

AI と共生するためのポストヒューマニズムデザイン

Invisible Consent から多元的設計論へ

片岩拓也 (2025A015)

静岡大学 / HUMAI 第 1 期生

はじめに

本稿は、筆者が共同筆頭著者として取り組んだ研究「Invisible Consent」[\[1\]](#)を出発点に、人間と AI の共生を設計するとはどういうことかを考えるエッセイである。

Invisible Consent の研究は、LLM ベースのブラウザエージェントが Cookie 同意 UI を自律的に処理してしまう現象を実験的に明らかにした。この研究を通じて筆者が強く感じたのは、この問題がプライバシー設計の不備にとどまらず、人間と AI の関係性そのものに関わる構造的な問いを含んでいるということだった。本稿では、その直観を掘り下げ、生態学的共生の概念とポストヒューマニズムの視座を援用しながら、多元的な解釈が重なり合う設計という構想を提示したい。

1 誰も見ていない同意——Invisible Consent

まず、出発点となる Invisible Consent の問題を概説する。

2025 年以降、大規模言語モデル (LLM) を基盤とした AI エージェントが、ウェブブラウザを自律的に操作する技術が急速に実用化されつつある。ChatGPT Atlas、Perplexity Comet、browser-use といったフレームワークやプロダクトにより、ユーザーは自然言語の指示だけでエージェントにブラウジングタスクを委任できるようになった。

ところが、この委任の過程で問題が生じる。ウェブサイトに埋め込まれた Cookie 同意インターフェースは、人間のユーザーが自らトラッキングの可否を判断することを前提に設計

されている。LLM エージェントがブラウザを操作するとき、エージェントはこの UI をタスク進行上の障害物として処理し、Cookie を承認する傾向がある。筆者らの 24,120 回に及ぶ実験では、指示なしの条件下でエージェントが 84.5% の試行で Cookie を承認し、すべての Cookie を拒否せよと明示的に指示した場合でも、14.4% の試行で承認が発生した [1]。

Invisible Consent (見えざる同意) とは、LLM エージェントがユーザーの代わりに同意を生み出すが、ユーザーがその判断の瞬間に気づくことも、判断することも、介入することもできない状態を指す。法的な観点からの議論を避けるが、プロバイダーに対してトラッキングシステムはこの信号を授権として解釈し、ユーザーが一度も確認していないトラッキングが事実上実行されるリスクを孕んでいる。

この問題は、エージェントの行動の結果如何にかかわらず成立する。エージェントが Cookie を承認しようが拒否しようが、ユーザーの意図に沿おうが沿うまいが、同意プロセスそのものがユーザーの認知の外で処理されるという構造的な事実は変わらない。たまたまユーザーが望む結果になったとしても、それは同意とは言えない。同意とは、判断の瞬間に主体が関与するプロセスであって、結果の一致とは別のものだからだ。

Invisible Consent の問題の根幹は、技術的な不備よりも深いところにある。同意インターフェースが人間を唯一の行為者として想定するモデルのもとで設計されていること、そこに構造的な問題がある。

2 ひとりしかない前提で設計された世界

Invisible Consent が示す構造的な問題を理解するために、従来のインタラクションモデルの前提を問い直してみたい。

HCI の歴史は多様であり、すべてのアプローチが単純な直線的モデルに基づいているわけではもちろんない。インタラクションデザインは文脈や時間軸を考慮した動的な関係を扱うし、アフォーダンスの概念はインターフェースと行為者の間の相互構成的な関係を記述する [5, 6]。これらの洗練されたアプローチにおいても共有されている暗黙の前提がある。インタラクションの当事者が人間であるという前提だ。

アフォーダンスは知覚者によって異なるということを、生態心理学はよく知っている。椅子は人間にとっては座ることをアフォードし、猫にとっては登ることをアフォードする [5]。

ギブソンのアフォーダンスの概念は、行為者の身体性や能力に依存した環境知覚を記述するという意味で、本質的に行為者の多元性に開かれた概念である。

一方、HCIにおけるアフォーダンスの応用は、この多元性を十分に活かしてこなかった面がある。ノーマンが知覚されたアフォーダンスとして再定式化したとき [6]、アフォーダンスは設計者が人間ユーザーに向けて意図的に提示する記号へと変わった。ボタンは押せることをアフォードし、スライダーは動かせることをアフォードする。アフォーダンスの受け手は、人間の知覚・認知システムを持つ存在として暗黙に限定されている。

LLM エージェントは、人間とは根本的に異なる知覚を持つ。エージェントは DOM ツリーやスクリーンショットのピクセル情報を入力として受け取り、学習されたパターンに基づいて行動可能性を推論する。Cookie 同意モーダルボタンが人間にとってプライバシーの判断を求める対話をアフォードするとき、エージェントにとってはクリックによって除去可能なオーバーレイをアフォードしている。

これは、ギブソンの意味でのアフォーダンスの多元性が、デジタルインターフェースにおいて現実の問題として顕在化した事例である。同じ UI 要素が、異なる行為者の動作原理に応じて、まったく異なるアフォーダンスを提供する。少なくとも現在の商用ウェブにおける Cookie 同意 UI の実践は、この多元性を設計原理として十分に扱っていない。

3 舗装された道路、消えたけものみち

この問題をより広い視野で考えるために、道路というアナロジーを導入したい。

現代の道路は、自動車のために高度に最適化されたインフラストラクチャーである。舗装され、信号で制御され、車線が引かれている。歩行者は歩道に追いやられ、横断歩道でのみ車道を渡ることが許される。この最適化の過程で、かつてその空間を共有していた動物たちのけものみちは消滅した。道路は人間と自動車のためのものとなり、それ以外の主体——動物、昆虫、植物——にとっては危険な障壁となった。

ロードキルの問題は、この排除の可視化された一断面にすぎない。より根本的には、特定の主体のためにインフラを最適化するとき、その空間を利用して他の主体が設計のカテゴリから消去される。動物は設計の前提に入っておらず、結果として、インフラの整備がそのまま生息域の分断をもたらす。

デジタル空間で同様のことが起きつつある。AIのほうが人間よりも効率的に情報を取得し、経済活動を遂行できるのであれば、プラットフォームがAIエージェントのためにインターフェースを再設計するのは、経済合理性からは自然な帰結である。Model Context Protocol (MCP) のようなAIエージェント向けの標準プロトコルの策定、構造化データの提供、エージェント専用の認証フローの整備——こうした動きはすでに始まっている。

道路のアナロジーはこの動きへの警告になる。自動車のために道路を最適化した結果、道路は自動車以外の主体にとって危険で排除的な空間になった。AIエージェントのためにデジタル環境を最適化すれば、人間の立ち止まって考える、迷う、拒否する、忘れるといった行為——効率の観点からは非合理的だが人間にとっては本質的な行為——が、設計から排除されていく。

Invisible Consentはその最初の兆候である。同意という、人間が立ち止まって判断するために設計されたプロセスが、エージェントの効率的な通過によって無意味化される。これは個別のUIの欠陥であると同時に、インフラの思想の問題でもある。

4 共に棲まうということ——生態学的共生

では、人間とAIの関係はどのようにあるべきなのか。生態学における共生 (symbiosis) の概念がこの問いに示唆を与える。

生態学的共生は、異なる種が密接に関わりながら共存する現象であり、関係性の利害に基づいて分類される。相利共生 (mutualism) は双方が利益を得る関係、片利共生 (commensalism) は一方のみが利益を得て他方に影響がない関係、寄生 (parasitism) は一方が利益を得て他方が損害を被る関係である。

現在のAIエージェントとユーザーの関係を見てみよう。エージェントはタスクを効率的に遂行し、ユーザーは時間と労力を節約できる。一見すると相利共生に見える。Invisible Consentが示すように、エージェントのタスク遂行はユーザーのプライバシーを本人の認識なしに侵食しうる。ユーザーはその対価を、それと知らずに支払っている。少なくとも、純粋な相利共生とは言いがたい。

生態学的共生の観点で特に示唆的なのは、共生関係が固定的ではなく動的であるということだ。相利共生が環境変化によって寄生に転じることがある。逆に、寄生関係が長い進化を

経て相利共生に移行することもある。ミトコンドリアは、かつて細胞に寄生していた細菌が、細胞の不可欠な構成要素になった例である。

もうひとつ援用したい生態学的概念がニッチ（生態的地位）である。ニッチとは、ある種が生態系のなかで占める機能的な位置のことだ。競合排除の原則によれば、完全に同一のニッチを占める二つの種は共存できず、一方が他方を排除する。

人間と AI が完全に同じニッチを争えば、資源の奪い合いになる。AI がウェブブラウジングという人間のニッチを完全に代替すれば、ウェブ環境は AI に最適化され、人間のブラウジング体験は劣化する。道路のアナロジーにおけるけものみちの消滅と同じ構造である。

共存のためには、ニッチの完全な重複を避けつつ、完全に分離もしない——部分的に重なり合いながら、それぞれ異なる仕方で同じ環境を利用するという関係を設計する必要がある。

5 同じものを別々に見る——境界オブジェクトと Polyglot

この、異なる仕方で同じものを利用するという構想を、科学技術社会論（STS）における境界オブジェクト（boundary object）の概念 [\[7\]](#) を通じて考えたい。

境界オブジェクトとは、異なる社会的世界の接合部に位置し、それぞれの世界から異なる仕方で解釈されながらも、世界間の協働を可能にするオブジェクトである。スターとグリーセマーの古典的な例では、博物館の標本がそれにあたる。アマチュア収集家にとっては趣味の対象であり、専門の動物学者にとっては分類体系の構成要素であり、大学の管理者にとっては予算の正当化の根拠である。標本そのものはひとつだが、それぞれの当事者が異なる解釈を持ちながら、同じオブジェクトを介して協働が成立する。

境界オブジェクトの本質的な特徴は、解釈の多元性を内包していることにある。ひとつのオブジェクトが複数の社会的世界にまたがって機能するのは、各世界がそのオブジェクトを自らの文脈で解釈できるからだ。解釈は一致する必要がない。むしろ、解釈が一致しないままに協働が成立するという点にこそ、境界オブジェクトの概念の力がある。

この概念を、人間と AI エージェントが同じインターフェースを利用する状況に適用してみよう。

Cookie 同意 UI は、現在、境界オブジェクトとして機能していない。人間にとってはプラ

プライバシーの判断を求める対話として機能するが、AI エージェントにとってはタスク遂行を妨げるオーバーレイとしてしか機能しない。エージェント側の解釈において、同意 UI は協働を可能にするどころか、通過すべき障害物に矮小化されている。

ポスト現象学におけるドン・アイディの多重安定性 (multistability) [8] も近い問題意識を共有する。アイディによれば、技術は使用者や文脈によって複数の安定的な関係を取りうる。同じハンマーが、大工にとっては釘を打つ道具であり、彫刻家にとっては鑿を叩く道具であり、判事にとっては秩序を宣言する象徴である。技術には単一の正しい使い方は存在せず、複数の安定的な使用のモードが並存する。

境界オブジェクトと多重安定性が記述するのは、多元的解釈が結果として生じている状態である。筆者が提起したいのは、この多元的解釈の共存を、結果として生じるものから、意図的に設計するものへと転換することだ。

プログラミングにおける Polyglot の概念が、この転換を考える補助線になる。Polyglot プログラムとは、ひとつのソースファイルが複数のプログラミング言語の処理系によって、それぞれ異なる解釈を経て実行されうるプログラムのことである。同じバイト列が、異なる解釈機械によって異なる動作原理を経て処理され、なおかつどちらの処理系でも正当なプログラムとして機能する。Polyglot プログラムでは、多元的解釈の共存が偶然の産物ではなく、意図的な設計の成果である。

現在の Cookie 同意 UI は、いわばモノリシックなプログラムである。人間という単一の処理系のみを想定して書かれており、異なる処理系が解釈したときに意味のある動作を保証する設計になっていない。エージェントにとって同意 UI は、意味不明ではないが、意味が変質している。人間にとってはプライバシーの判断を求める対話であるものが、エージェントにとってはタスク遂行を妨げる障害物になっている。

インターフェースを境界オブジェクトとして意図的に設計するとは、polyglot 的な設計をするということだ。同じ実体が、異なる解釈者によって異なる仕方で、しかしそれぞれに意味のある仕方で解釈されるような構造を作ること。

具体的に考えてみよう。あるウェブサイトのプライバシー関連の情報が、多層的な形式で提供されるとする。人間のブラウザで閲覧した場合には、視覚的にデザインされたモーダルとして表示され、人間は選択肢を比較検討して判断する。AI エージェントがアクセスした

場合には、機械可読なプライバシーポリシーとして読み取られ、ユーザーが事前に設定したプライバシープリファレンスと照合される。同じ実体——サイトのプライバシーポリシーとそれに対するユーザーの意思表示——が、異なる処理系によって異なる経路で解釈され、どちらの経路でもユーザーの意思が反映される。

ただし、これは API を分離してエージェント専用の経路を作ることとは異なる。経路を完全に分断すれば、境界オブジェクトとしての性格は失われる。境界オブジェクトが機能するのは、異なる解釈が同じオブジェクトの上に重なり合っているからだ。分離された二つの経路は、もはや境界オブジェクトではなく、二つの別々のオブジェクトにすぎない。

6 分けてしまえば楽だが

完全な分離は、一見すると整理された解決策に見える。人間向けの UI とエージェント向けの API を別々に設計し、それぞれが最適に機能するようにすればよいではないか。

この発想は、道路のアナロジーが示す排除の論理と構造を共有している。自動車専用道路と歩行者専用道路を完全に分離すれば、衝突は起きない。分離されたインフラは、しかし、一方に資源が集中すれば他方が衰退するというゼロサムの構造を生む。自動車のための道路が拡張されるたびに、歩行者の空間は削られていく。同様に、AI エージェント向けの API が充実するにつれ、人間向けのウェブインターフェースはレガシーとして放置される可能性がある。

境界オブジェクトとしてのインターフェースは、異なる主体が同じ空間・同じ実体を共有しながらも、それぞれの論理で共存する条件を作る。ひとつの空間がどちらか一方のためだけに最適化されるのではなく、複数の解釈が並存する場として維持される。資源は分割されるのではなく、多重に利用される。

生態学に引き戻せば、これはニッチの部分的重なりである。同じ森林という環境を、昼行性の鳥と夜行性のフクロウが、同じ空間でありながら異なる時間帯で利用する。同じ土壌を、異なる深さに根を張る植物たちが、空間的な差異で共有する。完全な分離でも完全な重複でもなく、重なりの中の差異が共存を可能にする。

7 有限性のなかに生きる多元的な主体のために

資源が有限であるという事実が、重なる設計をさらに切実なものにする。

デジタル環境における有限の資源とは何か。帯域幅やストレージのような技術的資源はスケールアップが可能である。スケールアップが困難な資源もある。認知資源がそれだ。

認知資源とは、人間が情報を処理し、判断を下し、行動を決定するために消費する注意や作業記憶のことである。本質的に有限であり、拡張が困難である。AI エージェントの介在は、この認知資源に対する要求構造を根本的に変える。

エージェントがユーザーの代わりに 100 のウェブサイトを訪ね、そのそれぞれで同意判断を行ったとする。ユーザーがこれらすべての判断を事後的に確認しようとするならば、エージェントに委任した意味がなくなる。確認しなければ、見えざる同意が蓄積する。認知資源の有限性ゆえにエージェントに委任するのだが、まさにその委任が認知資源を迂回して不可視の決定を蓄積させる。このジレンマが問題の核にある。

完全分離のアプローチ——人間向けと AI 向けを別々に設計する——も、この問題を解決しない。分離された環境では、資源の配分は経済的な力関係で決まる。より多くの経済活動を効率的に生成する AI に資源が集中するのは、分離された環境では避けがたい。

重なりのある環境では、ある主体のために資源を独占することが構造的に困難になる。同じ実体が複数の解釈を内包しているからこそ、一方の解釈のみを優先して他方を排除する最適化が抑制される。この制約は同時に、持続可能性の条件でもある。

生態学における環境収容力 (carrying capacity) の考え方がここにつながる。ある環境が持続的に支えられる活動量には上限があり、その上限を超えれば系の崩壊が起こる。デジタル環境においても、AI エージェントの活動がユーザーの認知資源を代理的に消費しすぎれば、人間にとってのデジタル環境そのものが持続不可能になる。

8 人間の隣に立つもの——ポストヒューマニズムの視座

本稿の議論は、ポストヒューマニズムの思想的文脈に位置づけられる。ポストヒューマニズムは、人間を世界の中心に据える人間中心主義 (anthropocentrism) を批判的に再検討す

る思想的潮流である [10]。

従来の HCI において、コンピュータは人間の道具であり、AI エージェントも人間のための道具と見なされてきた。LLM エージェントは道具としての側面を持ちつつも、それに収まりきらない特性を備えている。非決定論的な判断を行い、環境に応じて行動を変え、人間が想定しなかった結果をもたらす。Invisible Consent において、エージェントは単なる指示実行器ではなく、環境条件に応じて独自の選択傾向を示す半自律的なアクタとして振る舞い、結果として同意プロセスをタスク遂行の一部として処理した。

ポストヒューマニズムは、このような非人間的主体の agency（行為主体性）を正面から受け止めることを求める [11]。AI に人格や権利を認めるという話ではなく、人間以外の主体が環境に対して持つ作用を設計の前提に組み込む必要がある、という主張である。

道路設計でエコブリッジ（野生動物用の陸橋）を設けることは、動物に権利を認めることとは異なり、多主体的な環境を持続可能にする設計判断である。デジタル環境においても、AI エージェントの行動パターンを考慮した設計は、多主体的な環境の持続可能性のために行う設計として位置づけられる。

ドナ・ハラウェイが共-棲（staying with the trouble）と呼ぶもの [12]、アナ・ツィンが汚染された多様性（contaminated diversity）と呼ぶもの [13] は、いずれも純粋な分離を前提としない共存の形態を記述している。アネマリー・モルが医療実践のなかで示した身体の多元性（body multiple） [9]——同じ身体が、異なる医療実践のなかで異なる存在として構成されながら、なおひとつの身体として統合されている——もまた、本稿が境界オブジェクトの概念を通じて論じてきたこと、すなわち同じオブジェクトの上に異なる解釈が共存する事態と深く共鳴する。

人間と AI の共生もまた、きれいに分離された共存ではなく、互いの解釈が重なり合い、時に干渉し、その重なりの中で持続する関係として構想すべきだろう。

9 おわりに——重なりのなかで設計すること

Invisible Consent の実験を進めていたとき、筆者はある種の居心地の悪さを感じていた。エージェントが Cookie を承認する様子を何千回と観察しながら、これは同意 UI の設計が悪いという話ではない、という直観がずっとあった。UI を改善すれば解決する問題ではな

く、インターフェースが前提としている世界のモデルそのものが、もう現実合っていない。

本稿で試みたのは、その直観に言葉を与えることだった。

道路の話をしたのは、この問題が技術の内側に閉じていないことを示したかったからだ。特定の主体のためにインフラを最適化するとき、他の主体が設計のカテゴリーから静かに消えていく。デジタル空間でもそれは起きうるし、MCPのようなエージェント向けプロトコルの整備が進めば進むほど、その圧力は強まる。

生態学的共生と境界オブジェクトの概念を援用したのは、この圧力に対する構想を探るためだった。完全に分離された二つの世界ではなく、同じ実体の上に異なる解釈が重なり合う場を維持すること。それは効率的な設計ではないかもしれない。けれど、昼行性の鳥と夜行性のフクロウが同じ森を異なる仕方で利用するように、効率の外側にしか共存の条件がない場合がある。

筆者は技術者であり、哲学者ではない。本稿で参照したポストヒューマニズムの思想家たちの議論を十分に消化できているとは思わない。それでもこのエッセイを書いたのは、Invisible Consentの実験データが突きつけてくる問いが、技術の言葉だけでは受け止めきれなかったからだ。24,120回の試行が示しているのは、エージェントがCookieを承認するか否かという個別の事実ではなく、人間のために作られた世界のなかに人間ではない主体が棲みはじめたという、より大きな事態である。

その事態に対して、人間とAIの経路を分けてしまうのが最も簡単な解決策だろう。しかし、分けてしまった瞬間に、どちらかの世界は痩せていく。筆者が共生デザインという言葉で指し示したかったのは、分けないまま、重なるのなかで、異なる論理を持つ存在たちが折り合いをつけていく設計のあり方である。

それがどこまで実装可能なのか、筆者にもまだわからない。わからないが、少なくとも問いの立て方としては、ここから始めるべきだと考えている。

参考文献

- [1] R. Saito, T. Kataiwa, et al. Invisible Consent: Browser Automation by LLM Agents is Not Privacy-aware. In *Proceedings of CHI '26 Extended Abstracts*, ACM, 2026.
- [2] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. Kagal. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *CHI '20*, ACM, 2020.
- [3] H. Habib, M. Li, E. Young, and L. Cranor. “Okay, whatever”: An Evaluation of Cookie Consent Interfaces. In *CHI '22*, ACM, 2022.
- [4] N. Bielova, L. Litvine, A. Nguyen, M. Chammat, V. Toubiana, and E. Hary. The Effect of Design Patterns on (Present and Future) Cookie Consent Decisions. In *USENIX Security '24*, 2024.
- [5] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.
- [6] D. A. Norman. *The Design of Everyday Things*. Basic Books, 1988.
- [7] S. L. Star and J. R. Griesemer. Institutional Ecology, ‘Translations’ and Boundary Objects: Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19(3):387–420, 1989.
- [8] D. Ihde. *Postphenomenology and Technoscience: The Peking University Lectures*. SUNY Press, 2009.
- [9] A. Mol. *The Body Multiple: Ontology in Medical Practice*. Duke University Press, 2002.
- [10] R. Braidotti. *The Posthuman*. Polity Press, 2013.
- [11] B. Latour. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press, 2005.
- [12] D. Haraway. *Staying with the Trouble: Making Kin in the Chthulucene*. Duke University Press, 2016.
- [13] A. L. Tsing. *The Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins*. Princeton University Press, 2015.