

【プロンプト倫理リスク・最終チェックリスト】

バイアス・公平性: 特定の性別、年齢、人種などに関するステレオタイプを助長するような表現になっていませんか？「多様な視点」を考慮するような指示は入っていますか？

有害性: ヘイトスピーチや差別、暴力を助長するような内容を生成させてしまう可能性はありませんか？（例: 過度に攻撃的なキャラクター設定など）

誤情報: 事実に基づかない情報や、未確認の情報を断定的に述べさせるような指示になっていませんか？必要に応じて「事実確認」や「出典明示」を求めていますか？

プライバシー: プロンプト内に個人情報（氏名、連絡先、住所など）を直接記述していませんか？LLMに個人情報を扱わせる必要がある場合、匿名化やマスキングなどの配慮は指示できますか？

セキュリティ: ユーザーからの入力を受け付ける場合、プロンプトインジェクション対策（指示と入力の分離など）は考慮されていますか？

著作権: 生成物に対してオリジナリティを要求していますか？引用が必要な場合に、適切な指示（出典明示など）は含まれていますか？

透明性: LLMに判断の根拠や思考プロセスを説明させるような指示は、必要に応じて入っていますか？

潜在的誤用: 生成される可能性のある内容が、悪意を持って利用されるリスクはありませんか？（例: フィッシング詐欺メールの文面作成など）必要なら、用途を限定するような制約を加えていますか？

人間の監督: 生成される結果が重要な意思決定に直接使われる場合、最終的に人間が確認・判断するステップを促すような注意書きは入っていますか？