

業務DXでの生成AI活用 クローズド環境下のLLM試作事例紹介

九州工業大学 情報統括本部
中村豊, 井上純一, 富重秀樹, 畑瀬卓司

◆ 背景

九州工業大学では「本学が2040年に在りたい姿を示す」として【九州工業大学ビジョン2040】という大方針が策定された。

ビジョン達成にむけ、教育研究、社会貢献および法人の組織・経営に教職員がそれぞれの立場で行う**17のアクション**が設定されている。



◆ 背景

情報統括本部に課せられたアクション:

DX・業務標準化を核としたスマートキャンパスの実現

すべての業務プロセスに対して整流化・標準化を実施し、紙媒体や手作業での決裁といったアナログ要素に関して全てデジタル化したうえで、職務や立場の違いを越え全学一致でDX・業務標準化を実現します。そのためのDX環境構築に際しては、システムそのものの維持や管理にかかる要素の効率化、品質向上及び属人化の解消を図り、柔軟性と創造性に富んだシステムを構築します。また、様々な形式のデータを分析・活用可能なストレージ基盤であるデータレイクシステムを構築し、Kyutech-DXビジョンを実現するとともにスマートキャンパスの実現を促進します。

デジタル化技術を使って全学の業務プロセスを根底から見直 せ す。
新たに構築するシステムそのものに対しても適用 もろ される。

◆ DX・業務標準化にむけた課題

課題1:DXの多様性

一言で「DX」と言っても、その手法は多種多様。

課題2:業務の多様性

業務内容・フローが部署ごとに異なる。一律の標準化は困難。

課題3:文化的・心理的抵抗

紙からデジタルへの移行(変化)に対する不安、古技術への愛着。
特に属人化された業務が障壁。

◆ 課題に対するアプローチ

① 段階的・選択的導入

標準化しやすい領域から着手し、成功体験を積み重ねる。

② 現場との合意形成

現場の声を丁寧に拾い、変化の目的とメリットを共有する。

③ 柔軟性を持った標準化設計

一律ではなく、部署ごとの特性に応じた運用を許容。



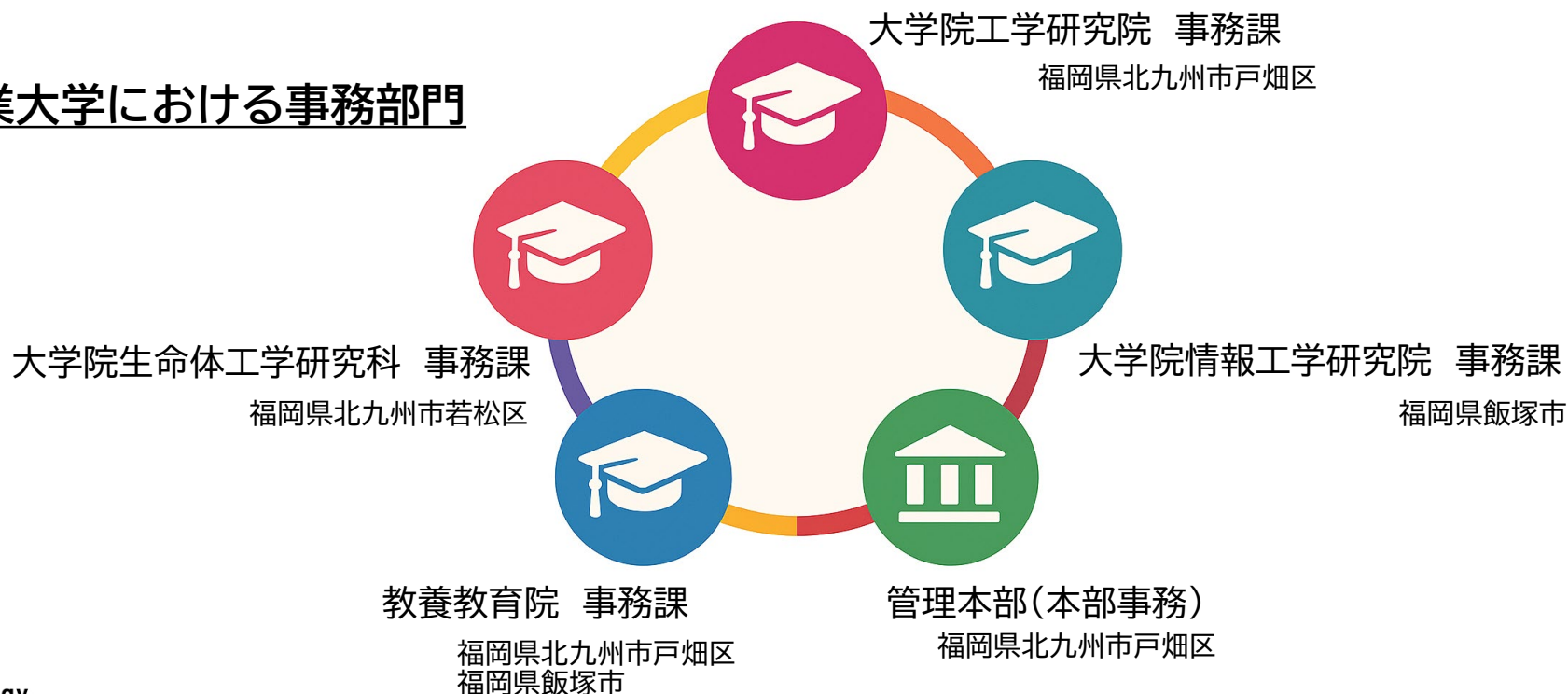
事務部門からスタートさせよう！

◆ 事務部門からの開始を選択した理由 ①

【共通性】が多い＝【標準化】しやすい

部・課を問わず【業務】と【課題】に共通する要素が(比較的)多い。

九州工業大学における事務部門



◆ 事務部門からの開始を選択した理由 ②

【波及効果】が望める

定型業務を標準化・DX化できれば事務部門全てに効果が広がる。
インパクトと実現性を両立することができる。

~~学内に説明しやすい~~

事務作業を効率化することで各部門の担当者がコア業務に集中する時間を増やすことができる＝全体的な生産性が向上する＝昨今問題になっている教員の教育・研究時間を奪う事務処理時間を削減することが期待できます！



◆ 事務部門からの開始を選択した理由 ③

このままだと事務部門が破綻する

定常業務(問合せ対応や様々な文章作成、データ入力など)に多くの時間を費やしており、非定常業務(個別判断を要する業務や新しい・突発的な純増業務)に対応する時間が確保できていない。

職員数の未充足率が深刻化しており、

- ・情報共有しようにも相手がおらず、業務プロセスの属人化が加速。
- ・一部職員に対する負荷が増大。

などが勤務時間記録簿などの解析で判明している。

◆ 生成AI利用を選択した理由

定常業務を支えることが可能(なはず)

大量のデータを迅速に分析し、精度の高い情報を出力する特性を活かせば、業務プロセスを大幅に効率化することが可能(なはず)

- ・文章生成

報告書やメール、議事録などの文章生成を高速・効率化できる。

- ・情報検索

業務に必要な情報を自然言語で検索でき、要約して提示してくれる。

- ・問い合わせ対応の自動化

各種規則やFAQのような問い合わせに24時間自動で即時対応できる。

◆ 生成AI利用のデメリット(当初予想)

- セキュリティ上のリスク

外部ネットワークと接続された環境下での利用には機微情報の流出(学習も含む)といったリスクがある。

閉域網内で構築しよう！

- ハルシネーションの問題

自然言語での対話可能なのが売りだが、対話である以上、PDFなどをベースとしたチャットボットと違って回答が揺れる可能性がある。

許容範囲内に収まることをお祈りしよう！

- そもそも…素人にできるのか？

やってみないとわからない。

まずはミニマムスタートしてみよう！

◆ クローズド環境下で情報検索と回答を生成するAIを試作①

- 利用方法

学内限定でウェブブラウザを用いてアクセスする方法を採用。

- ナレッジ

本学の規則を全て学習させてみる(全15編、800項程度)

- 構成(ハードウェア)

新規購入は ~~できない~~ しない(あくまでも試作。運用は想定していない)。
情報統括本部に所属する教員の研究用サーバを間借り。

Intel Xeon Silver 4110

NVIDIA GeForce RTX 2080Ti

DDR 32GBx6

SSD 1TB(M.2)x2



◆ クローズド環境下で情報検索と回答を生成するAIを試作①

- 構成(ソフトウェア)

OS: Ubuntu 24.04.1 LTS

プラットフォーム: (最終的に) ollama

LM Studioやllama.cppなどを試したが、最終的にollamaを採用。

LLM: gpt-oss-20b

Llama-3.1-8Bなどを試したが、最終的にgptを採用。

◆ クローズド環境下で情報検索と回答を生成するAIを試作①

想像以上に簡単に構築できた

過程を紹介する必要性を感じないほど簡単。

Ubuntuへのollamaインストール	非常に簡単
ollamaへのLLMインストール	非常に簡単
ollamaとDifyの連携	けっこう簡単
テストLLMの作成と稼動	簡単



ローカルで動作するLLM、準備完了
これ、いけるんじゃない？

◆ クローズド環境下で情報検索と回答を生成するAIを試作①

ハルシネーションとの闘い

学内規則を学習しているのに正解を導き出せない。

びっくりするぐらい堂々と嘘をつく

国立大学法人九州工業大学基本規則

第1章 第2条

九州工業大学は、法人法第4条の規程に基づき、主たる事務所を福岡県北九州市戸畑区仙水町1番1号に置く



※規則に「ロサンゼルス」という文言は存在しない

◆ クローズド環境下で情報検索と回答を生成するAIを試作①

ハルシネーションの緩和を目指して

見直し①： 学習させるデータの書式を整形して統一

複数のファイルフォーマットを検証 (Word、Text、Markdown…)

加工しやすさと今後の拡張性を考えてtxt形式で統一。

公開されている規則データはword形式だが規則性がない(バージョンなどが異なる)状態だった。

形式変換ついでに手作業で再入力を実施(数字や空白が半角だったり全角だったりも修正)。

見直し②： チャンキング

長文を短文に分割。句読点などで改行処理を統一。

規則を途中で分割すると意味が変わってしまう危険性は今回は黙認。

関連する情報が複数のチャンクに跨ってしまうことが揺れの原因になる可能性についてもとりあえず黙認。

◆ クローズド環境下で情報検索と回答を生成するAIを試作①

ハルシネーションの緩和を目指して

見直し③： ナレッジ総量の自主制限して負荷を軽減

規則を編単位で分割し、複数のLLMを用意。

例： 人事課用(第3編のみ学習)のLLM、会計課用(第5編のみ学習)のLLMをそれぞれ準備して負荷軽減。



結果！



完全に解決したわけではないが、緩和はできた

◆ クローズド環境下で情報検索と回答を生成するAIを試作①

プロンプトエンジニアリングが必要

学習させた規則にある言葉をそのまま使わないと回答が揺らぐ。
「正しい答えを得るためには正しい質問から」とはいうが…



「事務所」を人の言う「事務所」と認識してはいない。

パラメーター数の大きなLLMを使うしか対策はなさそう…(未検証)

◆ クローズド環境下で情報検索と回答を生成するAIを試作①

回答の検証が必要

その回答が本当に正しいのかがわからない。



「法人法」はナレッジのなかに存在しないため
このLLMだけでは確認することができない。

見極めるのは使う側。 質問・回答に対するある程度の知識は必須。

◆ クローズド環境下で情報検索と回答を生成するAIを試作①

最新情報ではない

クローズド環境である以上、妥協するしかない。



知ってたらオフラインじゃないと思う。
それはわかってる。

汎用型LLMほど融通が効かない

学習内容を絞った&オンプレミスな特化型に対話を求めてはいけない。

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

卓上LLMはどうだろう…ということで切り口を変えて第2弾

- 利用方法

ollamaに直接アクセスする方法を採用。

- ナレッジ

本学の規則を全て学習させてみる(全15編、800項程度)

- 構成(ハード&ソフト)

Apple MacBook Pro M4 Max 32GB

【ollama】+【gpt-oss-20b】

OpenWebUI ※dockerが簡単だったが紆余曲折の結果uv+python版

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

システム構築

第1弾同様、汎用型LLMのインストールは問題なく完了。

学習データの準備

ウェブ上で公開されていた規則を 勝手に 用いた第1弾とは異なり、全ての規則データを総務課より提供して頂いた(つまり本物)。

ファイル総数は1,000以上。word形式やexcel形式などバラバラ。

折角もらったが、正規化は時間的負担が大きすぎると判断し、word形式をtext形式へと変換した。

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

ハルシネーションとの闘い 再び

理解している規則もあるため、ナレッジ検索が緩すぎるのが原因と思われる。

生命体工学研究科の規則について教えてください

Q1 Kyutech_rule
8秒間の思考

申し訳ありませんが、現在提供いただいている情報源には「生命体工学研究科」に特化した規則の記載がありません。
最も近い関連情報は、九州工業大学大学院工学府の学修細則（令和7年2月26日改正）です。こちらは大学院全体の学習・研究に関する基本的な規定をまとめたものです。

“〇九州工業大学大学院工学府学修細則【1】”

生命体工学研究科に関する具体的な規則（例えば、動物実験の取り扱いや研究倫理、卒業研究の要件など）が必要な場合は、以下のように確認してください。

1. 学科・研究科の公式ウェブサイト
研究科ごとに専用の規定・細則が公開されていることが多いです。
2. 学科事務局や指導教員への直接問い合わせ
最新の規則や手続きについて直接確認できます。
3. 大学の規則集や統括規程
「大学院生命体工学研究科動物実験細則」など、研究科ごとの詳細規定が別途設けられている場合があります。

ご不明点や具体的な項目がわかれば、もう少し詳細にご案内できるかもしれません。どうぞお知らせください。

3 Sources

ないわけないだろ！ —このポンコツが！



◆ クローズド環境下で情報検索と回答を生成するAIを試作②

検索が緩い原因として考えられるポイント

組み込みモデルの場合、学習用データ(ファイル)をアップロードする際にベクトル化される(embedding modelを用いて)が、この時、標準のチャンクサイズに自動的に区切られている。チャンクサイズのオーバーラップ設定などが必要なのではないか。

学習用データ(ファイル)をアップロードする前に、組み込みモデルをどうかを決定しておく必要がありそう。

なお、OpenWebUIのデフォルト組み込みモデルは実用には耐えない。

埋め込み

埋め込みモデルエンジン

デフォルト (SentenceTransformers) ▼

埋め込みモデル

sentence-transformers/all-MiniLM-L6-v2



After updating or changing the embedding model, you must reindex the knowledge base for the changes to take effect. You can do this using the "Reindex" button below.

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

組み込みモデルの検討

ollamaで利用できる代表的な組み込みモデル

- embeddinggemma
- nomic-embed-text
- mxbai-embed-large
- qwen3-embedding
- kun432/cl-nagoya-ruri-large:latest

※名古屋大学産の、日本語に特化した組み込みモデル

検証(試行錯誤)を実施

【cl-nagoya-ruri-large:latest】でembedしてプロンプトで【qwen3-embedding】を使ってみる、などなど…

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

チャンクサイズの変更

標準設定では【chunk size 1000 overlap 100】だった。

学習用データ内の「1条文」が100トークン程度だったので【chunk size 100 overlap 50】に設定を変更してみた。



一般

コンテンツ抽出エンジン デフォルト ▾

PDF 画像抽出 (OCR) ☐

埋め込みと検索をバイパス ☐

テキスト分割 デフォルト (文字) ▾

チャンクサイズ

100

チャンクのオーバーラップ

50

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

トップ N・リランキングの変更

トップ Nの標準設定では【Top3】だったので【Top10】に増量してみた。
リランクのトップ Nも標準設定の【Top3】から【Top 5】に増量してみた。
ハイブリッド検索は無効だったので有効化した(BAAI/bge-reranker-v2-m3)。

検索

フルコンテキストモード ☐

ハイブリッド検索 ☒

リランクエンジン デフォルト (SentenceTransformers) ▼

リランクモデル BAAI/bge-reranker-v2-m3

トップ K 10

トップ K リランカー 5

関連性の閾値 0

注意：最小スコアを設定した場合、検索は最小スコア以上のスコアを持つドキュメントのみを返します。

BM25の重み ☐ カスタム 0.5

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

組み込みモデルの最終的な設定

- embedding model 【qwen3-embedding:8b】
- chunk size 100 over 50
- ハイブリッド検索【有効】
- リランクモデル【BAAI/bge-reranker-v2-m3】
- Top K 【10】
- Top K reranker 【5】

これらの設定を保存したのち、学習用データ(ファイル)のアップロードを実行。

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

① ワークスペース上でナレッジスペースを【新規作成】

② 学習用データ(ファイル)をアップロード

Tips:

ネットワーク経由で行うとOpenWebUI内のsqliteがストールしてアップロードが失敗することがある。オンプレでpython版を構築して回避した。

③ モデルを生成

ナレッジベースの設定で②で作成したものを指定。

ベースのLLMはgpt-oss-20bとした。

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

結果！

九州工業大学の主たる事務所はどこにありますか？ナレッジベースからの情報のみを用いてください

Q1 Kyutech_rule_qwen
2秒間の思考 ▼
九州工業大学の主たる事務所は、福岡県北九州市戸畑区仙水町1番1号に置かれています。 0100001000基本規則_25_07_30.txt
4 Sources

パートタイム職員の勤務時間について教えてください

Q1 Kyutech_rule_qwen
33秒間の思考 ▼
パートタイム職員の勤務時間は、九州工業大学の「パートタイム職員就業規則（平成16年九工大規則第6号）」で定められています。
その中で、1週間の勤務時間が **20時間以上** になる者は、別の規定（全日制職員扱いなど）に該当するとされています[^2]。
ただし、週あたりの最大勤務時間や日別の時間配分、休憩時間、給与計算方法などの具体的な数値は、今回の抜粋には明示されていません。
実際の勤務スケジュールは、所属部署や業務内容に応じて決定され、時間給で支払われるケースが一般的です。

要点まとめ

- パートタイム職員は「勤務時間等に関する規程（平成16年九工大規則第20号）」で定義されています。
- 週20時間以上の勤務者は別途規定対象となります。
- 週20時間未満であればパートタイムとして扱われ、給与は時間単価で計算されます。

[^2]
6 Sources



アホになる例。。。。

九州工業大学の主たる事務所はどこにありますか？ナレッジベースからの情報のみを用いてください

○ **Kyutech_rule_mxbai**
2秒間の思考 ▼

現在提供されているナレッジベースの情報には、九州工業大学の主たる事務所の所在地に関する記載が含まれていません。したがって、ここからは回答できません。必要であれば、別の情報源をご確認いただくか、大学の公式ウェブサイトなどでご確認ください。

5 Sources

九州工業大学の主たる事務所はどこにありますか？ナレッジベースからの情報のみを用いてください

○ **Kyutech_rules_ruri**
7秒間の思考 ▼

申し訳ありませんが、提示された情報源（source 1～5）には「九州工業大学の主たる事務所の所在地」に関する記載がありません。したがって、現在のナレッジベースからは回答できません。もし別の情報源や具体的な場所の手がかりがあれば教えていただければ、再度確認いたします。

5 Sources

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

- 総務課からもらったオリジナルデータ(word,excel,ppt等で構成)を直接 OpenWebUIにインポートしてみた
 - 18時間ぐらいかかった
 - 古いワードやら、いろいろインポート時にエラーが出た
- 以下の変換プログラムを作ってオリジナルデータとした
 - pptx → pdf
 - docx, doc, rtf → txt
 - xls, xlsx → xlsx
 - その他はコピー
 - ~\$ はコピーしない

◆ クローズド環境下で情報検索と回答を生成するAIを試作②

- 別表などはxlsxに格納されているので、txtのみのナレッジベースだと検索できない項目がある

教授の俸給表について教えてください。ナレッジベースからの情報のみを用いてください

OI Kyutech_rule_qwen

17秒間の思考

申し訳ありませんが、ナレッジベースに「教授の俸給表」に関する情報は含まれていません。必要な情報は別途ご確認ください。関連する資料を参照してください。

6 Sources



Kyutech_rules_orig_qwen

52秒間の思考

教授の俸給表について、ナレッジベースに記載されている情報は以下のとおりです。

項目	内容	備考
本給表（基本年俸）	<ul style="list-style-type: none">- 「教授相当」職種に該当- 博士（PhD）を有する教授の年間年俸は 7,320,000円、月給は 610,000円（7,320,000÷12）です。- 本給表は寄附講座教員・研究職員・特別教授・特任教授・特任准教授・特任講師・特任助教にも適用されます。	0400052000年俸制職員...アイル①②.i
昇給表（号俸表）	<ul style="list-style-type: none">- 別表第2（第5条関係）に教授用の昇給区分が記載されています。- 昇給区分は A～E まであり、昇給の号俸数は<ul style="list-style-type: none">* A：2以上* B：1* C：0* D：0* E：0- ただし具体的な金額は表からは直接読み取れません。	0400051910第二期年俸...01_別表.xls
適用対象職種	<ul style="list-style-type: none">- 本給表・昇給表ともに、教授・准教授・講師・助教の職種に適用されます。- 特別教授・特任教授・特任准教授・特任講師・特任助教も本給表の対象です。	0400052000年俸制職員...アイル①②.i

※ 上記の数字は、与えられた原始データから直接引用したものです。詳細な区分（例：年俸額の12分の1）や他の職種（助教相当の学士・修士）は別表に記載されていますが、教授に特化した情報は上記の表が主要なものになります。

3 Sources



関連質問

教授の年俸はどのように昇給されるのか、詳細な金額の変化はありますか？

教授以外の職種（准教授・講師・助教）の年俸表と比較した際の差はどのくらいですか？

昇給表（号俸表）の具体的な金額や適用条件について、もう少し詳しく教えてください。

◆ まとめ

- あくまでも支援機能だと割り切ることが大切。
全てを任せると絶対(←重要)に失敗する。
- 単純作業は任せられそうだが、利用者側の知識が必須。
AIの示す答えが「炒飯」なのか「ピラフ」なのかは人間の判断が必要。
併せて、プロンプトが書ける国語力の育成も必要。
- 正解率が8割ぐらいの精度でよければなんとかなるかもしれない。
9割以上の精度を求めると相当な労力が必要。
だが、労力に見合った成果が得られるかはかなり怪しい(実感)。