

# ローカル環境での生成 AI を活用する教材検討

## Teaching Materials Using Generative AI in a Local Computing Environment

宮治 裕<sup>†1</sup>

Miyaji Yutaka<sup>†1</sup>

<sup>†1</sup> 青山学院大学 社会情報学部

<sup>†1</sup> School of Social Informatics, Aoyama Gakuin University

キーワード：生成 AI, ローカル LLM, RAG

**Keywords:** Generative AI, Local LLM, RAG

## 1. はじめに

生成 AI のなかでも特に大規模言語モデル (LLM) の企業における業務内での活用が一般化し、実績が報告されるようになってきた。たとえば情報システム開発においては、既に生成 AI の活用がすすめられており、その適用範囲は要件定義・設計・開発・テストなどの全ての工程におよんでいる。それに伴い、開発期間の短縮と少人数化や開発ライフサイクルそのものにも変化を与え、業務への関わり方も変化し始めている。ICT 業界以外でも利用は活発化しており、商品企画での活用、社内情報検索、キャッチフレーズや説明文作成やデザインの提案など多岐にわたる。業務内容や職種にもよるが、既にその利用は前提であり、それを踏まえた DX や人手不足解消への取り組みがなされている。

その一方で、大学教育や学生の生成 AI の利用は限定的な印象がある。チャット AI の登場時には、学生の利用については是非が問われ、その後その全面的な利用を全学として推進するようなコメントが出されていないこともあり、基本は利用しないものという前提で捉えられているようである。その結果、以下の3パターンの学生に別れてしまっている。A: 生成 AI をまっ

たく利用しないで以前の手法で学習をする学生、B: 何も考えずに単純に生成 AI を「使うだけ」の学生、C: 生成 AI サービスの特性を把握し、適切な指示をした上で得られた結果を活用する学生である。チャット AI 登場時に恐れていたのは B タイプの学生の利用の仕方であるが、それ以上に気をつけなければならないのが A タイプの学生であろう。

社会が大きく変わろうとする中では、その社会に影響を与えているものを「知る」「体験する」ことによって本質的な部分を捉え、その「有効活用する方法」を学ぶ必要があるのではないだろうか。本取り組みでは、以上の背景から、様々な生成 AI を実際に利用する過程をととして、その知見を教材開発、教育環境に生かすことを目的とする。

## 2. 教材検討の条件

### 2-1. ローカル計算機環境下での LLM

この生成 AI を用いる際には、LLM の場合であれば一般的には性能の高いモデルである ChatGPT や Gemini, Claudeなどをサービスとしてあるいは API

をとおして利用することになる（これらのネットワークの向こう側（サーバー側）で動作する LLM を、便宜上クラウド型 LLM と記載する）。性能が高い反面その内部モデルは明らかにされておらず、その学習（事前学習）には大規模な計算機資源が必要となる。本取り組みでは、背景からもわかるように LLM 自体を新規に作ることを想定しない。

また、クラウド型 LLM を利用者としての工夫の余地はプロンプトエンジニアリングと RAG (Retrieval-Augmented Generation) が主なものとなる。この中のプロンプトエンジニアリングについては昨年度の取り組みにて把握済みであることから、RAG について検討を残すのみとなる。

ここで、クラウド型 LLM は、その利用時の入力を学習しない契約もあるが、個人のプライバシーや業務内容を含むデータの取り扱いが難しいといった問題が存在する。また、従量制の課金であるため、プロンプトやパラメータなどを確認しながら、トライアンドエラーで学習をすることには注意が必要となる。

以上のことからローカルの計算機環境下での LLM（以降ローカル LLM）を教材候補として選定し、その最終的な目標をローカル LLM に RAG を適用したシステム構築とする。現状のローカル LLM の社会での利用例としても適当といえる。

## 2-2. ローカル計算機環境下での生成 AI

画像や動画を生成する AI に関しても教材の検討が必要である。LLM 選定と同様にクラウド型のものは非常に性能が高いが、トライアンドエラーに対する利用料金などを考えても授業での利用は難しい。

したがって、ローカル計算機環境で動作する生成 AI を採用することになる。しかしながら、その生成対象の幅広さ、利用方法の多様さ、種類の多さから、そもそも選定が難しい。さらに、それぞれが様々なバージョンの環境を利用することから、どの動作環境を整えること自体がむずかしい。

以上から、ローカル計算機環境で動作する生成 AI を選んで動作させることのできるソフトウェアの紹介と、それを利用して 2～3 の生成 AI を利用する方法を紹介し、その他の生成 AI については興味を持った

ものについて各自が利用してみる形態を教材候補とした。

## 3. 教材 1: ローカル LLM と RAG

ローカル計算機環境下で LLM を動かす様々な手法があるが、LM Studio[1] や Ollama[2] などのソフトウェア環境下で利用するのが容易である。これらのソフトウェアの中から、様々なオープンソースの LLM のモデルをダウンロード、インストール、実行することができる。

世界中の人々が AI のモデル（とその重みファイル）や学習データセットなどを共有している機械学習プラットフォームの Hugging Face[3] から、LLM のモデルをダウンロードして利用する。しかしながら、モデル（とその重みファイル）は、そのままではサイズが大きすぎて各自が所有するローカル計算機環境上では（RAM や VRAM 不足のため）実行することができない。

そのため、既に学習済みのモデルを量子化（Quantization）の処理をすることによって、動作可能なサイズにまでスケールリングすることができる。量子化は、Ollama（正確には中で動いている llama.cpp[4]）で処理可能である。また、メジャーなモデルは、有志が適度なサイズに量子化済みの形式に変換して配布されているが、自分のローカル計算機環境での適切な（最大かつ動作可能な）量子化サイズに変更したモデルのほうがパフォーマンスが良い。

また、LLM を動作させる際には、いくつかのパラメータの設定により好ましい出力を得ることが可能となる。Temperature（温度）や Repetition Penalty 他さまざまな値を調整し、その効果の確認する試行錯誤が必要である。

さらに、このようにして動作させることができたローカル LLM を RAG と組み合わせる。RAG は LLM の知識を外部の情報源で補強するための手法であり、より正確で最新の情報に基づいた回答を生成することが可能となる。LLM は、そのモデル作成時に学習させた事項しか内部に知識を持っていない。学習時以降の事柄や、そもそも学習していない事柄については、

回答ができなかったり間違った回答をしったりしてしまう。RAGを用いることによって、LLM 単体では持っていない最新情報や特定の分野の知識を必要な時に外部（ウェブ検索や外部ドキュメント）から取り込んで回答に反映させることができる。それにより回答の根拠を示すことができたり、ハルシネーションを回避することができたりする上に、非常に限定的な知識についてモデルをファインチューニングすること無く回答することができるようになる。

このような内容は、実際の社会で利用されている実例とも近く、自分の身近な要件にも適用することができる。LLM を活用するシステム開発の練習とも捉えることができる教材である。

## 4. 教材 2: さまざまな生成 AI を経験

生成 AI は、画像や動画、音楽や声の生成など様々なモデルが存在している。Pinokio[5] は、これらが動作する環境構築、モデルのダウンロード、インストール、実行環境の提供を提供するソフトウェアである。

Pinokio のソフトウェア上から、様々な利用可能な生成 AI について確認することができ、生成 AI の適用先の広がりを体感することができる。

さらに実際に Pinokio を利用し、Stable Diffusion web UI [6] や ComfyUI[7] などをワンクリックでインストールし、その利用を体験することができる。

一度ローカル計算機環境に生成 AI を入れることが

できれば、利用料金を気にせずにプロンプトの使い方をトライアンドエラーすることができる。また、ある程度自由度のある形でのパラメータを設定しておけば、自分の意図に近い表現がなされるまで何回も生成をさせるような使い方も可能である。

これにより自分の興味のある領域に関して、生成 AI を活用する方法を模索することができる。さらに自分が直接興味を持っていない対象や LLM 以外の生成 AI についても理解を深め、その組み合わせ利用を含めて様々な可能性を検討することにまでつなげることができる。

## 5. まとめ

本取り組みは、様々な生成 AI を実際に利用する過程をととして、その知見を教材開発、教育環境に生かすことを目的とした。クラウド型 LLM に対して API 呼び出し型の利用法は、既に検討済みであった為、今回はローカル計算機環境に生成 AI を導入する形での教材を想定した。学生がこれらの教材に触れることによって生成 AI を「知る」「体験する」し「有効活用する方法」を模索することができるようになることを期待する。

謝辞

本取り組みの一部は、本学部社会情報研究センターの支援によりおこなわれたことを感謝します。

## 参考文献

- [1] 溝 LM Studio, <https://lmstudio.ai> (2015 年 1 月取得)
- [2] Ollama, <https://ollama.com> (2015 年 1 月取得)
- [3] Hugging Face, <https://huggingface.co> (2015 年 1 月取得)
- [4] llama.cpp, <https://github.com/ggerganov/llama.cpp> (2015 年 1 月取得)
- [5] pinokio, <https://pinokio.computer> (2015 年 1 月取得)
- [6] Stable Diffusion web UI, <https://github.com/AUTOMATIC1111/stable-diffusion-webui> (2015 年 1 月取得)
- [7] ComfyUI, <https://www.comfy.org/> (2015 年 1 月取得)