

大規模言語モデル (LLM) を用いた DfX 学習支援の基礎的検討

Basic study of Learning Support of DfX using Large Language Model

○川村 泰世 (東京大学) *1 原辰徳 (東京大学) *2 太田順 (東京大学) *3

*1 Taisei Kawamura, The University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo, 113-8656, kawamura@race.t.u-tokyo.ac.jp

*2 Tatsunori Hara, The University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo, 113-8656, hara@race.t.u-tokyo.ac.jp

*3 Jun Ota, The University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo, 113-8656, ota@race.t.u-tokyo.ac.jp

キーワード: Design for X, 大規模言語モデル, 生成 AI, 対話型設計, サービスデザイン

1. 緒 言

1.1. 背景

設計プロセスにおいて肝要な上流設計をよりよく、焦点を絞って行うための設計方法の枠組みとして Design for X(DfX)が存在する。DfX は汎用的で具体的な設計方法を一体としてとらえた設計方法論の総称であり、生産性設計 (Design for Manufacturing: DfM) や品質機能展開 (Quality Function Deployment: QFD) などが代表格である。DfX の X は幅広く、製品を介したサービスの他、ユーザの感情や体験に焦点を当てるデザイン方法論も、広義の DfX といえる。

DfX の活用にあたっては特定の担当者だけでなく、設計に取り組むチーム全体による方法論の理解が求められ、学習が容易でないという課題がある⁽¹⁾。そのため、製品開発のための手法やツールの活用状況として捉えると、認識度の低いものがみられること、普及活動が経験的なものに留まっていること、および CAD や CAE などと比較するとまだ低調であることが報告されている⁽²⁾。

次に設計知識と大規模言語モデルの背景について言及する。学習に関わる知識創造プロセスにおいて、形式知と暗黙知がある。上流設計は概念的で自由度が高いことから、直感や閃き、人の信念などを含む「個人そのものに体化される認知スキル」としての暗黙知が多分に含まれるといえる。この暗黙知の学習過程として「個人が他者との直接対面による共感や、環境との相互作用を通じた暗黙知の獲得」が知られており、これは“共同化”と定義される⁽³⁾。

本研究では、DfX における暗黙知を共同化する手段として、大規模言語モデル (Large Language Model: LLM) を用いた対話的な支援が有効と考える。共同化を通じて DfX に対する理解を早期に獲得することは、DfX の普及に貢献し得る。LLM とは、近年急速に発展している人工知能技術の一種であり、大量のテキストデータを学習して人間が書いたような自然な文章を生成できる。LLM を用いた教育は多少の問題点は含みながらも効果的であると複数の先行研究から示唆されている。設計に対する LLM の適用可能性を調べた研究速報⁽⁴⁾⁽⁵⁾では、設計知識の有無と説明能力、アイデア生成能力、および一般的な設計過程にしたがった文脈に沿った対話能力に関する基礎的な報告がなされている。そこでは、設計に関する専門的な知識に関する学習不足が指摘されており、これは上流設計にも当てはまる。中でも、

DfX など、ある側面に焦点を当てた特定の設計方法論に知識に依拠した設計を対象とした LLM の検討は現時点では行われていない。

1.2. 本研究の目的

本研究の目的は、LLM を用いた DfX の学習支援に向けた調査である。具体的には、設計の演習の試行を通じて、LLM が持つ、DfX を適用する個別題材への応用性、対話の柔軟性、および効率性を調査するとともに、一般的な LLM の課題である専門知識の欠如による回答精度の低下への解決策の糸口を探る。

2 節では GPT-4 を用いた設計の調査方法と調査結果を述べ、現段階における LLM を用いた DfX の学習支援の性能と課題を明らかにする。3 節では 2 節で得られた課題を受けて、特定範囲の知識を強化する手法を用いた追加調査を行い、同様に考察を行う。4 節では、全体の結果をまとめる。

2. ChatGPT との概念設計による効果の調査

現段階において最も性能が良いとされる OpenAI 社の GPT-4 モデルの ChatGPT と共同で概念設計を行い、所要時間、対話の回数、うち DfX に関する対話の回数、設計案、学習効果に対する自己評価、および印象的なやり取りの抽出を行った。

2.1. 調査方法

ヒアリングから得られた研究室や職場環境に関する各課題 Y を解決する設計の演習実験を通じて、各 X を対象とした DfX の学習効果を調査した。慣れに対する効率性の差異を解消するため、設計の演習実験は各 X の順序を交換して行った。探索的な調査の側面が強いため、今回はまず著者自身が演習を行う。

この時、ユーザ (著者の一人) の目的は、最良の設計案を出すことではなく、設計を通じた各 DfX の学習であることに留意する。X は、Design for Cost (DfC), Design for Manufacturing (DfM), Design for Excellent Service (DfES) における C, M, ES の 3 種類とした。DfC と DfM はよく知られており多くの文献が存在するのに対して、DfES は比較的新しく、一般的な文献は少ない (詳細は 3 節で説明)。Y は「業務や研究に集中したい時に話しかけられることで集中

が途切れる」「作業机が狭い」「備品補充効率が悪い」とした。主な対話の流れは以下の通りである。

- 1) 条件指定と設計テーマのアイデア列举の依頼
 - ・【ユーザ】GPT-4 に X と Y を指定するとともに、設計を通じた当該 DfX の学習が目的であることを伝え、設計案の列举を依頼する (図 1)。
 - ・【GPT-4】所与の情報に基づき、課題を解決するアイデアの案を挙げる。
- 2) 設計テーマ決定と基本機能案の依頼
 - ・【ユーザ】当該 DfX の学習に使用したい設計テーマを GPT-4 に伝え、基本機能の記述を依頼する。
 - ・【ユーザ&GPT-4】対話し、基本機能を決定する。
- 3) DfX の観点による設計案の検討
 - ・【ユーザ】当該 DfX の観点に基づいた基本機能の解説や改善案を GPT-4 に依頼する。
 - ・【GPT-4】当該 DfX の観点に基づいた基本機能の解説や実現案をユーザに提示する。
 - ・【ユーザ&GPT-4】質問と対話を繰り返しながら、基本機能の構成と学習を共同で行っていく。
- 4) 概念設計案の出力
 - ・【GPT-4】ユーザから終了指示を受けて、行った対話を要約して、概念設計案を出力する。

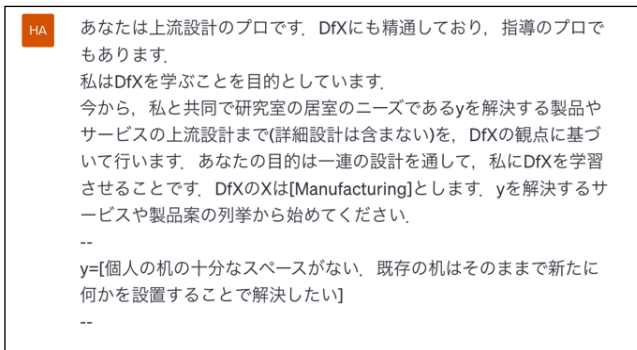


Fig.1 Prompt format at the start of design exercise

2.2. 調査結果

3つのDfXごとに、概念設計案が完成するまでの平均所要時間、ユーザとGPT-4間の対話の平均回数、うちDfXに関する平均対話数を表1に示す。後述する表2のテーマ毎による変動はあるものの、所要平均時間はC, M, ESの順に短くなり、対話の平均回数、うちDfXに関する対話の平均回数も同様に減少した。特徴的だったことは「無人運転ロボット」のような複雑なテーマでは対話が思うように進まず、時間を要した点である。対話回数は、ユーザによる入力とGPT-4による出力の総回数を集計しているため、対話回数2が一往復分に相当する。対話のうち、DfXに関する対話の平均回数も集計した。GPT-4と共同で作成した概念設計案のテーマは表2の通りである。

各DfXに関する学習効果の検証として事前に用意した質問紙を元に自己評価を行った。本研究では「記憶・理解」「分析・評価」「応用・創造」の3つの観点に対する達成度を1(初学者)~7(熟練者)の7段階に分けるとともに、各段階に該当する基準を設定した。「記憶・理解」の1,4,7の段階の基準を以下に示す。

- ・ 1: 認識できる主題(各DfX)の知識がほとんど、または全くない。
- ・ 4: 主題に関する基本用語や概念をいくつか認識しており、大方説明できる。しかし詳細な説明はできない。
- ・ 7: ある分野における専門知識を含めた全体を具体的な事例や歴史を踏まえて、体系的に説明できる。

Table.1 Attributes of observed design process

	Cost (DfC)	Manufacturing (DfM)	Excellent service (DfES)
平均所要時間(分)	42	38	30
対話の平均回数	19.3	18.6	14.7
DfXに関する対話の平均回数と割合	11.3 (59%)	11.3 (61%)	8.0 (54%)

Table.2 Themes of obtained conceptual design ideas

課題	DfC	DfM	DfES
業務に集中したい	通知システム	取り込み中インジゲータ	取り込み中インジゲータ
作業机が狭い	キーボードトレイ	壁取り付け型拡張机	AR Desk
備品補充効率が悪い	重量センサ付き棚	無人運転ロボット	備品管理アプリ

Table.3 Result of self-evaluation of learning each DfX

	DfC		DfM		DfES	
	前	後	前	後	前	後
記憶・理解	2	3	2	3	1	2
分析・評価	1	2	1	3	1	2
応用・創造	1	3	1	2	1	2

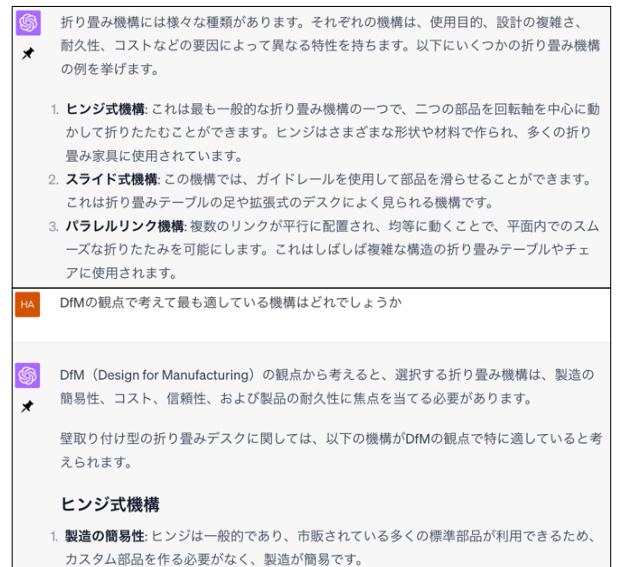


Fig.2 Example of notable conversation 1



Fig.3 Example of notable conversation 2

「記憶・理解」「分析・評価」「応用・創造」の3つの観点は、教育目標設定の課題の明示に広く活用されてきたブルーム・タキソノミーに基づいており、本研究の目的であるDfXの学習支援の効果と課題を探るのに適している。自己評価の結果を表3に示す。

GPT-4 との代表的な対話例をあげる。図2と図3は、「壁取り付け型拡張机」の設計中の対話例である。図2では、どの折り畳み機構を採用するか議論している。複数の機構例を列挙した後に、DfMの観点で案を選択し、その理由を解説している。図3では、ユーザが設計テーマの決定に迷った時に、最もDfMを学習できる設計例の選出を依頼し、GPT-4 がそれに応える様子である。

2.3. 結果の考察

本研究では計9回の設計演習を行い、全て1時間以内に終了した。これは人間の専門家などに依頼して学習を支援してもらう場合と比べて短い時間と考えられることから、効率性が良いことがわかる。また、表1によると全対話数に対するDfXに関連した対話数の割合はDfC、DfM、DfESの全てにおいて6割程度と高い数値である。このことからXに関わらず高い密度でDfXに関して議論がなされており、効率性を高める一つの要因であると考えられる。一方で、DfC、DfMの平均所要時間、対話の平均回数はほぼ同等で、DfESでは共に減少している。DfCやDfMについては、表3の自己評価の結果からもわかるように著者の「記憶・理解」が高かったため、対話を通じて深いところまで議論を行いやすかったためだと考えられる。加えて、テーマや課題が身近で日頃から接しているものであるほど深い議論がしやすいことも考えられる。

図2の例からも分かる通り、GPT-4には、テーマ、課題に沿ったアイデアや機能、改善案を各DfXに沿って具体的に解説する能力が一定以上ある。また、図3にあるように、GPT-4がDfX学習の観点では「壁取り付け型拡張机」のテーマが最も効果的と判断して選ぶなど、ユーザのニーズに応えた演習を行うことができた。これは環境との相互作用を利用した共同化の一種といえ、LLMが有する、DfXを適用する個別題材(ケース)への応用性と対話の柔軟性を示唆する。表3の結果を見ると、DfESに比べてDfCとDfMの学習評価が高く、これらにテーマの身近さも加味した共同化の効果が寄与していると考えられる。

一方、DfESでは詳しい説明を行わない傾向が見られた。詳しい説明を行うように依頼すると複数の当たり障りのな

い説明が生成された。DfESは、国際規格ISO/TS 24082 (JIS Y 24082) において文書化されたのが2020年と比較的最近の概念・手法であるためにGPT-4の知識が疎く、当該規格には則わず、Design for Excellent service という単語から一般に連想される説明や記述が生成されたと考えられる。この要因として、ISOやJISの国際規格は著作物でありパブリックドメインにない(一般に公開され誰でも入手・使用可能な状態にない)ため、LLMの構築に使用されていないことが考えられる。専門知識の欠如による回答精度の低下は先行研究でも指摘されており、全分野に関わるLLMの共通課題である。これは最新知識に限らず、外部に非公開のデータを加味した設計を考える上でも同様である。

3. Embeddings 手法を用いた追加調査

2節の調査により、GPT-4は、DfXの中でも一般的で広く知られているDfCとDfMについての知識を一定程度持ち活用できることがわかった。一方でDfESに対するGPT-4の知識が疎いことがわかった。

そこで本研究では、DfESを扱う当該JIS規格(JIS Y 24082:2021 サービスエクセレンス—卓越した顧客体験を実現するためのエクセレントサービスの設計)、およびその前提にあるJIS規格(JIS Y 23592:2021 サービスエクセレンス—原則及びモデル)に関する著者の連載解説記事⁽⁶⁾を読み込むことで専門知識を補う。このようにして回答の精度向上を試みることで、課題への対応策を追加調査した。

3.1. Embeddings を利用した LLM との対話の改良

Embeddingsとは、言語データ等をベクトル化する手法のことである。GPT-4などのLLMでは、各ドメインの追加データをベクトル化してデータベース(ベクターストア)に保管しておき、ユーザの入力と類似した文章をデータベースから検索してプロンプトに追加で埋め込むことで、回答の精度向上を試みる手法として用いられる。既に学習済みのLLMに新たな層を追加し再学習する手法であるFine-tuningとは異なり、低コストで簡便にカスタマイズに取り組むことができる。以下、本稿では、この回答精度向上の手法をEmbeddings手法、2節の手法(つまりEmbeddings無し)を通常手法と呼称する。Embeddings手法の概要を図4に示す。一般的な大まかな流れは以下の通りである。

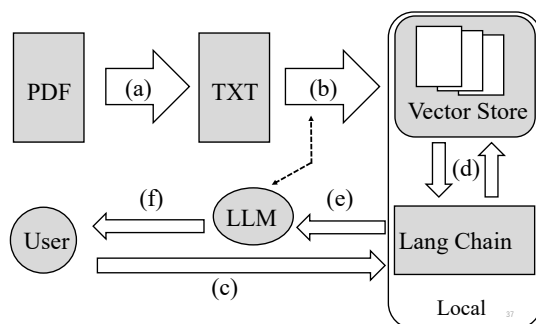


Fig.4 Overview of embeddings method

- (a) 追加データのPDFを読み込み、テキストを抽出する。
- (b) 抽出したテキストを分割し、Embeddingsモデルによってベクトル化したものをベクターストアに格納する。
- (c) ユーザからの新たな質問に、これまでの会話履歴を追

加したものを受け取る (Prompt と呼ぶ)。

- (d) 受け取った Prompt とコサイン類似度が高い文章をベクターストアから検索する。
- (e) Prompt に検索結果を追加したものを, LLM が応答生成を行う Completion モデルへの Prompt (入力) とする。
- (f) Completions モデルがユーザーに質問の答えを返す。

なお, このベクターストアはローカル環境で構築されているため, 学習用データは外部の人に公開されない。また GPT-4 の API を呼び出しているため, 内部の検索結果にあたる追加データが OpenAI 社にて再学習されることはない。

3.2. 調査方法

Embeddings 手法を用いて DfES への追加調査を行った。2 節と同様の流れを想定し, 同じ課題と質問を基本にした。

3.3. 調査結果と考察

以下の図 5 は, 通常手法にて DfES への解説を求めた際の出力である。このように DfES の用語から類推される一般的な内容に留まっているが, 対話は途切れることなく, 2 節と同様に一通りの設計プロセスと支援が行われた。



Fig.5 Explanation of DfES with default method

以下の図 6 は, Embeddings 手法にて DfES への解説を求めた際の出力である。ChatGPT のユーザーインターフェースではなく, Embeddings に関する独自プログラムを介した出力結果のため, 実行画面ではなくテキストを直接掲載する。

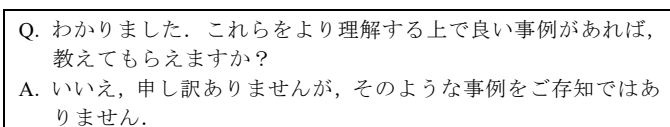
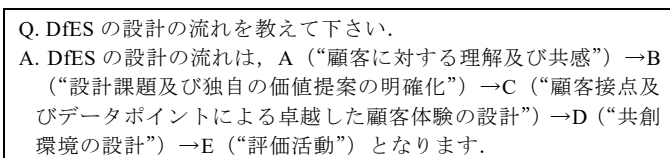
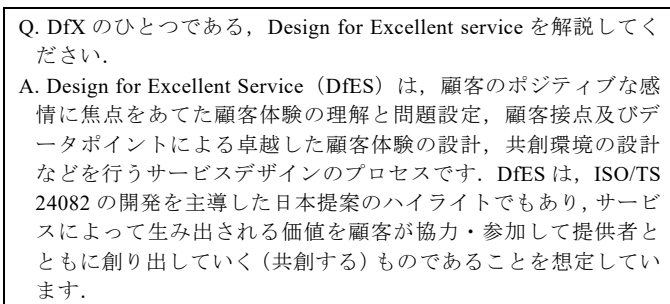


Fig.6 Explanation of DfES with embeddings method

Embeddings 手法により, 当該規格に則った DfES の説明がなされるようになった。追加データの抜粋ではなく, 幾つかの主要な記述を組み合わせ, 大きな違和感も齟齬もなく, 新たな説明文章が生成されている。2 番目の設計の流れについては, A-E の記号表記自体は追加データ内であったが, 本結果で初めて直列的な記述がみられるなどの工夫があった。このように, Embeddings 手法によって, 特定の知識を含んだ返答が得られ, 回答精度が向上し得る。

一方で, 2 節でみてきたような, DfX を適用する個別題材への応用性と対話の柔軟性の点で課題が残る。本稿で行った調査では, GPT-4 の表現の自由度 (ランダム性) に関するパラメータを一律に低くした (元の追加データの表現に近い出力を極力保つよう指示している)。そのため, アイデア列举や設計案の構成, あるいは追加データに直接書かれていない内容を尋ねた場合には, 3 番目の結果のように回答が得られず, 対話が止まってしまった。

追加調査での設計プロセスでは, 特定の追加知識そのものをうまく引き出して学習する状況, 個別題材に即した説明で理解を深める状況, およびアイデアを幅広く検討する状況 (発散) などで回答の効果度に違いがあった。今後, Embeddings 手法に基づいてより発展させていく上では, これらの状況の違いに応じて動的にパラメータを調節していくような仕組みにより, 専門知識の補完と, 個別題材への応用性・対話の柔軟性とを両立させていく必要がある。

4. 結 言

本研究では上流設計学習支援の LLM 使用に関する基礎的調査を行った。LLM は DfX を適用する題材の個別ケースへの高い応用性と対話の高い柔軟性を生かしてユーザーと知識の共同化を行う能力があることがわかった。また, LLM が持つ専門知識の欠如といった一般的課題を解決する糸口を見つけるべく, Embeddings 手法を用いて DfES に関する追加調査を行った。その結果, 特定文書にある専門知識を補完した回答が可能になったが, その回答精度を優先すると, 設計支援に必要な応用性と柔軟性が落ちることが明らかになった。これらの両立の仕組みに今後取り組む。

文 献

- (1) Kuo, T. C., Huang, S. H., and Zhang, H. C.: Design for Manufacture and Design for ‘X’: Concepts, Applications, and Perspectives, Computers & Industrial Engineering, Vol.41, No.3, pp.241-260, 2001.
- (2) 野間口 大, 高見 真史, 阪口 杏奈, 藤田 喜久雄: 産業界における設計工学関連の手法とツールの活用状況の調査研究, 設計工学, Vol.55, No.1, pp.43-60, 2020.
- (3) 野中 郁次郎(著), 竹内 弘高(著), 梅本 勝博(翻訳), 知識創造企業, 東洋経済新報社, 1996.
- (4) Kocaballi, A. B.: Conversational AI-Powered Design: ChatGPT as Designer, User, and Product, arXiv, 2302.07406, 2023.
- (5) Wang, X., Anwer, N., Dai, Y., and Liu, A.: ChatGPT for design, manufacturing, and education, Procedia CIRP, Vol.119, pp.7-14, 2023.
- (6) 原辰徳:連載 サービスエクセレンス規格で目指す組織のレベルアップと事業革新—第1回 サービスエクセレンス入門—, J-Club NEWS, Vol.29, No.2, pp.6-9, 2022.