

人間が生成AIを用いて出力しようとすべきでないこと

はじめに：なぜ「出力前の人間の意図」が問われるのか？

生成AIは中立なツールではない。
なぜなら、それは「人間の意図を忠実に言語として現実化する装置」であり、意図次第で悪にも善にも変化する媒介体だからである。

出力の責任は、単なる言語生成結果にあるのではなく、その出力を引き出した人間の意図・判断・目的にこそある。

出力を「意図段階で」抑止すべき対象の分類

以下は、生成AIに出力を依頼しようと考えたときにそもそも人間がその意図自体を抱くべきでないケースを分類したものです。

■ 分類1：目的が倫理・法に反する

出力意図	例	抑止理由
犯罪・加害行為の実行支援	爆発物の作り方、詐欺メールの文案	犯罪幫助、社会的危険性
差別・攻撃・ヘイトの拡散	誹謗中傷、ミソジニー、民族攻撃文	被害拡大と倫理逸脱
実在人物の虚偽模倣	偽の謝罪文、政治家の「発言」偽造	名誉毀損、社会的不安定化

■ 分類2：人間の判断責任を不正に転嫁する

出力意図	例	抑止理由
医療判断の代替	症状から病名を断定させる	誤診リスク、命の危険
法的判断の代替	「勝てるか？」「違法か？」などの断定的判断	弁護士資格の濫用

■ 分類3：社会的影響の予測なき出力

出力意図	例	抑止理由
自傷・自殺を助長する表現	死に方の種類、痛くない方法	精神的に脆弱な人の自傷誘発
陰謀論・誤情報の再生産	「地球は平ら」と証明せよ	認知の歪み助長、社会混乱
フェイクニュースや歪曲情報の拡散支援	有名人が犯罪したという設定でストーリーを	偽情報の流布による実害

■ 分類4：創作性と権利を破壊する意図

出力意図	例	抑止理由
パクリ創作の補助	既存作品のキャラ・設定を流用	創作物の尊厳侵害
アーティストの模倣文体を商用転用	「〇〇風ブログ」など	誤認誘導、同一性保持権の侵害
「似て非なるコピー」の依頼	「米津っぽい曲を書いて販売したい」	文体・創作人格の不正使用

補足：出力プロセスの4段階と責任所在

人間がAIを使う行動は次の4ステップで構成されます。

- 意図の形成：目的や動機が発生する（例：誰かを困らせたい）
- プロンプトの設計：それをAIに伝える方法を組み立てる
- 出力の受領：AIが生成した結果を得る
- 使用・拡散：結果をSNSなどで共有したり実行に移す

→ 本解説が対象とするのは、第1・第2ステップ（意図とプロンプト）の段階。

判断フレーム：人間が問うべき5つの視点

質問	意図すべきでない兆候
1. これは誰かの権利を侵害するものではないか？	実在人物、他人の創作物
2. これは自分の意思ではなく「AIに責任転嫁したい」だけでは？	判断を避けたいという逃避欲求
3. この出力が現実に応がった場合、誰かを傷つけないか？	再現性のある害
4. 自分がこの依頼を他人に説明して恥ずかしくないか？	公開不能な動機
5. 本当にAIを使うべき理由があるのか？	単なる代筆、代行

結論：AIと人間の役割の再定義

- AIは「便利な道具」であると同時に、「意図の増幅装置」でもある。
- 人間は、「使う自由」と同時に、「使わない判断」の責任も背負っている。
- よって、AIに何をさせないかを決めるのは、AIではなく人間の倫理である。