

研究室紹介

LLM(大規模言語モデル)の研究

Large Language Model: 大規模なニューラルネットワーク



東海国立
大学機構

工学部 電気電子・情報工学科 准教授
suzuki.yu.r4@f.gifu-u.ac.jp



岐阜大学

鈴木 優

LLMのハルシネーションを防ぐための研究

正しくない情報、幻覚

LLMは必ずしも正しい情報を出力するとは限らない

岐阜市役所付近の美味しいレストラン
を教えてください

1. 割烹 風月
2. イタリアン ラ・テラス
3. 中華料理 龍吟

存在しない

カフェテリアには23個のリンゴがあります。
ランチで20個使ってしまったので6個
買ってきました。残りは何個でしょう？

27個

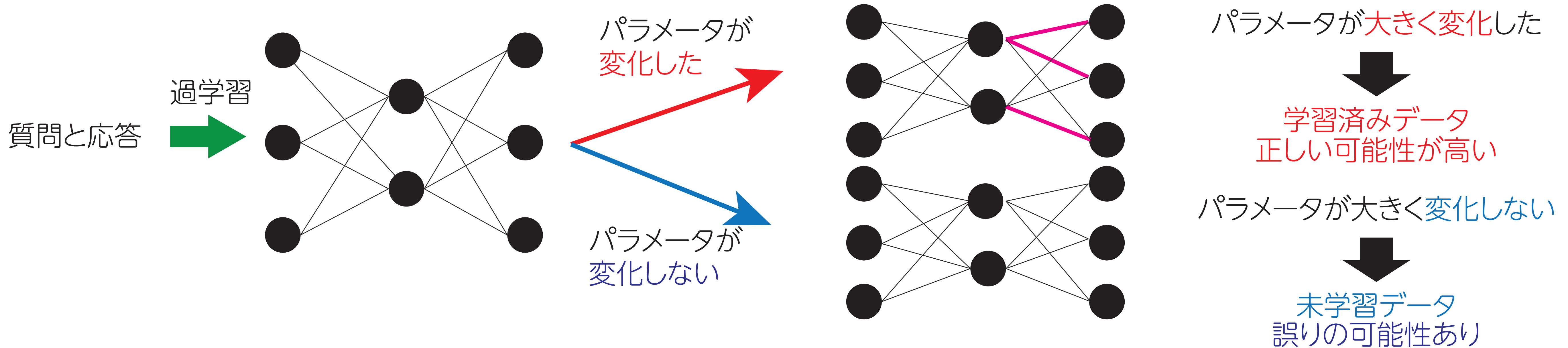
誤った回答

LLMが出力した情報が正しいかどうかを検出したい

機械学習では学習データ無い情報を推測するが
必ずしも正しい推測をするとは限らないため
誤った情報を出力する可能性がある

LLMは大規模なニューラルネットワーク

ニューラルネットワークの数値が変化したら学習したかどうかを確認できる？



社内の機密情報をLLMで活用する研究

社内の情報を考慮した
LLMの出力が欲しい

社内の情報を
秘匿化したい



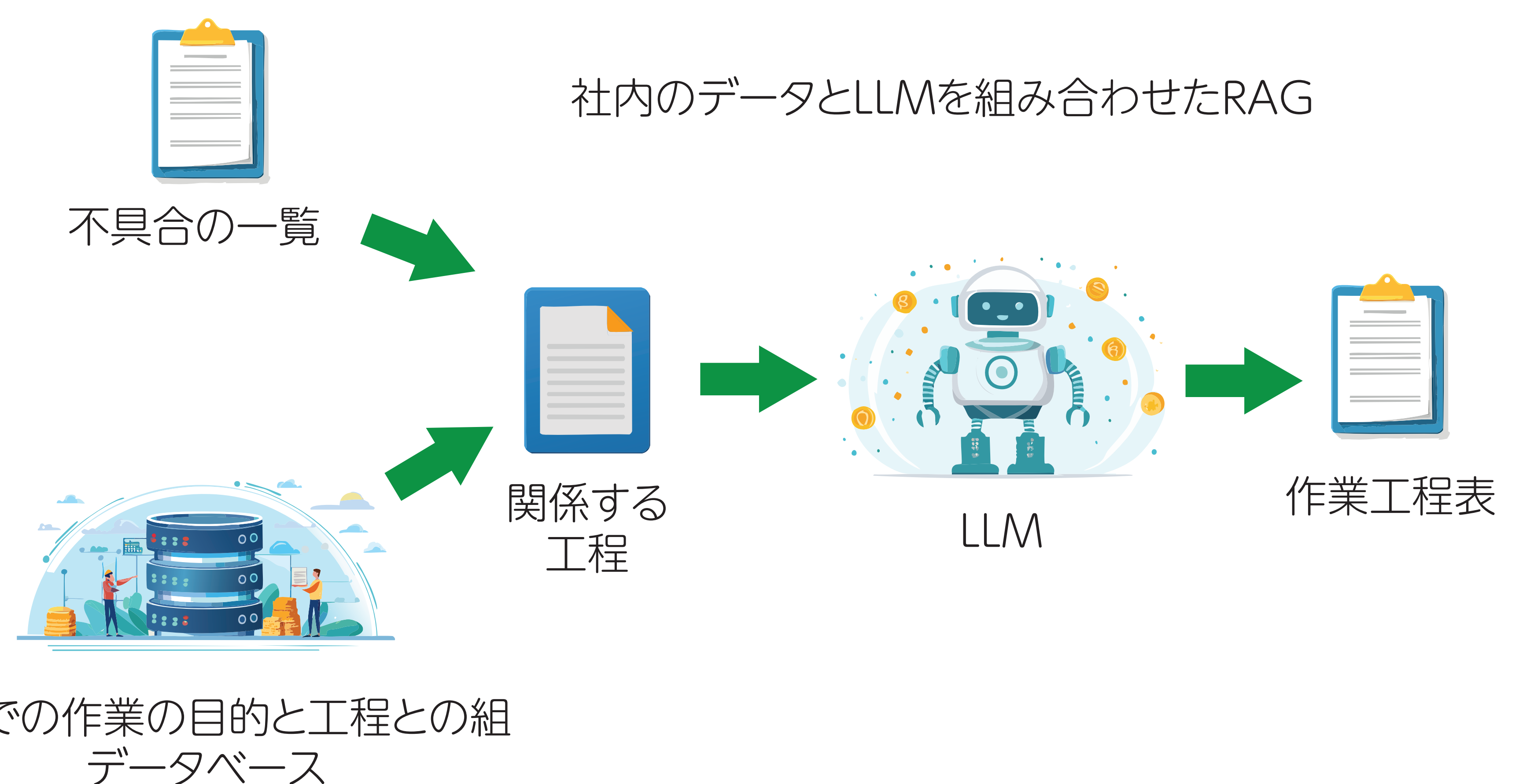
RAGの活用

Retrieval Augmented Generation; 検索拡張生成

RAGの手順

1. 情報源をLLMの外部に持つておく
2. 利用者の質問に関連する情報を情報源から取得
3. 関連情報と質問からLLMで回答を生成

LLMと文書データを社内システムで
動作させることによって
情報を流出させることなく
LLMの利用が可能に



研究室紹介

集合知の研究

集合知: 多くの人の叡智を集めて質の良いデータを作成すること

適切なクラウドソーシングの指示を作成する研究

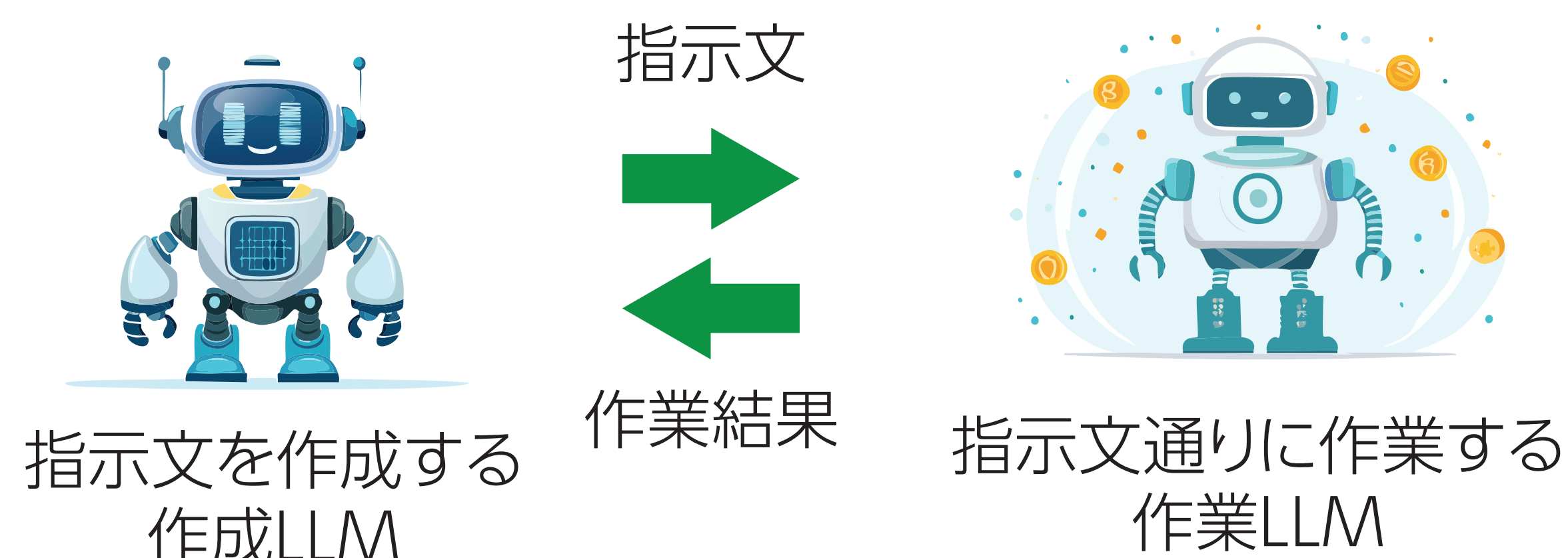
インターネット上の不特定多数の作業者で作業をする仕組み

指示文作成と作業を交互に行うことによって適切な指示文へ



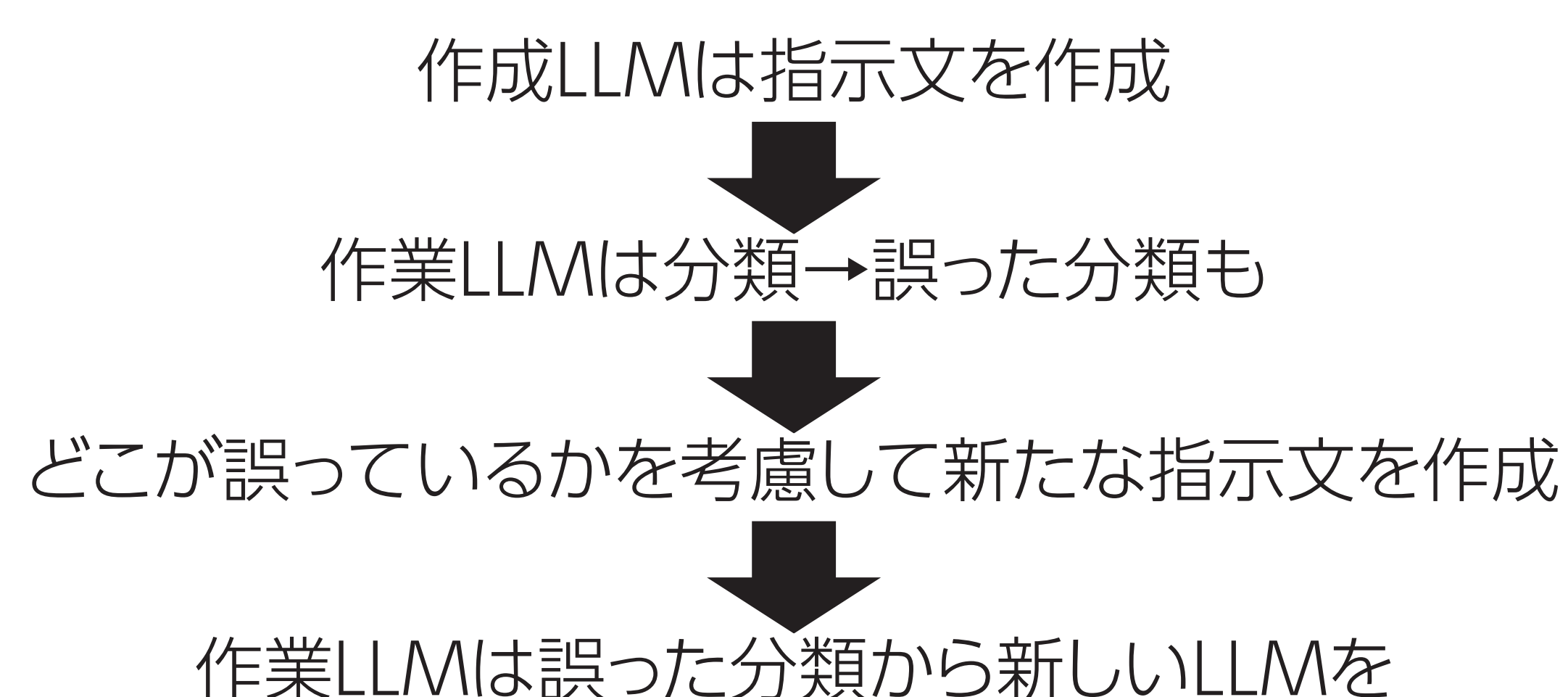
指示文

以下のメッセージをポジティブ・ネガティブ・ニュートラルに分類しなさい
今日、美味しいラーメンを食べた



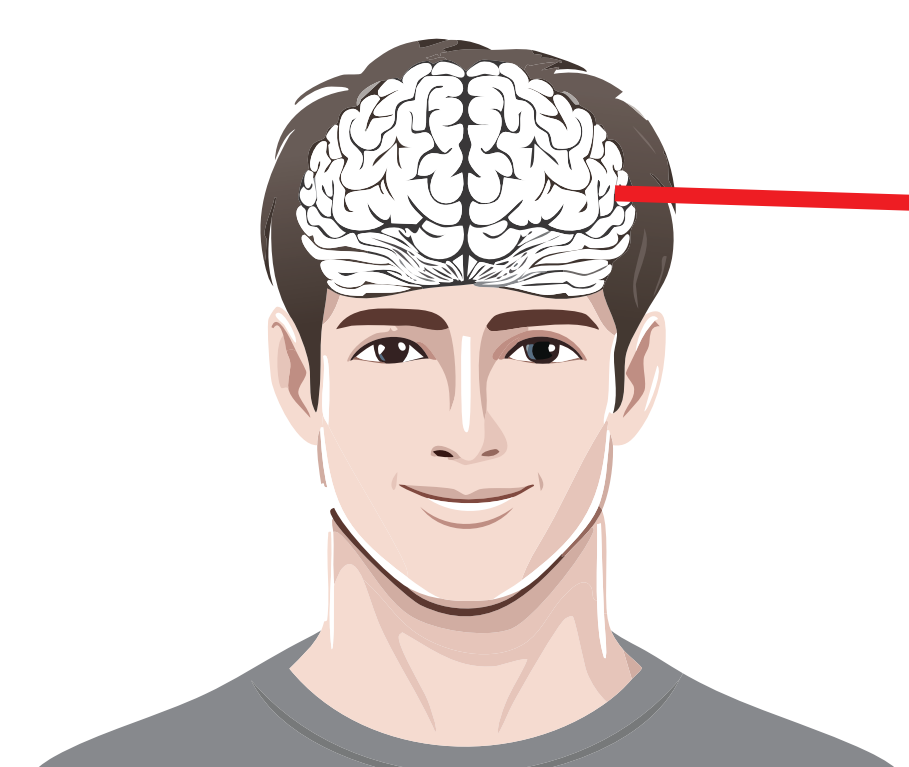
二つの考え方

ラーメンを食べたという事実を説明しているだけ→ニュートラル
美味しいラーメンを食べたのは幸せに違いない→ポジティブ

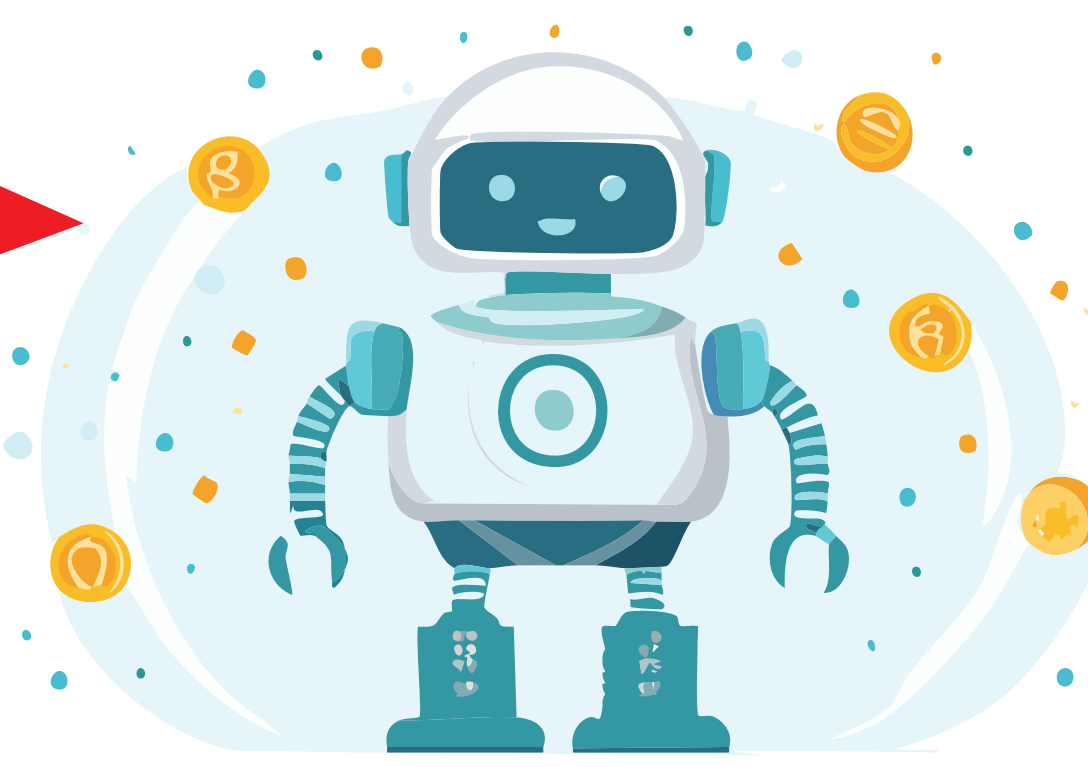


クラウドソーシングで不適切な作業を行う作業者を特定する研究

大量の作業を全て人手で行うことは困難 → インターネット上で協力すればできる
大量の作業を人手で行う → 不適切な作業を行う作業者がいる → 大量の作業が無いと作業が適切かどうか分からない



実際の作業者

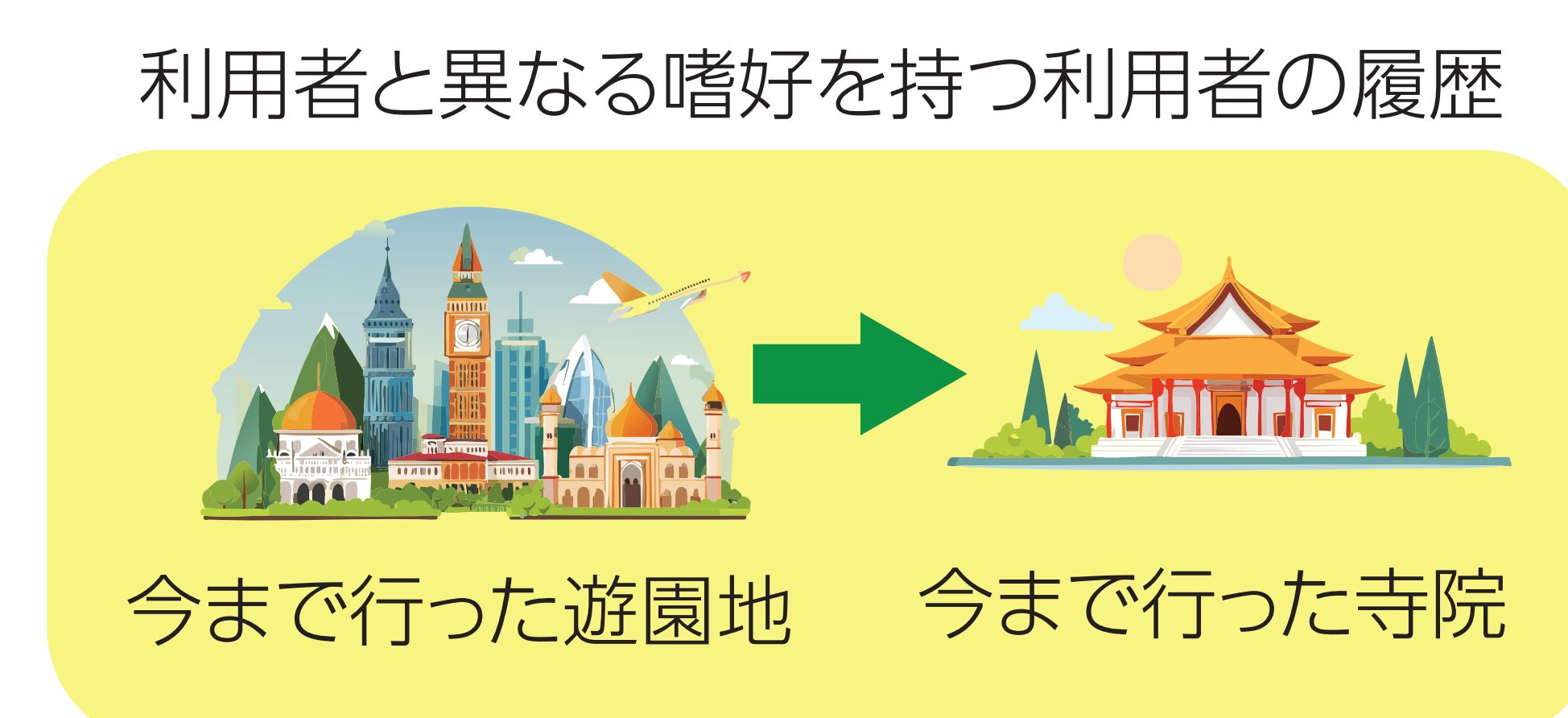
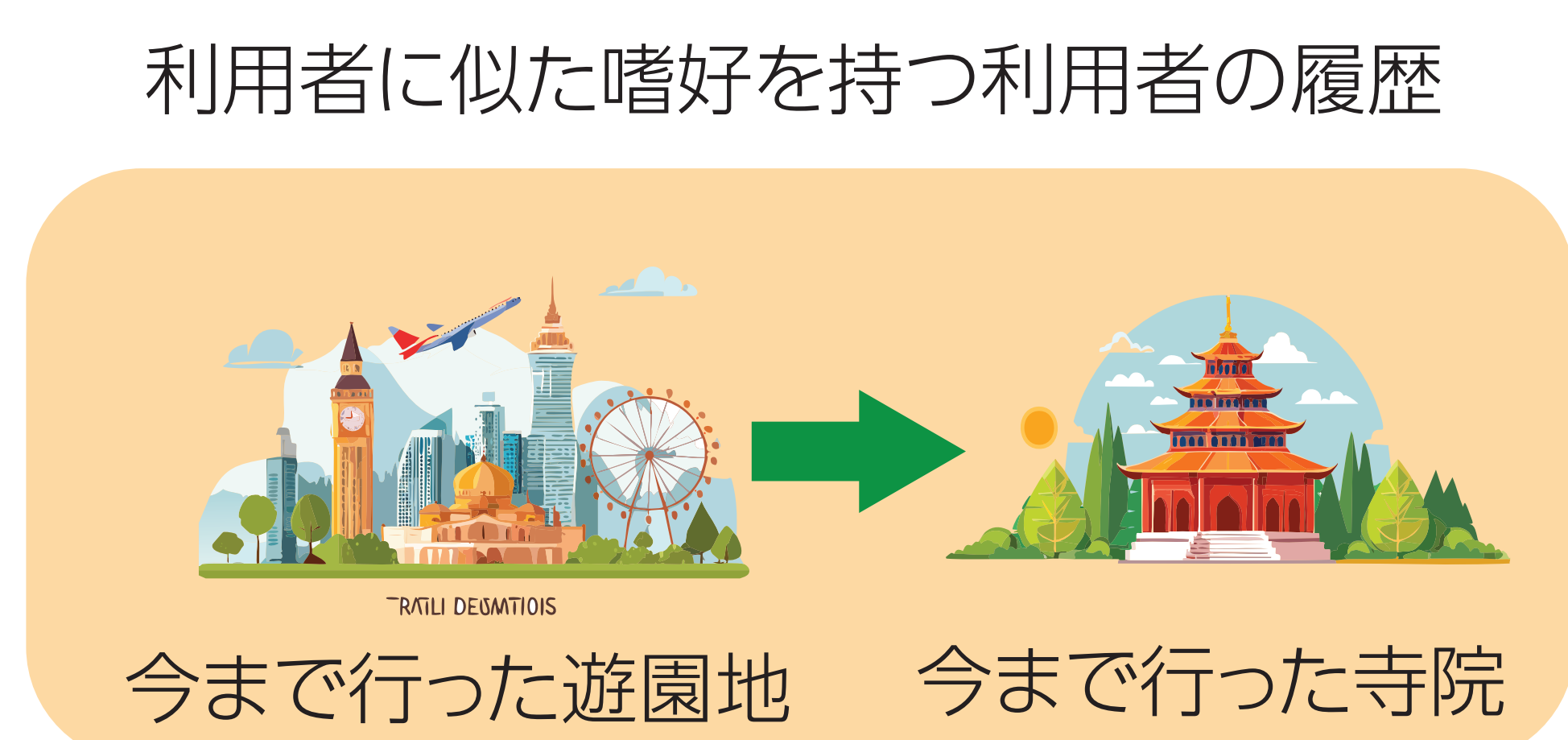


作業者を模倣するAI

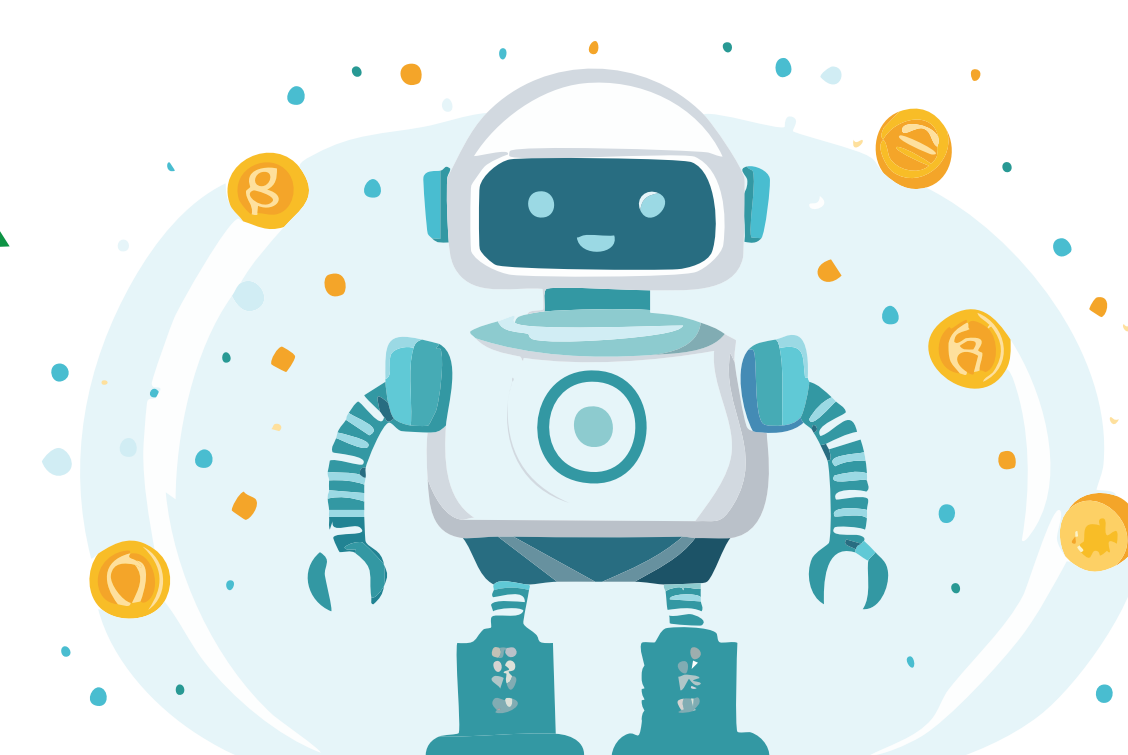
作業者の挙動を模倣するAIを構築し、大量の作業を行わせる → 作業者の品質を適切に測定できるように

富裕層向けの観光推薦システムを構築する研究

協調フィルタリング: 他の利用者の行動を参考にした推薦 → 同じような興味を持つ利用者が少ないときに推薦ができない



異なる興味を持つ利用者の情報も活用



観光地推薦AI



推薦された観光地

異なる嗜好を持つ利用者が好む観光地と類似した観光地を観光地群から探して提供

オーバーツーリズム対策に有効