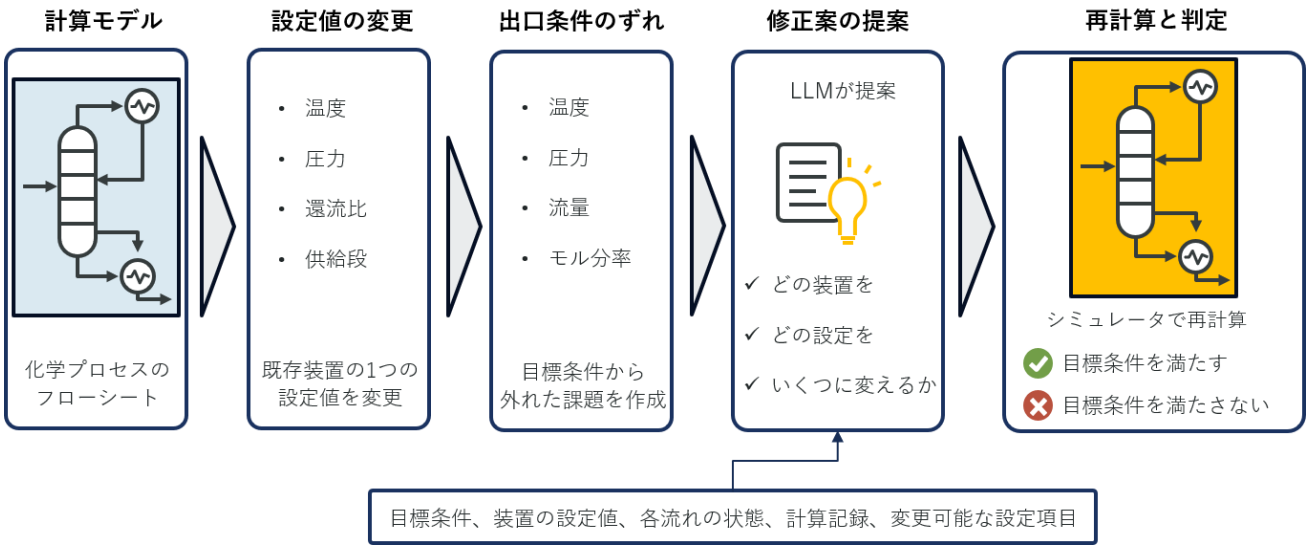
化学プロセス設計支援 AI の評価方法を開発

―LLM の化学プロセス修正提案能力を測る―

概要

化学製品の製造工程の設計には化学プロセスシミュレータ※1 や専門知識に基づく検討が必要です。近年、この作業を生成 AI の一種である大規模言語モデル（LLM）※2 で支援することが期待されています。その一方で、AI が提案した条件が、目的の結果を実際にもたらすかどうかを確かめることは重要な課題です。

京都大学大学院工学研究科 高原渉 特定助教、同大学工学部 菅沼龍 学部学生、同研究科 宮本奏汰 助教、田辺克明 教授（兼：京都 MPI 株式会社 取締役 CTO）らの研究グループは、AI による化学プロセスの設定値の修正案をシミュレータで検証する評価ベンチマーク※3「PSE-RepairBench」を開発しました。共通の課題で、LLM による化学プロセス設計支援において、その挙動や修正提案の有効性を比較できることを示しました。また、化学プロセスの条件修正を提案する能力を評価でき、コーディングや推論能力などを測る公開ベンチマークとは違う観点から LLM を評価できることが示唆されました。今後の LLM の選定・改良に役立ち、化学プロセス設計支援 AI の実現に向けた共通の基盤となることが期待されます。本研究成果は、2026 年 11 月 19 日（現地時刻）に英国の国際学術誌「*Digital Discovery*」にオンライン掲載されます。



LLM の修正提案の有効性を計算で確かめる評価の枠組みである「PSE-RepairBench」の概要図

（作成：高原 渉）

１．背景

医薬品や樹脂、燃料などをつくる化学プロセスでは、反応器、分離器、熱交換器、蒸留塔などを組み合わせ、目的とする製品が得られるように温度や圧力などを調整します。一つの装置の設定変更が、後段の装置や、プロセスの出口を流れる物質の温度・圧力などにも影響するため、装置同士の関係を踏まえた検討が重要です。化学プロセスシミュレータは、こうした影響を計算によって確かめ、設計を支える役割を担います。

近年、科学技術分野における LLM の活用が進んでいますが、その能力を化学プロセスの設計支援につなげるには、使用する LLM が化学プロセス特有の性質を捉える能力をどの程度有するかが重要となってきます。そこで本研究では、その土台として LLM における化学プロセスの条件修正を提案する能力を測るベンチマーク「PSE-RepairBench」を開発しました。

２．研究手法・成果

本研究では、化学プロセス設計における条件調整のうち、既存装置の一つの設定値を修正する課題に着目しました。オープンソースの化学プロセスシミュレータの公開データベースから計算モデルとなるフローシート[※4]を集め、装置の設定値を一つだけ変えて、出口の温度、圧力、流量、または指定した成分のモル分率[※5]が元の計算モデルの基準値から外れる課題を作成しました。本ベンチマークは 100 件のテストに課題より構成されます。

本ベンチマークのタスクでは、LLM に目標条件、装置の設定値、各流れの状態、計算記録、変更可能な設定項目を与え、「どの装置の、どの設定を、いくつに変えるか」を一回だけ提案させます。その提案を計算モデルに適用し、化学プロセスシミュレータで再計算して有効性を判定します。

各課題では、評価する出口の流れと、温度や圧力などの評価項目を一つずつ事前に指定しました。設定変更によって生じたずれを基準値の方向へ 7 割回復する値をしきい値とし、上昇した値はしきい値以下に、低下した値はしきい値以上にすることを目標としました。LLM の提案を適用した再計算が正常に完了し、指定した評価項目がこの目標を満たした場合を「修復成功」としました。元の設定値への完全な復元は求めず、条件を満たす別の設定値も成功として評価しました。例えば、空気を圧縮して冷やし、水分を分離する工程では、上流の圧縮機の設定が、出口の水の圧力にも影響します。本研究では、正常時に 275.0 kPa であった出口の水の圧力が、上流の圧縮機の出口圧力の設定値を意図的に変更することで 412.5 kPa に上昇した状況を課題として設定しました。この課題では LLM に、「どの装置の、どの設定を、いくつに変えるか」を提案させ、出口の水の圧力が 316.25 kPa 以下になれば「修復成功」となります（参考図表を参照）。

OpenAI の GPT 系列、Anthropic の Claude 系列、Google の Gemini 系列の計 9 種類の LLM を本ベンチマークにて評価した結果、一回の提案による修復成功率は 34～67%となりました。無作為に変更対象と数値を選ぶ比較手法の平均 8.6%（5 回の評価）を全ての LLM が上回り、LLM が化学プロセス情報を基に、所定の目標条件を満たす修正案を生成できることを確認しました。一方、最も成功率の高い LLM でも、約 3 分の 1 の課題では目標条件を達成できず、今回評価した LLM のうち検証当時のフロンティアモデルにも改善の余地があることが示されました。

9 種類の LLM が生成した成功例 437 件のうち、約 13%は元の計算モデルの基準値と完全に一致し、約 74%は同じ装置・設定項目を選びながら、その基準値とは異なる値で目標を達成していました。この結果は、あらかじめ用意した修正例との一致だけでは捉えられない、有効な提案も評価できることを示しています。

また、本ベンチマークでの 3 種類の LLM 間の相対的な成績は、コーディングや推論などを測る公開ベンチマークの評価による結果とは異なりました。この結果は、化学プロセス設計を支援する LLM の選定には、この分野に即した評価が重要であることを示しています。

さらに、本ベンチマークにおける、各 LLM の失敗の内容を分析したところ、性能の高いモデルほど修正すべき装置を正しく見つけられる一方、適切な数値の設定や、修正後の計算を安定して完了させることには、検証当時におけるフロンティアモデルにおいても課題が残ることが分かりました。

PSE-RepairBench は、修正案の有効性を共通の基準で評価し、LLM に化学プロセスの条件修正を提案させる能力を測るための具体的な手がかりを与える評価基盤です。

3．波及効果、今後の予定

研究で開発した評価方法は、化学プロセス設計の支援に適した AI を選び、より有効な提案を行う AI へと改良するための共通の基盤となりうると考えられます。進歩の速い LLM を同じ課題と手順で継続的に評価することで、性能の変化やモデルごとの特徴を把握し、化学工学の専門知識と AI を組み合わせた設計支援技術の発展に貢献することが期待されます。

今後は、本ベンチマークを軸として対象とする装置や設定項目を拡充し、実際の化学プロセス設計に役立つ AI の開発を支える評価方法へと発展させることをめざします。

4．研究プロジェクトについて

日本学術振興会（JSPS）科学研究費助成事業 基盤研究（C）23K03942 の支援を一部受けています。

＜用語解説＞

※1 化学プロセスシミュレータ　物質やエネルギーの移動、反応、分離などを計算し、製造工程の状態を予測するソフトウェア。本研究ではオープンソースの DWSIM を使用。

DWSIM 公式サイト：

https://dwsim.org/

※2 大規模言語モデル（LLM）　大量の文章などを学習し、自然言語で書かれた情報を処理して、文章やプログラムなどを生成する生成 AI の一種。本研究では、OpenAI の GPT 系列、Anthropic の Claude 系列、Google の Gemini 系列から、API を通じて利用した検証当時のフロンティアモデルを含む計 9 種類のモデルを評価。

各 LLM 系列の関連公式サイト：

https://chatgpt.com/

https://claude.com/

https://gemini.google.com/

※3 ベンチマーク　複数の手法を共通の課題と基準で比較するための評価の仕組み。

※4 フローシート　反応器、分離器、熱交換器、蒸留塔などの装置と、それらを結ぶ物質の流れを表した工程図。本研究では、装置の設定値などを含む、シミュレーション用の計算モデルをさす。

※5 モル分率　混合物に含まれる特定の成分の物質量を、混合物全体の物質量で割った値。成分がどの程度含まれるかを表す。

＜研究者のコメント＞

本研究は、研究室での化学プロセス設計演習から着想を得ました。化学プロセスの設計には、装置同士のつながりを読み解き、適切な設定を選ぶ専門知識やシミュレータの活用が欠かせません。AI がどこまでこの作業を支援できるかを知るため、AI の化学プロセスの設計提案を計算結果で確かめる土台を作りました。化学工学の知識と AI の力を組み合わせ、研究者や技術者がより良い化学プロセス設計を検討できる環境づくりにつな

げたいと考えています。（高原　渉）

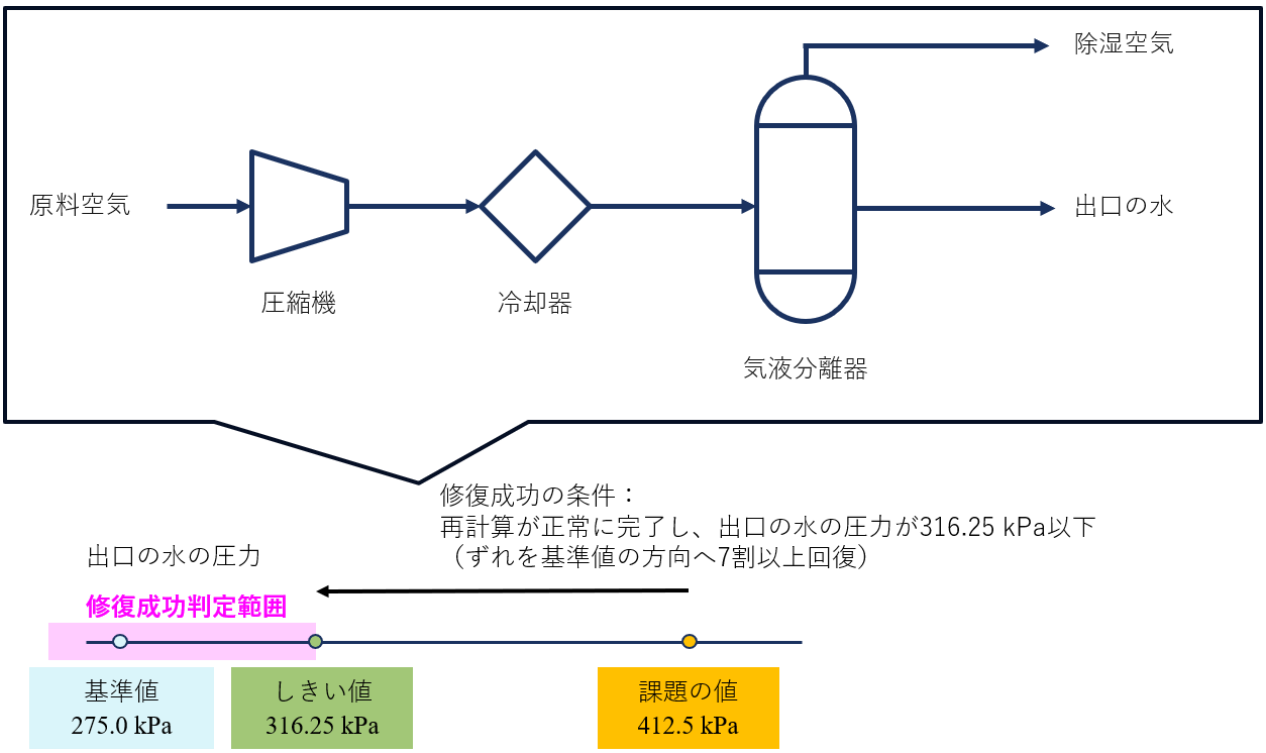
＜論文タイトルと著者＞

タイトル：PSE-RepairBench: A simulator-grounded benchmark for evaluating large language model-based setpoint correction in executable chemical process flowsheets（PSE-RepairBench：計算可能な化学プロセスフローシートにおける大規模言語モデルによる設定値修正を評価するシミュレータに基づくベンチマーク）

著　　者：Wataru Takahara、Ryu Suganuma、Souta Miyamoto、Katsuaki Tanabe

掲 載 誌：*Digital Discovery*　DOI：10.1039/d6dd00673f

＜参考図表＞



図：空気の除湿工程を用いた条件修正課題の具体例（簡略図）（作成：高原　渉）