

生成 AI を活用した充填包装機ของผู้ใช้支援 システムの開発（第 1 報）

大野 宏* 石澤 賢太* 長谷川 直樹* 丸山 隆成** 本間 克美** 滝原 正貴** 澤田 恒輝**

Development of a User Support System for Filling and Packaging Machines Using Generative AI (Part 1)

OHNO Hiroshi*, ISHIZAWA Kenta*, HASEGAWA Naoki*, MARUYAMA Takashige**, HONMA Katsumi**
TAKIHARA Masaki** and SAWADA Kouki*

抄 録

株式会社悠心が販売している製造装置に関するカスタマーサポートでは、客先トラブルの電話対応に長時間を要し、担当者も長時間拘束される。この課題を解決するため、生成 AI を活用し操作マニュアルやトラブル対応マニュアルを参照し、ユーザーの質問に回答し支援するシステムの開発に取り組んだ。まずは、入社間もない社員が支援システムを使うことで、トラブル対応をベテラン社員並みにできるシステムを開発した。さらに、回答精度の向上と想定外の質問への対応を進め、最終的には顧客が直接使用できるシステムへの発展を目指す。

1. 緒 言

株式会社悠心の主力製品である液体充填包装機は、連続するフィルムを袋状に加工するとともに、一定量の液体を充填し、温度と圧力を一定時間かけて溶着・切断する。充填する液体の種類・状態により袋の温度が変わり、またロットの品質のばらつきにより、溶着の最適条件が変わることもあり、液漏れがまれに発生する。トラブルが発生すると、顧客からの電話に対して、装置に詳しいサポート担当者が状態を聞き対策を伝える。この際、顧客の説明が不十分で状況が正確にサポート担当者に伝わらないことがあり、復旧までに時間がかかり、担当者も対応で拘束される。また、担当者により助言内容が若干変わることもある。

そこで、日々性能が向上している生成 AI を活用し、カスタマーサポートの課題を解決するため、客先の現場作業員に対し、装置の操作方法とトラブル対応方法を提示する支援システムを開発した。顧客が装置の操作やトラブル対応を

円滑に行える環境を構築し、対応の迅速化による高付加価値化とサポート業務の効率化を目指す。まずは、入社間もない社員でも本システムを利用することで、顧客からの質問にベテラン社員並みの回答ができるようにする。これにより、多くの社員が効率よく顧客対応が可能になり、ベテラン社員は開発等にかかる時間を確保できる。次年度は、想定外の質問や外国語にも対応できるように改良し、顧客が直接質問し、回答を得てトラブルに対応できるシステムを開発する予定である。将来的には、顧客はトラブルが発生しても 24 時間・祝休日も利用できることを目指す。

2. システムの概要

開発するシステムの概要を図 1 に示す。液体充填包装機の現場における操作支援およびトラブル対応の迅速化を目的として、生成 AI を用いた質問回答機能を構成の中心とした。システムは、操作方法マニュアルやトラブル対応マニュアルを参照して質問に回答し、音声でも入出力できるようにした。

* 技術統括センター

** 株式会社悠心

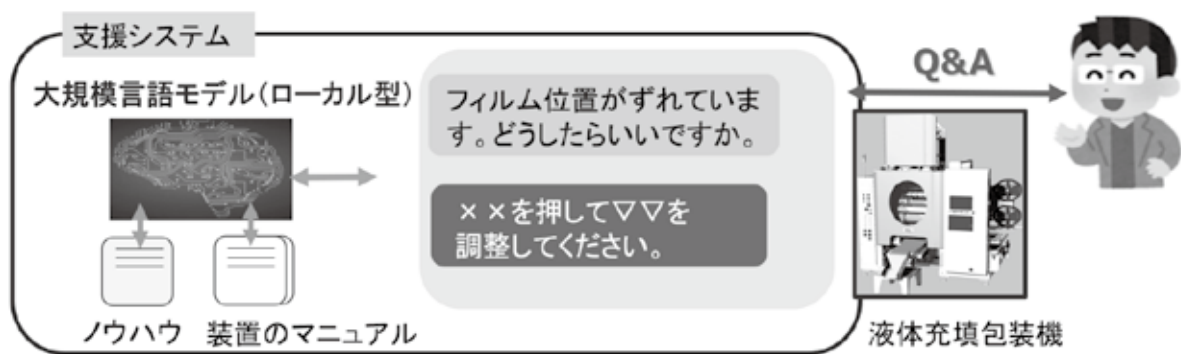


図1 システムの概要

回答を生成する大規模言語モデル(Large Language Model : LLM)は、社内のパソコンにダウンロードして使うローカル型の大規模言語モデル Llama-3-ELYZA-JP-8B¹⁾を利用した。このモデルは若干古いものの日本語に強く回答速度が速いことが特徴である。ローカルモデルを使う場合は、社内情報をクラウドに上げる必要がないために情報漏洩の心配がなく、本システムをインターネットに接続できない工場でも使うことができる。

3. マニュアルの整備

既存の操作マニュアルには、装置の操作方法、設定方法、取り扱いの注意点などが記載されているが、顧客は不明な点があると、マニュアルを読まずに電話で問い合わせることが多い。そのため初版から改訂があまりなされておらず、生成AIが誤解してしまう表現や情報不足が確認された。そこで、曖昧表現の排除や文言の統一など、生成AIの知識基盤として再整理した。

具体的には、各部の名称と働きにおいて、機能の詳細説明があるものの、基本的な役目の説明が不十分な場合は、これを追記した。また、同じ機能について、違う単語が使われていることがあり、適切な文章が検索されやすいよう、これらを統一した。

トラブル対応では、症状→原因候補→確認→対処の流れが重要であるため、液漏れを含む代表的トラブルを対象に、現場で実施できる確認

方法と対処手順を記したマニュアルを作成した。また、症状別にマニュアルを作成し、1個ずつ確認を行い、該当する場合は対処し、症状が解消されたか確認するようにした。例として「横シールが溶着しない」、というトラブルについて、図2に対応フローチャートを、図3に対応マニュアルを示す。トラブル対応マニュアルを読んで1つずつ原因候補を確認するようLLMに指示すると、フローチャートのよう1項目ずつ確認しトラブルを解決することができる。

4. 質問から回答を生成するシステムの開発

4.1 概要

本章では、生成AIを用いてユーザーからの質問に回答する仕組みを設計・実装した内容を述べる。回答の一貫性と現場で実行可能な具体性を満たすことを目標とした。システムは、まず質問から操作に関する内容か、トラブルに関する内容かを判定する。質問が操作に関する内容の場合は操作マニュアルを参照し、トラブル対応に関する内容の場合は、対象のトラブル対応マニュアルを参照して回答する。

4.2 質問分類機能

質問から、装置の操作方法や設定方法に関する内容か、トラブル対応に関する内容かを正確に分類できれば、参照する文章が限定されるので、回答精度が向上する。そのため、質問を分類するプログラムを開発した。当初は、LLMに

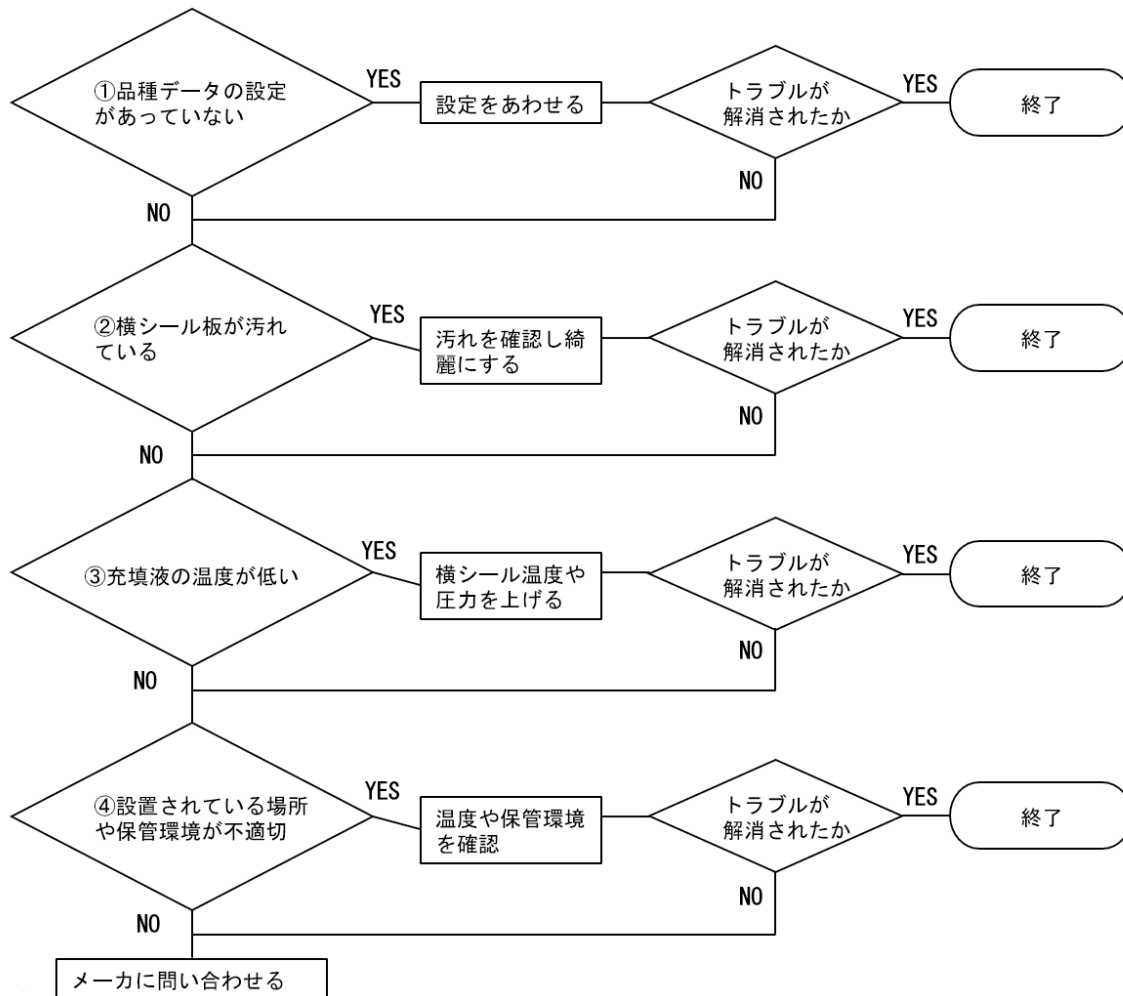


図2 トラブル対応フローチャート(横シールが溶着しない)

現象：横シールが正しく溶着しない。そのため液漏れが発生する可能性がある。その原因として、次の4個が考えられる。

原因候補1：フィルムと品種データの番号があていない。

確認方法：フィルムと品種データの番号はあっているか確認する。データ番号や生産条件の確認は自動運転モニタの画面で行う。

対処手順：フィルムと品種データの番号があっているか照合し、解決できるか確認する。

原因候補2：横シール板にフィルムや液が付着して汚れている。

確認方法：横シール板にフィルムや液が付着して汚れていないか確認する。横シール板の汚れは次の画面を参考に確認する。

対処手順：横シール板にフィルムや液が付着して汚れている場合は、板表面の汚れを除去し、解決できるか確認する。汚れは濡らしたメラミンスポンジで清掃する。

原因候補3：充填液の温度が通常より低くなっている。

確認方法：機械左側の液温度パネルを見て液温が基準より低下していないか確認する。

対処手順：液温を直接上げることは難しいので、横シール温度を上げるか、横シール周速を伸ばし、解決できるか確認する。横シール温度の設定は自動運転モニタの画面で行う。横シール周速の設定は生産条件1設定で行う。

原因候補4：装置が設置されている場所やフィルム保管環境が寒冷な状況である。

確認方法：装置が設置されている場所やフィルム保管環境が寒冷な状況でないか確認する。理想の温度は20℃～25℃である。

対処手順：縦シール温度やプリヒート温度の設定を高くし、解決できるか確認する。縦シール温度の確認と設定は生産条件1設定の画面で行う。

* 4つの原因全てを確認して解決されない場合は、メーカーに問い合わせる。

図3 トラブル対応マニュアル(横シールが溶着しない)

質問を分類させようとしたが、「横シール温度の設定方法を教えてください。」という設定方法の質問をすると、横シールという単語に引きずられて「横シールが溶着しない」というトラブルと分類されてしまう。そこで、トラブル対応に関する質問には、「どうしたらいいですか。」という定型表現を含める運用とし、当該表現が含まれる場合のみ類似度計算によって、どのトラブルに相当するか判断する方式に改めた。

4.3 RAG の利用

生成 AI は汎用知識を持つ一方、装置固有の仕組み、操作方法や設定方法など、LLM が学習していない事には、正確に答えられない。そのため、操作マニュアルの質問回答では RAG (Retrieval-Augmented Generation : 検索拡張生成) を用い、ユーザーの質問に関連する箇所をマニュアルから 10 件検索し、これを基に回答を生成する。具体的にはマニュアルを 50~100 字程度に分割し、エンベディング関数でベクトル化し、質問文との類似度を計算し、その値が高いものを参照した。また、回答時に根拠箇所を提示可能とすることで、利用者が参考にしながら作業できるようした。なお、操作マニュアルはワードで作成され PDF に変換しているが、補足情報としてページ数から PDF の該当ページを表示させるため、ページ情報を持つ PDF からテキストを抽出した。

4.4 ICL の利用

本システムの LLM である Llama-3-ELYZA-JP-8B は LLM の中では比較的小さなモデルであり、液体充填包装機自体の知識が不足しているせいか、当初は回答精度が低かった。そのため、図 4 に示す装置の概要を毎回読ませるようにした。すると、装置の概要を踏まえて回答を生成するため、回答精度が大幅に向上した。これは、ICL²⁾(In-Context Learning) とよばれる方法である。LLM には、入力できる文字数に制限がある

概要：本装置は、タレを充填した小袋を製造する装置である。ロール状のフィルムを供給し、折り曲げて端部を重ね合わせ、熱と圧力で溶着して袋状に成形する。次に液体を充填し、上部を溶着した後、切断して製品を完成させる。

図 4 装置の概要

ので、装置の概要を 100 字程度に簡潔にまとめた。

4.5 会話履歴の参照

人間同士の会話では、「さきほど質問した件について」や「この回答をもう一度分かりやすく説明して」など、過去のやり取りを踏まえて話を進めるのが一般的である。生成 AI との対話でも、過去の経緯や前提が引き継がれない場合、背景説明の繰り返しや認識違いが起こりやすい。特にトラブル対応では、確認すべきチェック項目が複数あるため、「項目 1 と項目 2 を確認済みであること」を踏まえ、次に「項目 3 の確認」を促すといった手順に沿った応答が求められる。

本システムでは、トラブル対応の場合、会話履歴の参照範囲を直近 10 往復分（ユーザーの質問とシステムの回答を 1 往復として数える）とする設定にした。これにより、直近の流れを維持しつつ、LLM への入力文字数の上限を超えにくい形で運用できる。一方、操作マニュアルに関する質問回答では、ほぼ一問一答で終わることが多いので、少なめの 3 往復分にした。

4.6 音声入力と音声出力

本システムに質問を音声で入力する機能と、回答を音声で出力する機能を追加した。音声入力では、マイクで音声を録音し、音声認識モデル faster-whisper³⁾で日本語テキストに変換する。これは OpenAI が開発したモデルで、高速かつ省メモリで動作する。モデルのサイズは 5 種類あり、大きいサイズのモデルは認

❖ 質問を入力してください。

🔊 音声入力

「充填液量の調整方法を教えてください。」

❖ 充填液量の調整方法は、7-2-2-7 生産条件3の「液絞り設定」で行うことができます。具体的には、液絞り回転速度、液絞り隙間(充填時)、液絞り隙間(停止時)を設定することで、充填液量を調整することができます。

🔊 音声出力

図5 支援システムの画面の例(操作方法に関する質問)

識精度は高くなるが動作に時間がかかるため、GPUがあると快適に動作する。音声出力は、テキストを読み上げるアプリ VOICEVOX⁴⁾を利用した。これを起動した状態で、Python プログラムからテキストファイルを渡すと音声ファイルに変換し、音声として出力される。設定番号を変えると、様々なキャラクターの音声を使うことができる。

4.7 支援システムのプログラム

システムは Python で開発し、ChatGPT のような GUI(Graphical User Interface)を簡単に構築できる Chainlit⁵⁾を利用した。実際の画面の例を図5に示す。プログラムを起動すると、各種モデルなどの読み込みを行い、準備が整うと「質問を入力してください。」と表示される。「充填液量の調整方法を教えてください。」と質問すると、操作方法に関する質問なので操作マニュアルを検索し質問と関連の高い部分を参照し、LLM が出力した回答を表示する。この時、検索された参照部分も表示することができ、この調整方法が記載されている操作マニュアルの 7-2-2-7 を表示する。また、質問を音声で入力する時に使用する「音声入力」ボタンと、回答を音声で出力する時に使用する「音声出力」ボタンが表示される。これを押すと音声で質問を入力し、音声で回答を出力する。

5. 精度検証と改良

実際の運用を想定し、システムの精度検証と改良を行った。100個の質問と理想の回答を考え、実際の回答と比較したところ、10問で満足する回答が得られなかった。これらが不正解となった原因を分析し、回答精度向上に向けた改良を実施した。

操作マニュアルに関する質問では、RAGを用い、ユーザーの質問に関連する箇所をマニュアルから10件検索し、これを基に回答を生成している。質問に対応する箇所が正しく検索されれば、正しい回答が生成されるので、この検索が鍵になる。RAGの対象となる文章は4.3節に述べたように、文字を定数ごとに区切り、前後をオーバーラップさせる方法が一般的であるが、文章が途中で途切れてしまう。そこで、文章を句読点や改行で文章の切りの良いところで区切り、文頭にマニュアルの章と節をつけて、その文章が何に関するものかより明確にすることで、不正解を減らすことができた。

トラブル対応に関する質問では、4.4節で述べたICLの利用により回答精度が上がった。また、参照ファイルの文字数が多くなると、指示文の指示を守らずに、すでに確認した項目を再度確認するよう回答する。そのため、再度、明確に指示をした。最初の指示文は、「参考情報と過去の会話履歴を踏まえて、で

きるだけ正確に回答してください。」と「原因候補を1つずつ確認するよう回答してください。」であったが、これに「すでに確認したことを再び確認しないようにしてください。」を加えたところ、正しく回答するようになった。

誤回答する質問については、マニュアルに記載がなかったり、表現が分かりにくかったりするものがあり、その都度加筆修正した。

6. GENIAC-PRIZEへの申請

経済産業省とNEDOは「国内の生成AI開発力を底上げして、社会実装まで加速させる」ために、プロジェクト生成AIアクセラレータ・チャレンジ⁶⁾(Generative AI Accelerator Challenge : GENIAC)を令和4年度から実施し、国内企業の言語モデルや視覚言語モデルの開発を支援している。また、令和7年度にGENIAC-PRIZEという、実証・コンテスト型プログラムが実施された。生成AIの利用を活発にするためには、現場のニーズに基づき活用されるシステムの開発が必要であり、生成AIアプリやサービスを開発し実証したものが応募対象となる。募集対象は4分野で、領域1「国産基盤モデル等を活用した社会課題解決AIエージェント開発」の「カスタマーサポート生産性向上」に本システムの研究開発を申請し、特別賞（地域賞）を受賞した。

7. 結 言

- (1) 生成 AI を活用し、マニュアルを参照して質問に回答する支援システムを開発した。本システムにより、入社間もない社員でも高い水準のカスタマーサポートが可能となった。
- (2) 回答精度が上がるようにマニュアルの加筆修正を行った。
- (3) 直接ユーザーが本システムを活用してトラブル対応ができるよう、回答精度の向上、予期せぬ質問へ対応など、今後も改良に取り組む。

参考文献

- 1) ELYZA, <https://huggingface.co/elyza/Llama-3-ELYZA-JP-8B>, (閲覧 2026-2-11)。
- 2) Adrian de Wynter, “Is In-Context Learning?”, arXiv:2509.10414, 2025.
- 3) faster-shisper, <https://github.com/SYSTRAN/faster-whisper>, (閲覧 2026-2-11)。
- 4) VOICEVOX, <https://github.com/VOICEVOX>, (閲覧 2026-2-11)。
- 5) Chainlit, <https://github.com/Chainlit/chainlit>, (閲覧 2026-2-11)。
- 6) GENIAC, https://www.meti.go.jp/policy/mono_info_service/geniac/index.html, (閲覧 2026-2-11)。