動機付けが効果的に担保される（S.741-742）。もっとも、自律性のリスクに対する責任は、権利外観責任により一般的に限界が画される。つまり、信頼は、特に、思慮分別のある取引参加者が構成要素の不正確さを知っていたか知らねばならなかった場合には保護されない（BGB173条〔悪意あるいは過失による不知の場合の効力存続〕類推）。これは、とりわけ、明白な AI の不具合の場合に妥当する。それ以外の場合には、客観的な手がかりが問われるべきである。一部で認められる見解とは異なり、その場合、利用者にとっての AI の取引の「予見不可能性」は重要ではなく、決定的なのは、相手方の視点のみである（S.742-743）。

26) たとえば、BGH NJW 2005, 976, 977では、表意者が正しく入力した数値（2,650€）がソフトに欠陥があるために245€と変造され受領者に伝達された場合は表示錯誤だが、計算ソフトの誤りなどで入力すべき数値を誤っていた場合は動機の錯誤となるという。

27) たとえば、Köhler, Die Problematik automatisierter Rechtsvorgänge, insbesondere von Willenserklärungen, AcP 182 (1982), 135は、プログラムの誤りは様々な原因があるが、いずれも、事態に照らして考えれば、いわゆる計算錯誤でしかなく、原則として顧慮されない動機錯誤として整理されるべきだと言う。

⑮ 錯誤について

この見解によれば、AI を用いた取引では基本的に錯誤取り消しは認められないと考えられる。また、AI が設定逸脱行為を行っても動機錯誤として顧慮されない。ただし、相手方が AI の明白な不具合について悪意または善意有過失であれば、利用者へ効果が帰属しない（おそらく契約不成立となる）。AI 利用契約が合意により有効となる場合についてはどうなるか不明である。

3. 代理人説

これは、AI に権利能力を認めて代理に関する規定を用いて処理する見解である。ここでは、A. ツインマーマンによる整理と同人の見解を紹介する²⁸⁾。

① 代理と AI の権利能力

この見解は、AI の自律性を強調し、BGB164条〔代理人の表示の効力〕以下の代理に匹敵する利益状況が生じ、これらの規定の類推適用が許されると考える（S.497）。ここで引き合いに出されている、トイブナーは、これに関して次のように言う²⁹⁾。「経済実務における現実の機能に正確に対応させて、経済学的に言えば、プリンシパル＝エージェント関係において、AI は法的に自己責任により意思表示をなすが、それは自己の名ではなく、プリンシパルの名においてなすのである。コンピューターの表示は、それ自体法的に拘束力があり、直接にプリンシパルを義務付ける。代理の法を類推して、自律的な AI がその人間のプリンシパルの代表者として扱われるべきである」。もちろん、AI には人間の代理人とは異なり、いまのところ権利能力はないが、この見解の基礎にはエールリッヒによる権利能力は機

28) A. Zimmermann, S.490. 本節で引用の際は本文でページのみ挙げる。他に、トイブナー論文や Schirmer, Rechtsfähige Roboter? JZ 2016, 660がこの見解を支持している。

29) Teubner, S.181f. ここでは Softwareagent という語が用いられているが、便宜上 AI と表記する。

能概念として分割できるという考え方³⁰⁾がある（S.497）。つまり、代理権限という意味での分割された権利能力を取り出して考えることができる³¹⁾。

② 白紙表示

しかし、代理規定類推の支持者は、BGB164条〔代理人の表示の効力〕以下の準用の過程で、いくつかの修正が必要なことを認めている。用いられているアプローチは、AIの表示を白紙の表示と同一視する（S.497）。たとえば、ノイナーは次のように言う³²⁾。「AIは通常は使者でも代理人でもない。単なる使者となると、AIの創造的機能に合致しない。ただし、AIが表示の受け手の決定的観点からみて、純粋な伝達者として現れ、意思形成の関与者として現れたのではない場合は別である。代理人として特徴づけること（あるいはBGB164条〔代理人の表示の効力〕以下の類推適用）は、同じく事態に即していない。なぜなら、BGB179条〔無権代理人の責任〕に基づく潜在的請求に対するAIの責任財産は（少なくとも今は）欠如しているので、取引保護が空洞化するし、特別な保護が必要なのはBGB179条〔無権代理人の責任〕3項の未成年者のみだからである。解釈論としても（*de lege lata*）、自己の意思表示の主観的構成要素の点で大きな矛盾が生じてしまう。その代わりに、学術的には白紙表示の原則に従った効果帰属が考えられる。両者の場合には、白紙の発行者またはAI利用者は、法律行為上の表示を意図しているが、その段階で具体的内容を自ら確定してはいない。意思表示は、事前の作業や事前の設定（コンピューターの起動、ネットワーク接続）を基礎としており、また包括的なコミュニケーション意思及び参加意思に基づいているので、表示の補充や具体的な作成が委任されるとして

30) Ehrlich, Die Rechtsfähigkeit, 1909, Neudruck 1973, S. 1 f. エーアリッヒ訳／川島武宜・三藤正訳『権利能力論〔改訳〕』（岩波書店、1975年）14頁以下参照。

31) Teubner, S.182も、法人としての完全な権利能力を与える必要はなく、機能的な観点から単に部分的な権利能力、つまり代理能力で十分であると指摘する。

32) Neuner, BGB AT, 13. Aufl. 2023, § 50 Rn. 115f. ここでは Softwareagent/Agent という語が用いられているが、便宜上 AI と表記する。

も、原則として発行者や利用者に効果が帰属する。AI がその利用者によって定められた限界を超える場合、白紙表示で指示に反した補充がされる場合のように、BGB172条〔代理権証書〕2項を類推して権利外観責任が考慮に値する（錯誤取消の選択肢はない）。AI のそのような関与に対しては、BGB165条〔制限行為能力者の代理人〕の規律も妨げとならない。これは、（実態および体系に違反した）行為無能力者が知能の不足を理由に、代理人として法取引へ参加することを完全に一律に排除しているだけだからである。これに対して、AI は人工知能を備えており、それは法律行為における意思形成に関与する能力がある。BGB172条〔代理権証書〕2項を類推した権利外観責任によって表示の受け手が包括的にかつ行為無能力の場合よりも手厚く保護されることもこれに加えられる。それゆえ、この利害状況は、AI の表示の不確定無効を裏付けはしない。（もっとも、この見解でも AI 自体が錯誤した場合には取消可能である（脚注20参照））。準代理（Quasi-Stellvertretung）を認めると主張する見解との違いは、さほど大きくない。つまり、白紙の補充にも BGB164条以下が準用される³³⁾ (S.498)。

③ 代理アプローチと白紙アプローチ

代理アプローチ及び白紙アプローチは、取消可能性に対して明白な法定の規律モデルを提供している。つまり、BGB166条〔意思の瑕疵；悪意の場合の帰責〕1項の類推適用は、意思の瑕疵にとって原則として AI が基準となる。顧慮される AI の誤りがあると、利用者が表示内容について具体的な観念を有していなかったとしても、契約は取り消される。利用者の意思は重要ではなく、特に、締結された契約と利用者の期待（それが具体的なのだとしても）との乖離があっても、それは取り消しを正当化しないだろう。利用者が誤って入力したデータは、AI において表示錯誤ないし内容錯誤（BGB119条〔錯誤に基づく取消可能性〕1項）、あるいは性状錯誤（同条2

33) MüKoBGB/Schubert, 10. Aufl. 2025, BGB § 172 Rn. 2. なお、白紙補充は他人の名前での行為が欠如するので代理ではなく、事実行為であるとされる (aa.O)。

項）を引きおこす場合にのみ、取消を正当化するだろう（S.498）。

④ AI の誤りに基づく取消可能性の見解

AI の誤りに基づく取消可能性の見解は多様であるが、少なくとも取消を完全に排除するという立場は、有益ではない。取消権に関する法定のリスク分配に適していないためである。それに対して、代理アプローチによる BGB166条「意思の瑕疵；悪意の場合の帰責」の類推適用は、適切かつ分化した解決のアプローチを提示している（S.498）。

⑤ 道具説への批判

錯誤の問題は、自律的な AI の表示を自動化された表示と同じように扱う点で疑問をより強くする。そこでの顧慮される誤りと顧慮されない誤りとの区別は空っぽである。AI に特有の本来的な問題は、AI の自律性から生じる誤りをどう扱うか、という問題だからである。これは自動化された表示の学説では重要ではない。なぜなら、そのような表示は、固定してプログラムされたシステムを対象として発展してきたもので、そうしたシステムは独自の誤りを犯すことはあり得ないからである。自律的 AI を用いる場合の決定的な誤りの類型は、自動化された意思表示に関する原則によって、盲点（blinder Fleck）である。同時にこのアプローチは AI 利用者のリスク負担を、古典的な自動化の利用者に比べて、十分な理由なく高めてしまっている。古典的な自動化された表示では、少なくとも、表示内容に対する人間の直接の指示が明らかに頻繁に行われており、また指示に誤りがある場合は、取消権が認められるからである（S.498-499）。

⑥ 独自の信頼責任説への批判

信頼責任アプローチはそれ自体優れているが、私見によれば、ドイツ民法の価値判断には適さない。たとえば、民法は、代理人に錯誤がある場合、BGB166条「意思の瑕疵；悪意の場合の帰責」1項の取消を明示的に認め

ている。つまり、本人は、自らの補助者の誤りのリスクを無制限に負うことを出立点にしておらず、本人が代理人を注意して選任または監督 (ausgebildet) していたことを、取消可能性に対して要求もしていない。一般的にも、取消権は取消者が注意して行動するかどうかによって左右されない。民法起草者は、自覚的³⁴⁾ かつ長い議論³⁵⁾ を背景に重大な過失がある錯誤ですら取消原因として認めることを決定していた。契約から解放されるという取消者の利益を、有責性のある錯誤の場合ですら、契約を維持するという相手方の利益よりも重視している。——たとえば代理人のより適切な選任ないし監督 (Ausbildung) により——意思の瑕疵を回避するための動機付けとしてドイツ民法は、消極的な利益の賠償義務 (BGB122条 [取消者の賠償義務]) で十分だとみなした。これを前提にすれば、AI の誤りにおいて、契約の維持および契約で定められた積極的な利益の維持という形での強い動機付けを要求する十分な契機はない。むしろ、現行法のリスク判断の規則的な適用にとっては、まさに逆に、AI の表示においても少なくとも取消を原理的に認めることが要求されるように思われる。もちろん、技術的に未成熟な AI に対しては、取消の完全な排除よりも別の方法で対応することができる (S.499-500)。

⑦ BGB166条 1 項の類推適用と取消可能性

AI の表示の原理的な取消可能性を確実にし、かつ同時に AI 利用者に過分の特権を与えないために、最も説得力あるアプローチは BGB166条 [意思の瑕疵：悪意の場合の帰責] 1 項の類推適用である。当該規定は、取消にとって、表示を形作る意思を形成した実在 (Entität) の意思に照準を合わせている。ここでは、AI である。BGB166条 [意思の瑕疵：悪意の場合の

34) Mudan, Die gesamten Materialien zum Bürgerlichen Gesetzbuch für das Deutsche Reich, Bd., I, 1979, S. 712-716.

35) Savigny, System des heutigen römischen Rechts, Band 3, 1840, Beylage VIII, S. 335 und S. 264.

帰責] 1 項の錯誤への適用は、AI による表示に一般に代理法を適用するという立場も支持する。しかし、同条の適用は、少なくとも、AI と代理人を完全に等価なものと認めることと必ずしも結びついているのではない。契約への拘束を自動化された表示を類推して構成する者は、技術の自律的な誤りにおいてはその理論の欠点に出くわすが、その欠点は事態に即して補うことが重要である。そのためには、BGB166 条 [意思の瑕疵；悪意の場合の帰責] が適している。同条を代理法を超えて拡張することは全く新しいことでもないだろう。実務では交渉補助者の事例に類推適用されている (S.500)³⁶⁾。

⑧ 例 外

この見解には、システムに内在する二つの例外がある (S.500)。一つは裁量の余地のない AI (KI mit gebundener Marschroute) である。利用者が例外的に指示を与え、それが直接表示に取り込まれ、その限度で AI から自律性が奪われる。BGB166 条 [意思の瑕疵；悪意の場合の帰責] 1 項のルールは、代理の場合の表示が代理人の意思行為であることでもって正当化されている。しかし、決められた筋道 (Marschroute) が指示されている限りで、表示においては代理人ではなく、本人の意思が実現する。BGB166 条 [意思の瑕疵；悪意の場合の帰責] 2 項の考え方によれば、裁量の余地のない人間の代理人については、本人の意思の瑕疵に基づく取消が認められる。これは AI による意思表示にも転用されるべきである。それは純粹に自律的な AI の表示ではなく、自律的な表示と電子的に伝達された表示または自動的な表示の融合体 (Amalgam) であるという考え方に支えられている。たとえば、AI 利用者が、固定価格の入力の際に打ち間違える場合、BGB119 条 [錯誤に基づく取消可能性] 1 項 2 選択号に従って取消権が生じる (S.501)。

36) MüKoBGB/Schubert, 10. Aufl. 2025, BGB § 166 Rn. 19ff. 状況を唯一人把握している補助者に売主が交渉や実際の契約締結を完全に委ねる場合、同人は、BGB166 条 1 項に従って、補助者が知っていることについて帰責されねばならない (BGH NJW 1992, 900)。

もう一つが AI の技術的な不完全さ (Technische Unzulänglichkeit der KI) である。代理法は、技術的に未熟な AI から保護するアプローチを法的取引に提供している。人工知能とはいっても、技術的に未成熟なシステムも含む。独自の信頼責任説のように、取消を排除することは、原則として AI 利用者に技術的な改善を動機づける確実な手段ではない。しかし、AI 利用者は、技術的に明白に不十分な AI のリスクを負担するべきである。これは取消可能性のない契約への拘束を必要とする。そのような評価は行為無能力者を代理人として用いることを排除している BGB165条 [制限行為能力者の代理人] から引き出すことができる。民法は、代理人に意思表示をするための知的能力を要求している。一見すると、同条の法律効果が取消可能性のない契約への拘束に至るということに疑問が生じるかもしれない。通常は、無能力者を代理から排除することは、契約が初めから無効であることを意味するからである。しかし、これには例外がある。行為無能力者であると認識されていない有限会社の経営者の事例³⁷⁾で、権利外観の観点から契約への拘束を出発点としている。このアプローチは AI に転用できる。契約の当事者は、AI の技術的な性質を通常は判断できないし、技術的に十分な AI という外観は利用者に効果帰属させることができ、その結果、権利外観の観点が契約の拘束を正当化する³⁸⁾。もっとも、AI の技術的な性質への要求水準は低く設定されるべきである。これに対して基準となる BGB104条 [行為無能力] 2号でも、行為無能力は極めてまれな例外事例だからである。たとえば、ある者がその表示の射程を詳細に予測できないとしても、同人には行為能力がある。わずかな程度の理解力しかなくてもよい。また、行為能力にとって必要な能力は包括的である必要はなく、部分的な行為能力でも考慮に値する (S.501-502)。

37) BGH NJW 1991, 2566 (2567). 本件では行為無能力者ではなく会社が権利外観責任を負うので、行為無能力者保護の利益のために権利外観を排除するという一般的に認められた考え方は問題にならないという (Abs.5)。

38) A. ツインマーマンはこの点でドルニスに従うと述べている (Fn.95)。

Ⅲ．むすびに代えて

上記で紹介した見解をまとめて、AIの錯誤という観点から日本法への示唆を考えてみたい。

1．各学説のまとめ

道具説によれば、AIによる表示はコンピューター表示と同様に利用者の意思表示に還元されるため、法的効果は最終的に利用者に帰属する。しかし、ここで問題となるのは錯誤取消との関係である。錯誤取消が認められるのは、人間の意思が直接に表示へ反映される場合に限られる。しかし、AIの自律性からこれは起こりえないので、AIの錯誤による取消は事実上排除される。

独自の信頼責任説（以下「信頼責任説」とする）では、AIの技術的自律性に基づき、その行為を人間の意思と切り離すため、AIの行為により直ちに契約は成立しない。AIは包括的行為無能力者とみなされ、その表示に対応する利用者の意思がないからである。しかし、白紙表示と利害状況が類似し、BGB172条〔代理権証書〕を介して、権利外観による保護が相手方には認められる。ただし、相手方の信頼の対象はAIのEメール等での具体的行為表現となる。これによれば、AIの錯誤に基づく取消は認められず、意思によらないAI利用の場合、あるいはAIに明白な不具合があり、取引相手がその不具合を悪意または善意有過失で利用した場合に、BGB173条〔悪意あるいは過失による不知の場合の効力存続〕類推により、効果帰属が否定されるだけである。もっとも、事前に合意でAI利用の契約の有効性は確保できるので、信頼責任による保護は例外である。

代理人説³⁹⁾は、AIを自律的に行為し得る存在として捉え、BGB164条以

39) 代理人説の支持者は少なくない Dornis, Fn.17; A. Zimmermann, Fn.55参照。

下の代理規定を類推適用する見解である。AI には権利能力はないが、機能的に分割された権利能力としての代理権限を認めている。

第一に白紙表示アプローチでは、AI 利用者は表示を意図するが、その内容を確定せず、AI に補充をゆだねる。同人には包括的な関与の意思があるので、効果は利用者に帰属する。AI が定められた限界を逸脱する場合は BGB172 条〔代理権証書〕2 項の類推により権利外観責任が考慮される。ここでは錯誤取消は選択肢とならない。ただし、AI 自体に錯誤があれば、BGB166 条〔意思の瑕疵；悪意の場合の帰責〕1 項の類推・準用により取消は可能となる。

第二に代理アプローチでは、AI を代理人と考える。交渉補助者にも BGB166 条〔意思の瑕疵；悪意の場合の帰責〕1 項が適用されることに照らして同条の類推により、意思の瑕疵において AI を基準とする。AI に錯誤があれば、利用者が具体的な観念を欠いていても契約は取り消しうる。ただし、利用者の指示により AI に裁量の余地がない場合は、本人の錯誤として扱われる。また、技術的に明白に不十分な AI の場合には、BGB165 条〔制限行為能力者の代理人〕の趣旨から、契約は取消不可能な形で拘束力が維持され、外観責任により効果帰属が正当化される。

2. 日本法への示唆

わが国で、AI を用いた取引を行い、AI に「錯誤」があった場合、95 条に基づく錯誤取り消しは可能であろうか。本稿では、詳細に検討する余裕はないので、以下では若干の問題提起にとどめる。

① AI の「錯誤」をどう考えるか

仮に AI の錯誤を認めるとしても、実際に問題になるのは、AI による設定逸脱行為があった場合に限られるだろう。たとえば、利用者が AI を用いて中古車甲を 200 万円で購入した場合に、利用者は事故車でないものを欲しており、AI は甲を事故車ではないと判断したが、実際は事故車であり、

市場価値は180万円であるという事例で考えてみる⁴⁰⁾。利用者がその旨の設定をしていないならば、AI が甲を事故車でないと思って購入したとしても、AI にとっての主観的因果関係を満たさないだろう。それは利用者の錯誤でしかない⁴¹⁾。仮に利用者が、これを AI の錯誤だと主張するとしても、後述のように AI への指示を怠ったことは重大な過失になりうる。従って、実質的には、事故車を除くと設定していた場合に、その逸脱行為があってはじめて「AI の錯誤」が問題になる。ベジタリアン A 所有のスマート冷蔵庫が、「ステーキが野菜に属する」と誤解して、ステーキを注文するような場合⁴²⁾ にも設定逸脱行為が見て取れる。

② 問題提起と試論

AI の錯誤では、このようなリスクを利用者が負担するか相手方が負担するかが問題となる。道具説、信頼責任説では、基本的に利用者が負担する。道具説では、AI の錯誤はそもそも観念できないし、信頼責任説では錯誤がプログラムエラーに由来するものでも顧慮されず、第三者が操作したような場合か相手方が当該エラーについて悪意・有過失の場合にのみリスクを負担（おそらく契約不成立となる）することになる。もっとも、道具説では AI の自律性を考慮してないという問題が生じる（Ⅲ2⑧、Ⅲ3⑤）。また、信頼責任説では、事前の合意により契約を有効とする場合、「AI の錯誤」をどう扱うのかは当該意思表示主体の決定を経なければならない⁴³⁾。

40) A. Zimmermann, S.506, 508を参考にした事例。なお、A. ツインマーマンは、確率判断という観点に人間の錯誤との利益状況の類似性を見出し、AI には意識はないが錯誤すると考える（S.506f）。

41) 本人が錯誤に陥っていても、代理人が錯誤に陥っていなければ無効を主張できない（札幌地判昭和42年2月24日判時485号58頁）。反対説として、於保不二雄編『注釈民法（4）総則（4）』（有斐閣、1967年）38頁〔浜上則雄〕。

42) A. Zimmermann, S.503

43) 基本的に主体は利用者である人だと考えるべきであろう。United Nations Commission on International Trade Law, UNCITRAL Model Law on Automated Contracting with Guide to Enactment (United Nations 2025) Part Two, Remarks on Art 7, para 61 (https://

代理人説は、論者によれば、AI の錯誤を直接問題にでき、それがドイツ民法の価値に合致し、利用者と相手方との利益調整に適しているとされる(Ⅱ3④⑥⑦)。しかし、わが国でこの説を採用する場合、AI に権利能力を認めるか否かという大きな問題を別にしても、なお複数の問題点が想定される。

白紙アプローチでは、白紙表示をわが国でどう位置付けるかが問題となる⁴⁴⁾。代理アプローチでは、101条をドイツのように拡張して適用するとしても(Ⅱ3⑦)、それ以外の代理規定の適用如何を検討する必要がある。たとえば、AI の設定逸脱行為が無権代理や表見代理規定の適用対象になるのか、である。

また、日本とドイツとでは錯誤規定が異なるため⁴⁵⁾、少なくともドイツ法と同じような当事者間の利益調整は働かない。まず、95条2項の表示要件が挙げられる。ドイツにはこのような規定はないし、一般に解釈上も要求されていない⁴⁶⁾。これに対して、わが国では、基礎事情(動機)の錯誤の顧慮にとって、この要件の充足を考える必要がある。AI を人間の代理人と同じと考えるなら、一般に動機が表示され法律行為の内容になっていることが必要だと解されている⁴⁷⁾。もっとも、人間の場合でもそうであるが、AI においても動機が表示され法律行為の内容になるということの意味は明確

uncitral.un.org/sites/uncitral.un.org/files/2424674e-mlautomatedcontracting-ebook.pdf
(最終アクセス2025年10月9日)。

44) 白井豊「白紙書面の濫用補充と交付者の法的責任(2・完)」立命366号(2016年)590頁以下参照。

45) ドイツの錯誤法の概観については、ディーター・ライボルト(円谷峻訳)『ドイツ民法総論[第2版]』(成文堂、2015年)267頁以下参照。

46) MüKoBGB/Armbrüster, 10. Aufl. 2025, BGB § 119によれば、動機錯誤は基本的に取消の対象にならない(Rn.108)。ただし、性状錯誤は例外的な顧慮される動機錯誤だと理解されており、そこでは錯誤が取引本質的な性状にかかわること、その錯誤が主観的に客観的にも重要であることが要件となる(Rn.113f)。また、外部的計算錯誤について計算を交渉の対象にした場合(動機の表示に相当)に顧慮されるというライヒ裁判所の考え方を、連邦通常裁判所は明示的には引き継いでいないとされる(Rn.93f)。なお、アルムプリュスター自身は別の考え方を支持している(Rn.120ff)。

47) 東京地判平成28年9月12日 Westlaw Japan 文献番号2016WLJPCA09128015。

ではない⁴⁸⁾。少なくとも、錯誤顧慮にあたってはAIによる（動機表示としての）表示価値のある何らかの行為が必要だと考える⁴⁹⁾。

さらに3項柱書の重過失の問題もある。ドイツでは表意者に重過失があっても取消が認められ（Ⅲ3⑥⁵⁰⁾、相手方にはBGB122条〔取消者の賠償義務〕に基づく信頼利益の賠償が認められるに過ぎない。また、A. ツインーママンによれば、利用者の過失（より適切な選任監督の懈怠）もAIの錯誤において考慮される（Ⅱ3⑥）。これに対して、わが国では、一般に代理人に重過失がある場合、錯誤取消は認められない⁵¹⁾。ここで問題になるのは、AIの重過失を観念できるか、AIの錯誤に対する利用者の重過失は取消を妨げるか、である。

動機錯誤で仮に2項要件を充足する場合や表示錯誤（スマート冷蔵庫がセンサーエラーで魚ではなく肉を注文する⁵²⁾）の場合には、無重過失要件が問題になる。重大な過失を、錯誤を回避すべき義務を怠ったことだと理解する場合⁵³⁾、プログラムエラーによる逸脱行為に基づく錯誤の過失をAIに帰することはできないだろう。たとえば、前述の事例で、AIが甲は20%の確率で事故車であると判断していたが、プログラムエラーによりこれを考慮する

48) さしあたり中谷崇「サブリース契約における賃料減額の合意の効力」現消62号（2024年）101頁以下参照。

49) 東京高判令和3年12月23日金判1666号8頁は、被保険者の居住地に関する保険代理店の錯誤に関して、代理店の用意した確認用紙記載の「海外居住の方……は、お引受けできません。」という文言は、動機の表示とは認定されていないものの、錯誤の因果関係（要素性）の判断の決め手になっている。

50) MüKoBGB/Armbrüster, 10. Aufl. BGB § 119, Rn57によれば、たとえば、表意者が契約書を読まずに署名し、証書の内容について誤った観念（特定の別の文書であるとの誤解など）を抱いていた場合、重過失があっても、内容錯誤に基づく取消が認められると指摘されている（RGZ 62, 201（205））。

51) 四宮和夫＝能見善久『民法総則 第9版』（弘文堂、2018年）368頁。

52) Schirmer, (o. Fn. 28), S.664.

53) 中谷崇「錯誤者の過失の位置づけ（1）」立命395号（2021年）209頁。無重過失要件が錯誤取消の可否の実質的判断基準となっていると評価できる事例もあるとの分析もあるため、積極的要件の如何にかかわらず、この要件の検討は必要になろう。

ことなく200万円で購入する場合である⁵⁴⁾。これは利用者（設定を怠る場合など）または設計者（設計段階での欠陥）の過失と捉えるべきである。利用者の過失だと考える場合、AIには重過失がないので錯誤取消主張の妨げにはならないと解釈すべきか、それとも利用者には重過失があるので錯誤取消の主張ができないと解釈すべきか。取消を認めるのは妥当ではないと考えられるが、その法的構成が問題となる。101条3項を類推適用する見解⁵⁵⁾と101条1項を縮小解釈する見解⁵⁶⁾がある。また、仮にプログラムエラーによらず錯誤が生じる場合、AIの過失だと考えることができるだろうが、AIは人間よりも錯誤回避の要請が強く働くと考えられ、重過失が認定されやすいであろう⁵⁷⁾。さらに関連して、相手方の錯誤の悪意・重過失（3項1号）や共通動機錯誤（同2号）も問題となる。以上のように、代理人説は克服すべき課題が多い。

③ま と め

AIによる「錯誤」の問題は、そこから生じる不利益の分配をどう考えるのかという価値判断とそれを解釈論としてどう落とし込むかという2段の論点が生じる。取引安全を重視し、AIの錯誤に基づく取消は認めない方向性での議論（道具説、信頼責任説）は十分あり得る。特に信頼責任説はAI利

54) 意識または警戒しているリスクに関する錯誤であり、AIが表意者ならば重過失の認定が認められやすいと考えられる（中谷崇「法律行為の内容とリスク分配」滝沢昌彦他編『円谷峻先生古稀祝賀論文集 民事責任の法理』（成文堂、2015年）144頁参照）。

55) 大阪高判平成17年4月28日判時1907号57頁。改正前101条2項の類推適用により本人に重過失がある場合は、錯誤無効の主張はできない。

56) 東京高判昭和40年8月4日東高民時報16巻7・8号139頁。代理人に包括的代理権があったとされる事案で、改正前101条1項により直接適用されるのは改正前95条本文のみであり、本人に重過失がある場合には代理人の錯誤無効を主張できず、代理人に重過失があっても改正前101条1項の類推適用により錯誤無効の主張は認められない（代理人の錯誤には95条本文しか適用されないが、それを主張する本人には95条全体が適用されるということであろうか）。

57) 代理人が錯誤する場合でも、重過失の判断にあたり専門家であるかどうか、つまり錯誤を回避すべき構造的な要請が考慮されている（前掲（脚注46）平成28年9月12日判決）。

用リスクの事業者負担を志向する。この場合、プログラムエラーの原因が AI 作成者にある場合は損害賠償請求で対処することになるだろう。他方、当事者間の利害調整を図る意味で、特に一般利用者が AI を利用する場合には、錯誤取消の可能性を原則的に認める立場もあり得る。もっとも、代理人説は、わが国では課題が多い。この方向で考える場合は、AI の錯誤を観念するのではなく、むしろ、AI 利用による 1 つの意思表示を、利用者と AI の分業で行われるものと考え（Ⅱ1⑤）、AI の「錯誤」を利用者の錯誤に還元する構成を考える方が望ましいのではなかろうか。この場合、信頼責任説が言うように、AI 起動により利用者が行為無能力になり、AI も行為能力がないと考えるとしても（Ⅱ2⑧⑨）、AI 利用を、故意または重過失により意思無能力者になる状態と同視し⁵⁸⁾、有効な契約が成立するとの構成が一つ考えられる。上記の中古車の例で言えば（Ⅲ2①）、AI の判断の誤りに基づく意思表示を利用者の意思表示とみなすことになる。もっとも、この場合でも錯誤取消が認められる事例は現実的には極めて少ないと思われる。

現時点では、わが国では AI の「錯誤」を独自に観念する必要性は必ずしも高くない。もっとも、ドイツ法が示すように、技術的自律性を法的にどう位置づけるかという点は、今後の民法理論における新たな論点となり得る。この方向で検討する場合、AI が「錯誤」し得るか、その立証方法をどう構成するかが問題となる⁵⁹⁾。これは、詐欺における被欺罔性でも同様である。この点は、今後の研究課題としたい。

58) 民法〈債権法〉改正検討委員会編『詳解債権法改正の基本方針 1 序論・総則』79頁以下参照。心理留保の考え方の援用が指摘されている。他に原因において自由な行為、信義則違反を根拠とすることが考えられる。

59) A. Zimmermann, S.503ff 参照。

