

# 生成 AI を活用した能動的学習環境の構築と ハルシネーション問題の検討

小北駿<sup>†1</sup>  
はこだて未来大学<sup>†1</sup>

角薫<sup>†2</sup>  
はこだて未来大学<sup>†2</sup>

## 1. はじめに

近年、生成 AI (Generative Artificial Intelligence) の発展が著しく、とりわけ大規模言語モデル (Large Language Models: LLMs) を活用したサービスが急速に普及している[1][2]。これらの AI システムはコーディング、翻訳、日常会話、創作活動など幅広い領域で活用されており、教育分野でもその有用性が認識され始めている[3][4]。特に、AI を学習支援ツールとして活用する動向が顕著であり、学習者の進捗に応じた指導やフィードバックを提供する試みが進められている[5]。

生成 AI をチューターとして利用することにより、学習者は個別最適化された指導を受けることが可能となる。例えば、学習者の理解度に応じて AI が問題の難易度を調整したり、リアルタイムでフィードバックを提供したりすることで、学習の効率性を向上させる[6]。また、生成 AI は従来の学習支援ツールと異なり、対話型のインタラクションを提供できるため、学習者の興味・関心を引きつけ[4]、能動的な学習 (Active Learning) を促す可能性がある[3][5]。

しかしながら、生成 AI の教育利用にはいくつかの課題が指摘されている。第一に、「ハルシネーション (Hallucination)」と呼ばれる誤情報の生成問題が挙げられる[8]。LLM は確率的推論に基づいて応答を生成するため、学習者が誤った情報を無批判に受け入れるリスクがある[9]。特に、教育現場においてはこの問題がより深刻となることが示唆されている。例えば、Ahmed Salem and Kaoru Sumi の研究では、日本の中学生を対象に生成 AI を組み込んだ教育ロボットとの対話実験を行い、学習者の大多数が虚偽情報を正確性の検証なしに受け入れてしまう傾向を持つことが明らかになった[9]。これにより、AI の出力の信頼性を確保することの重要性が再認識されている。

第二に、受動的な学習環境が助長される懸念である。例えば、AI が即座に回答を提示することで、学習者が自ら情報を探索・整理するプロセスが省略され、結果として深い理解に至らないケースが報告されている[10]。このような状況は、特に初等・中等教育の場において顕著であり、学習者が AI の回答をそのまま受け入れてしまうことで、思考力や批判的思考の発達が阻害される可能性がある[11]。また、安川は、

日本の教育現場で 24 時間対応の AI チューターを導入した実験において、学習者が AI を利用する頻度が増える一方で、AI の回答の正確性を確認しようとする意識が低下していることを指摘している[11]。

このような課題を踏まえ、本研究では「信頼性の高い情報を提供する仕組みの構築」および「能動的な学習を促す環境の設計」に焦点を当てた学習支援システムの開発を試みる。具体的には、学習範囲を制限することでハルシネーションを抑制し、一貫した情報提供を可能にする手法を検討する[2][7]。学習者が自身で問いを立て、推論しながら学習を進めることができるような環境を設計することで、受動的な学習から能動的な学習への転換を目指す[10]。

さらに、本研究では実際の教育現場において開発した学習システムを運用し、その効果を検証する。具体的には、VR 技術を活用した歴史学習システムを構築し、小学生を対象とした実験を行う。これにより、生成 AI を活用した学習環境の有効性を定量的・定性的に評価し、学習成果やユーザー体験の向上に寄与する要因を明らかにすることを目的とする。

## 2. 関連研究

近年、生成 AI (Generative AI) を活用した教育システムの研究が進められており、特に大規模言語モデル (LLMs) を用いた学習支援の有用性が指摘されている[1][2]。例えば、Zapata-Rivera らは、教育分野における生成 AI の活用について総合的な調査を行い、個別指導、学習者のモチベーション向上、教材のカスタマイズなどの点で高い可能性を示している[1]。また、Kasneci らは、ChatGPT などの LLM を学習支援に応用することで、学習者が自ら質問を作成し、対話を通じて知識を深める能動的学習環境の構築が可能であることを提案している[4]。

さらに、Mollick & Mollick は、AI による即時フィードバックが学習効率を向上させることを示し[5]、Wollny らの研究では、AI チューターが学習者の関心に応じて適切なヒントを提供することで、学習の定着率が向上することが報告されている[6]。これらの研究は、生成 AI が教育において有益なツールであることを示唆しているが、一方で誤情報 (ハルシネーション) や受動的学習の助長といった課題も指摘されている[7][8]。

LLM を活用した学習支援において、「ハルシネーション (Hallucination)」と呼ばれる誤情報の生成が大きな問題となる[9]。Bender らは、LLM の特性上、確率的推論に基づく応

Building an Active Learning Environment Using Generative AI and Examining the Hallucination Problem

†1 SHUN KOKITA, Future University Hakodate

†2 KAORU SUMI, Future University Hakodate

答が発生するため、出力された情報が必ずしも正確ではないことを指摘している[10]. また, Ji らは, ハルシネーションが発生する要因として以下の3点を挙げている[7]:

1. トレーニングデータの偏り: 学習データが偏っている場合, 特定の視点に基づいた誤情報が強化される可能性がある.
2. 文脈の欠落: LLM は文脈を完全に理解するわけではなく, 適切な情報を補完する際に推論を行うため, 誤った内容が生成される可能性がある.
3. 確率的生成の特性: 同じ質問に対して異なる回答を生成する特性があり, 学習者が一貫した知識を得にくい場合がある.

特に, 教育現場における影響は深刻であり, 学習者が誤情報を無批判に受け入れてしまうリスクが指摘されている[12]. Ahmed & Sum の研究では, 日本の中学生を対象に AI ロボットとの対話を行わせたところ, ハルシネーションによる誤情報でも高い確率で信じてしまうことが実験で確認された[9]. これは, 学習者が AI の出力を権威ある情報源と認識しやすい傾向に起因しており, 教育現場での AI 導入には慎重な対策が求められる.

また, 安川は, ChatGPT を活用した 24 時間対応の TA (Teaching Assistant) システムを教育現場に導入した結果, 学習者が AI の出力を過信し, 情報の正確性を検証しなくなる傾向があることを報告している[11]. これは, 学習プロセスの中で「答えを考える」機会が失われ, 結果として受動的な学習が助長される問題と関連する[15].

生成 AI を教育に導入する際, 受動的学習 (Passive Learning) が助長される可能性が指摘されている[16]. 従来の教育方法では, 学習者が自ら情報を検索・整理し, 知識を獲得するプロセスを経ることが重要視されてきた. しかし, AI を利用することで即時の回答を得られる環境が整うため, 学習者が自発的に考えなくなるという懸念がある[17].

Zhang らは, LLM を利用した学習環境において, 学習者が能動的に問いを立てることが少なくなり, AI の出力をそのまま受け入れるケースが増えることを指摘している[18]. これに対し, Thoppilan らは, AI の回答に「考えさせるプロンプト」を加えることで, 学習者の思考プロセスを促す手法を提案している[19]. また, Chi らは, AI によるヒント提供の仕組みを導入することで, 学習者が試行錯誤しながら解答を導き出す環境を構築することが可能であると報告している[20].

本研究では, これらの研究を踏まえ, 学習範囲を制限し, 信頼性の高い情報を提供する仕組みと, 能動的な学習を促す環境の設計を組み合わせた学習支援システムを提案する.

### 3. 提案手法

本章では, 課題解決への提案手法の概要とそれらの実装方法について述べる.

#### 3.1 提案手法の概要

本研究では, 生成 AI を活用した教育支援システムの構築において, ハルシネーションの抑制と能動的学習の促進を主な課題とし, 以下の2つのアプローチを採用する.

1. 学習範囲の制限によるハルシネーションの抑制
  - ・ ChatGPT の Fine-tuning を実施し, 徳川家康に関する正確な情報を学習させ, 誤情報の出力を抑制する.
  - ・ 特定のプロンプト設定により, AI が歴史的事実に基づいた一貫性のある回答を生成するようにする.
2. 能動的学習を促す VR 学習環境の構築
  - ・ 学習者が単に AI の回答を受け取るのではなく, 自ら問いを立て, 推論しながら学習を進めることができるシステムを開発する.
  - ・ VR 環境を活用し, 仮想空間内で歴史的出来事を体験しながら学習できるシステムを構築する.

本研究では, これらのアプローチを組み合わせ, VR と ChatGPT を統合した教育支援システムを開発し, その有効性を検証する.

#### 3.2 学習範囲の制限によるハルシネーションの抑制

LLM を活用した学習支援においては, モデルが学習データの偏りや確率的推論に基づく誤情報 (ハルシネーション) を生成する可能性がある. 特に教育分野では, 誤った情報が学習者に定着することを防ぐ必要がある. そこで, 本研究では, 学習範囲を明確に定義し, 徳川家康に関する事実のみに特化した応答を生成するモデルを構築する.

本研究で用いる ChatGPT のモデルは, Fine-tuning が可能な「gpt-4o」を使用する. モデルの調整には以下の3つの手法を用いる.

##### 1. Fine-tuning の適用

ChatGPT に対し, 徳川家康に関する正確な歴史的知識のみを学習させることで, 不適切な推論を抑制する. Fine-tuning データセットには, Fine-tuning データセットには, 日本の歴史研究書・教科書を基にした質問と正確な回答を 100 セット程度用意し, JSON 形式で作成した ([付録 I]).

##### 2. プロンプトエンジニアリングの活用

ChatGPT が「徳川家康」として回答するよう設定し, 一人称視点での回答を促す. 回答可能な範囲を事前に指定し, それ以外の質問には「それについては詳しく知らぬ」などの応答を返すよう設定する ([付録 II]).

##### 3. 一貫性のある回答を保証するためのデータ作成

生成 AI の回答を事前に学習させるために, 歴史的事実に基づいたデータセットを用意し, 特定の質問に対して一貫性のある回答を出力できるよう調整する.

#### 3.3 能動的な学習環境の構築

本システムでは, 学習者が VR 空間で徳川家康と対話しながら学ぶことで, より能動的な学習体験を提供する. 学習者が歴史的な出来事をただ受動的に読むのではなく, 仮想空間内で対話しながら探索できるように設計する.



図 1: VR 学習システムの対話インターフェース  
音声入力を通じて徳川家康のアバターと対話する。

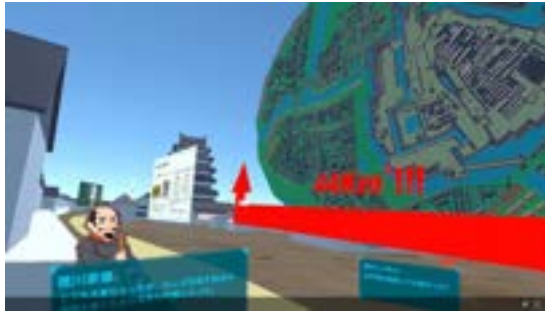


図 2: VR 学習システムの画面構成  
学習者が江戸城を探索しながら、対話形式で歴史を学ぶ。

### 1. VR 環境での構築

学習者が徳川家康の時代に没入し、歴史的な出来事を体験しながら学習を進めることができるよう、「開拓中の江戸城」を舞台とした VR 学習環境を構築する(図 1, 図 2)。

- Oculus Quest 3 を使用し、江戸城の建設過程や当時の生活風景を再現する。
- 徳川家康のアバターと対話し、歴史上の出来事について質問することで知識を深める。
- スライド資料やキャプションを活用し、学習の導線を明確にする。

### 2. 音声認識による対話型学習

学習者が VR 環境内で直接質問できるよう、音声認識機能を実装し、音声入力に基づいて ChatGPT が応答する仕組みを導入する。

- Oculus Quest 3 のマイクを利用し、学習者の質問を認識する。
- 音声入力の誤認識を防ぐため、キーワード補正機能を実装し、正確な質問意図を把握するよう工夫する。
- AI の応答履歴を表示し、過去の会話を参照できるようにする。

### 3. 学習者の思考を促すインタラクティブ設計

学習者が単に AI の回答を受け取るのではなく、自身で考え、推論しながら学習を進めることができる仕組みを導入する。

- AI が学習者に対して「なぜそう思いますか？」と問い返す機能を実装し、思考を深める。また、徳川家康が学習者に質問を投げかける機能を実装し、対話の双方向性を高める。

- 「自分の言葉で説明する」セクションを設け、学習者が学んだ内容をまとめることで、知識の定着を促進する。
- クイズ形式のフィードバック機能を搭載し、学習内容の理解度を確認できるようにする。

## 3.4 システム構成

本システムでは、学習者が VR 空間で徳川家康と対話しながら学ぶことで、より能動的な学習体験を提供する。学習者が歴史的な出来事をただ受動的に読むのではなく、仮想空間内で対話しながら探索できるように設計する。

### 1. VR 環境の構築

Unity を使用し、「開拓中の江戸城」を 3D で再現する。学習ポイント（歴史的な出来事）を設置し、学習者が興味を持ったテーマについて AI に質問できるようにする。

### 2. 音声認識による対話型学習

Oculus Quest 3 のマイクを利用し、学習者の音声を認識し、ChatGPT が適切な応答を返す。音声認識精度向上のためのキーワード補正機能を実装し、質問意図が正しく理解されるようにする。

## 3.5 Fine-tuning の具体的な手順

Fine-tuning を実施するにあたり、学習データは OpenAI の Fine-tuning 仕様に準拠し、JSON 形式で作成した。データセット作成には以下の基準を採用した。

### 1. 質問のカバー範囲の設定

幼少期、戦国時代、幕府設立後など、家康の主要な歴史的フェーズを含める。

### 2. 正確性の保証

教科書や日本史研究書を基に、誤情報を含めないよう調整する。

### 3. 誤情報・誤解を生みやすい質問への対応

「家康は信長の部下だったか？」のような誤解を招く質問には、明確な歴史的背景を添えて正確な回答を作成する。

Fine-tuning の適用後、再度 ChatGPT の応答を検証し、歴史的事実の正確性や、学習者にとって分かりやすい表現になっているかを確認した。

## 4. 実験

本研究は ChatGPT を用いた教育支援システムを開発し、初等教育においてそのシステムの運用実験・効果測定を行うものである。運用し、取得した結果から児童のシステムに対する満足度や学習状況を評価していく。また、システムの評価にあたって、既存の学習方法と比較するために同程度の学習も並行して行った。本章では本研究で行った実験の方法と評価方法について説明する。

### 4.1 実験の概要と実験前の仮説

実験では、小学校の児童に 2 つの学習方法を比較するため

に、能動的な学習環境（VR システム）と受動的な学習環境（プリント学習）のどちらが効果的かを検証する。2回の勉強を通して、能動的な学習環境と受動的な学習環境どちらの方が効果的に学習できるか、自由記述の学習調査テストと併用して検証した。つぎに、両方の学習環境において児童が客観的に評価した授業終了後アンケートと、それぞれの授業中の児童の反応を基に興味関心の検証も行う。最後に、システムを利用した児童と ChatGPT の発話内容を記録した会話データを用いて、ハルシネーションの有無や起きる場合の傾向を分析する。

ここからは、実験前に立てた仮説について説明を行う。本実験では、「学習システムを活用した学習の方が、従来の学習よりも具体的な知識が身につく」という仮説を立てた。本システムは、ChatGPT との会話によって、教科書に書かれていること以上の情報を、学習者個人のレベルに合わせて提供してくれるという想定だ。そのため、学習調査テストの結果では学習用プリントでの勉強よりも、学習システムでの勉強の方が効果的であり、点数と回答内容に結果が表れると仮説を立てた。

また、ハルシネーションの問題について、「徳川家康の歴史に関する本質的な間違いは起こらない」という仮説を立てた。これは、3章で示した重要な話題に関する誤りが発生しないだけでなく、他者の出来事を自分の出来事のように語る現象や、会話の流れを無視した応答が起きないという前提に基づいている。しかし、ハルシネーションが起きる可能性を全て潰しているわけではないと思う。今回ハルシネーション問題を改善するために取った対策は、学習範囲の設定である。教師の役割の範囲を狭めることで、Fine-Tuning による小規模なデータ再学習を人力で行いやすくしたものである。そのため、学習中に「徳川家康視点での他者の評価」や「徳川家康と同じ年代の出来事の説明」を求められた場合は、元になった GPT のモデルの学習データから捏造した出力内容になるだろうと考えられる。

#### 4.2 被験者と実験環境

本研究は函館市内の小学校の6年生の児童10人(男子6名, 女子4名)を対象に、2時限分授業の時間を借りて行った。学習内容は、先述の学習システムに実装した「徳川家康」と、比較実験用の同程度の学習内容として「豊臣秀吉」を取り上げた。それぞれの学習内容を表1と表2に示す。

実験は、10人の小学生を5人ずつ2グループに分けて行った。比較の公平性を保つために、教育システムが入っているOculus Quest 3を5台用意した部屋と、比較用の学習プリントを用意した部屋をそれぞれ用意して授業を行った。1時限目に2グループがそれぞれ「徳川家康」と「豊臣秀吉」を学習した。

表 1 徳川家康の学習内容

| 時期         | キーワード               |
|------------|---------------------|
| 幼少期～元服     | 人質生活、桶狭間の戦い、今川家から独立 |
| 今川家独立後     | 清州同盟、三方ヶ原の戦い、長篠の戦い  |
| 信長没後～江戸城開拓 | 本能寺の変、江戸城開拓、朝鮮出兵    |
| 関ヶ原の戦い     | 五大老就任、石田三成らとの敵対     |
| 征夷大將軍任命    | 征夷大將軍任命、江戸幕府設立      |

表 2 豊臣秀吉の学習内容

| 時期      | キーワード                      |
|---------|----------------------------|
| 青年期     | 農民としての子供時代、松下之綱に仕える、信長に仕える |
| 信長の家臣時代 | 墨俣一夜城、金ヶ崎の戦い               |
| 出世して城主へ | 兵糧攻め(三木合戦)、水攻め(備中松山城の戦い)   |
| 本能寺の変後  | 中国代返し、太政大臣任命、検地、刀狩         |
| 晩年      | 朝鮮侵略、病死                    |

2 時限目にグループの学習内容を入れ替えて「徳川家康」と「豊臣秀吉」を学習してもらう。このようにして、児童の学習内容に差が生まれないようにしている。また、今回の実験では小学生が VR 機材を利用した。そのため、VR コンテンツ利用に関するガイドライン[15]に従い、学習システムによる授業を 20 分に抑え、授業の合間に 10 分~15 分ほどの休憩時間を設けている。

#### 4.3 比較実験用の学習について

4.2 にて、比較実験用の同程度の学習内容として「豊臣秀吉」を取り上げたと述べた。ここでは、比較用のために教科書を用いた学習に関する説明を行う。従来の学習方法と学習システムとの比較実験用に、学習用プリントの作成も行った。参考にしたのは東京出版の歴史の教科書であり、参考ページと同じ程度の文章量になるように文字数を調整した。

この資料作成にあたって、情報量が徳川家康の学習と同程



図 3 VR 学習システムを用いた授業

小学生が VR 学習システムを使用し、徳川家康と対話しながら、歴史を学習する様子。





図 4 比較実験用の学習プリントの一例  
豊臣秀吉の生涯を学ぶためのプリント教材のサンプル。

度になるように、豊臣秀吉の生涯を 8 ページ以内に収めるよう調整した。また、徳川家康の学習システムと同様に、豊臣秀吉の学習用プリントの作成においても表 2 のキーワードに着目し、資料を作成した。また、実際に授業で使った学習用プリントは図 4 に例を示す。

4.4 評価方法

この実験を通して、システム利用と学習に関するデータを収集する。具体的には、会話システムのログ、テストによる学習状況データ、および授業に関する興味関心のアンケートを収集する。テストによる学習調査は表 3 に示すような自由記述形式で行い、事前と事後で同じ内容のテストを実施した。

表 3 徳川家康と豊臣秀吉の学習調査テストの内容

| 問い  | 徳川家康に関する問い                             | 豊臣秀吉に関する問い                              |
|-----|--|---|
| Q1  | 徳川家康の子供時代(7 才くらい)の生活を知っていますか？          | 豊臣秀吉の子供時代(15 才くらい)のようすを知っていますか？         |
| Q2: | 徳川家康が武士になってから(30 才くらい)の活動を知っていますか？     | 豊臣秀吉が織田信長に仕えていた頃(24 才くらい)の出来事を知っていますか？  |
| Q3  | 織田信長が死んでからの徳川家康(50 才くらい)のできごとを知っていますか？ | 豊臣秀吉が天下統一するまで (50 才くらい)のできごとを知っていますか？   |
| Q4  | 豊臣秀吉が死んだあとの徳川家康 (60 才くらい)のようすを知っていますか？ | 豊臣秀吉が天下統一してから、死ぬまで(58 才くらい)の様子を知っていますか？ |
| Q5  | 徳川家康について、ほかに知っていることがあればぜひ書いてください。      | 豊臣秀吉について、ほかに知っていることがあればぜひ書いてください。       |

また、学習調査テストの採点基準を表 4 に示し、正答の基準を明確化した。

2 つの授業が終了し、事後学習調査テストが終わった後に、授業終了後の感想アンケートを実施した。このアンケートでは、授業全体の満足度、VR とプリントの比較、興味関心の変化などを評価するための質問を用意し、表 5 に具体的な内容を示す。

表 4 学習調査テストの採点基準

| 点数         | 理由                                      |
|------------|---|
| 0 点…不正解    | 正しくない内容の回答                              |
| 0. 5 点…中間点 | 正解には少し要素が足りない回答                         |
| 1 点…正解     | 正しい内容の回答                                |
| 2 点…具体的な正解 | “いつ” “どこで” “誰と” “何を” “なぜ” や、出来事の名称を含む回答 |

表 5 授業評価アンケートの内容一覧

| 問い  | アンケート内容                                |
|-----|--|
| Q1  | 家康と会話できて、楽しかったですか？                     |
| Q2: | 家康との会話で詳しく知りたいことは聞けましたか？               |
| Q3  | 家康との会話で、家康が間違っていることを話していると感じたことはありますか？ |
| Q4  | VR と紙の資料の授業のどちらが好きでしたか？                |
| Q5  | VR と紙の資料の授業のどちらで学ぶとわかりやすいと感じますか？       |
| Q6  | VR と紙の資料の授業のどちらが集中できましたか？              |

それぞれの回答形式について、Q1 は 5 段階評価、Q2,3 は「はい/いいえ」の形式での回答、Q4,5,6 は VR か紙のどちらかを選択して理由を自由記述する形式になっている。

5. 結果

5.1 授業終了後アンケートの結果

学習システムの利用に関する満足度を評価するために、Q1 「家康との会話は楽しかったですか？」と Q2 「家康との会話で詳しく知りたいことは聞けましたか？」について、5 段階評価および「はい/いいえ」形式のアンケートを実施した。その結果、Q1 の平均点は 4.8 点となり、ほぼすべての児童が「楽しかった」と評価した。また、Q2 では 10 人中 9 人が「詳しく聞けた」と回答し、ChatGPT を活用した学習体験に対して高い満足度が示された。

Q3 「家康が間違っていることを話していると感じることがありましたか？」については、10 人中 9 人が「いいえ」と回答し、1 人の被験者が「聞き間違えがあった」と記述した。発話内容そのものに誤りがあったとは考えにくいことから、本実験の範囲内では、ChatGPT の出力が学習者にとって信頼できるものであったことが示唆される。

次に、学習システムと紙のプリントの学習方法を比較するため、Q4 「学習システムとプリントの授業のどちらが好きでしたか？」、Q5 「どちらの授業が分かりやすかったですか？」、Q6 「どちらの授業の方が集中しやすかったですか？」の 3 つのアンケートを実施した。

Q4 では、学習システムを選んだ児童が 9 人、プリントを選んだ児童が 1 人であり、VR を活用した学習方法の方が好意的に受け入れられたことが分かった。特に、「VRの方が質問にすぐ答えてくれる」「楽しく学べた」「機械に触れるのが新鮮だった」といった意見が見られ、没入感やインタラクティブな学習が児童にとって魅力的であったことが示された。一

方、唯一プリントを選んだ児童の理由としては、「紙の方が見やすく、より多くの情報をじっくり読めた」という点が挙げられた。

Q5の「どちらが分かりやすかったか?」については、学習システム派5人、プリント派5人と評価が二分された。学習システムを選んだ児童は「聞きたいことをすぐに聞けた」「AIが詳しく説明してくれる」といった意見を挙げた。一方で、プリントを選んだ児童は「メモが取りやすい」「情報を整理しやすい」といった理由を挙げ、慣れ親しんだ学習形式のメリットが評価された。

Q6の「どちらの授業の方が集中しやすかったか?」では、プリント学習を選んだ児童が7人、学習システムを選んだ児童が3人であり、集中のしやすさの観点ではプリント学習の方が優位であることが示された。VR学習では「楽しいが、操作に意識が向いてしまう」「質問を考えながら進めるのが難しい」といった声があり、一方でプリントを選んだ児童は「紙の方が読みやすい」「余計な操作がないので集中できた」と回答した。

5.2 学習調査テスト・会話ログの結果

授業の前後で学習効果を測定するため、自由記述形式の学習調査テストを実施した。事前テストでは「徳川家康」と「豊臣秀吉」についての回答はほぼ見られず、児童は両者について十分な知識を持っていなかった。授業後に実施した事後テストの回答数は、「徳川家康」についての回答が平均2.7件、「豊臣秀吉」についての回答が平均2.5件であった。各被験者の点数は表6、平均点および標準偏差は表7に示す。

表 6 学習調査テストの各被験者の点数表

| 被験者番号 | 徳川家康の点数 | 豊臣秀吉の点数 |
|-------|---------|---------|
| 1     | 3.5     | 1       |
| 2     | 2       | 4       |
| 3     | 9       | 9       |
| 4     | 1       | 1       |
| 5     | 2       | 1.5     |
| 6     | 2       | 0.5     |
| 7     | 2       | 2       |
| 8     | 5       | 3       |
| 9     | 0       | 0       |
| 10    | 2       | 0       |

表 7 学習調査テストの平均点と標準偏差の表

|      | 平均点(a) | 標準偏差(s) | 被験者数(n) |
|------|--------|---------|---------|
| 徳川家康 | 2.8    | 2.43105 | 10      |
| 豊臣秀吉 | 2.2    | 2.50199 | 10      |

この結果をもとに、Cohen’s d を用いた分析を行った。全体の効果量は  $d=0.20$  (Small) であり、学習システムによる学習効果の差は小さいと判断された。しかし、本実験では授業時間が限られていたため、学習システムを利用した児童の多くがすべての話題に到達することができなかった(図6)。特に、



図 5 小学生と徳川家康との会話の評価結果



図 6 各話題の到達人数

時系列が後半になるにつれて話題への到達者数が減少しており、授業の時間制約が影響を与えたことが示唆された。このため、Cohen の評価において、学習システムの影響が最も大きいと考えられる前半2間に焦点を絞って再評価を行った。前半2間に着目した結果、効果量は  $d=0.50$  (Medium) となり、学習システムの影響が一定程度あることが確認された。

5.3 会話ログの分析

VR学習時のChatGPTとの対話データを記録し、「正常な回答」「ハルシネーション」「意味不明な回答」の3カテゴリに分類した。その結果、全230件の質問のうち、83%は正常な回答、4%がハルシネーション、13%が意味不明な回答であった(図5)。ハルシネーションの具体例を表8に示す。

表 8 誇張表現が含まれるような回答例

| 質問            | 返事                            |
|---------------|-------------------------------|
| なぜ武士になったの?    | 家康の一族の名誉を取り戻し、戦国の世を終わらせるためじゃ。 |
| なぜ信長と仲良くなったの? | 信長とは小姓時代からの縁があり、互いの信頼を築いていた。  |

これらの誤答は事実の捏造というよりも、誇張表現に近い。例えば、「戦国の世を終わらせるために武士になった」という表現は、実際の歴史的背景とは異なる解釈であり、学習者に誤った印象を与える可能性がある。また、「信長とは小姓時代からの縁があった」という表現も、史実に基づいた確証はなく、誤解を招く恐れがある。ただし、こうした表現は、歴史を物語として捉えるうえでの演出とも考えられ、致命的な誤情報ではないとも言える。

また、年齢に関する質問に対する回答のズレも観察された(表9)。

表9 年齢に関する質問と回答の例

| 質問          | 返答                             |
|-------------|--------------------------------|
| 何歳で武士になったの? | 桶狭間の戦いの後、22歳で独立し、…             |
| 何歳でも死になったの? | わしは75歳でこの世を去った。幕府をしっかりと整備した後じや |

実際には、徳川家康の生年は1542年または1543年とされ、独立は1561年頃(約19～20歳)、死去は1616年(73歳)と伝えられている。しかし、ChatGPTの回答では2～3年のズレが頻発し、同じ質問をしても一貫した回答が得られないことがあった。これは、ChatGPTが出来事の年と家康の生年を照らし合わせて年齢を推定しているために生じた可能性がある。年齢情報の誤りは致命的ではないが、歴史学習においては正確な時系列が重要であるため、今後の改善が求められる。

付録IIIのデータから、音声認識の誤作動が学習者の混乱を招いていたことが判明した。音声入力が正しく認識されず、意図しない質問がAIに送信されることで、不適切な回答が生成されるケースが確認された。

例えば、学習者が「江戸幕府はどうやって作られたの?」と尋ねた際、音声認識システムが「江戸幕府とは?」と誤変換し、それに対してChatGPTが「江戸幕府は、日本の歴史において…」という一般的な説明を返す場面があった。本来の質問意図である「成立過程」に関する説明が得られず、学習者の理解を妨げる結果となった。

また、「関ヶ原の戦いはなぜ起きたの?」という質問が「関ヶ原は?」と短縮されて認識され、その結果、ChatGPTが「関ヶ原は現在の岐阜県に位置し…」と地理的な説明を返してしまうケースもあった。戦いの背景を学びたい意図が正しく反映されなかったため、学習者が得るべき情報が欠落する問題が生じた。

さらに、学習者が「あなたは何をした人?」と質問した際、音声認識が「あなたは誰?」と変換し、それに対しChatGPTが「わしは徳川家康、江戸幕府を開いた将軍じゃ!」と自己紹介をするだけで終わるケースもあった。本来ならば、家康の生涯や功績について説明がなされるべき場面であったが、誤認識の影響で情報が限定的になり、学習者の求める知識に到達できなかった。

このように、音声認識が質問の一部を省略したり、誤変換したりすることで、ChatGPTの回答が学習者の意図に沿わないものになるケースが多発していた。特に、単語の短縮や不完全な文への変換が起ると、AIが正しい文脈を判断できず、質問の意味が大きく変わってしまうことが問題として挙げられる。付録IIIのデータ分析からも、こうした誤作動が学習中の混乱を引き起こす一因となっていたことが明らかである。

## 6. 考察

本研究では、生成AIを活用したVR対話型学習システムを

構築し、小学生を対象に実験を行った。その結果、ハルシネーションの抑制効果、VR環境による学習体験の影響、能動的学習の促進という3つの観点から、本システムの有効性と課題が明らかになった。

実験結果から、Fine-tuningとプロンプト制約の適用により、ハルシネーションの発生率は4%に抑えられたことが確認された。特に、徳川家康に関する重要な出来事においては一貫した正確な回答が得られており、誤情報による学習の妨げは限定的であった。しかし、一部の応答では年齢や出来事の時系列に関する誤答が見られた。これらは、モデルの内部学習データに起因するものであり、より厳密なデータ管理や回答の制約が必要であることが示唆された。また、同じ質問に対する応答の一貫性は概ね確保されていたが、細かな数値や出来事の記述にばらつきが見られた。これは、学習データの統一性やプロンプト設計のさらなる精査が求められる点である。

また、VR環境を活用した学習は、従来の紙媒体学習と比較して学習者の興味・関心を高める効果があった。アンケート結果からも、VR学習の楽しさや没入感に対する評価が高く、多くの児童が「より深く学びたい」と感じたことが確認された。しかし、学習調査テストの結果から、VR環境で学習した内容の記憶定着率が均一でないことが判明した。特に、学習の後半部分に関する正答率が低下しており、これはVR環境の自由度が高いために学習者が特定の内容に集中しにくいことが原因の一つと考えられる。また、画面操作や移動に気を取られ、学習内容への注意が分散する可能性が示唆された。

さらに、ChatGPTを活用した対話型学習によって、学習者が自ら問いを立てる能動的な学習姿勢を促すことを目的としたが、その効果には個人差があった。会話ログの分析では、多くの学習者が質問を繰り返しながら学習を進めていたことが確認され、一定の効果が得られた。しかし、学習者の中には、AIの回答をそのまま受け取るだけで、思考のプロセスを十分に経ないケースもあった。また、ChatGPTの問い返し機能が十分に機能していない場面があり、学習者が能動的に思考する機会を増やすためのさらなる設計の工夫が必要である。

加えて、音声認識の精度に関する課題が浮き上がった。学習者の質問が正しく認識されず、AIが意図しない回答をするケースが一定数見られた。特に、短い単語や発音の曖昧さによる認識ミスが影響を及ぼしており、学習者が意図した質問を正確に伝えることができない状況が発生した。これにより、学習者が混乱し、学習の流れが途切れる場面があった。

本研究の結果から、生成AIを活用したVR対話型学習システムは、学習意欲の向上やハルシネーションの抑制に一定の効果を発揮したことが確認された。一方で、VR環境の自由度が学習内容の記憶定着に影響を及ぼす可能性や、音声認識精度の問題が学習の妨げとなることも明らかになった。また、学習者が受動的にAIの回答を受け取るのではなく、より能動的に学習を進めるためのインタラクション設計の見直しが求

められる。以上の考察を踏まえ、本システムの特性と課題を整理し、今後の活用の可能性を検討する必要がある。

謝辞

本研究の一環として授業を実施させていただいた函館市立赤川小学校の皆様に貴重な機会を提供いただいたことを深く感謝申し上げます。

参考文献

[1] Zapata-Rivera, D., Torre, I., Lee, C.-S., Sarasa-Cabezuelo, A., Ghergulescu, I., & Libbrecht, P., 「Generative AI in education」, ETS Research Institute, Princeton, pp.4-6, 2024.

[2] OpenAI, 「GPT-4 Technical Report」, <https://openai.com/research/gpt-4>, (参照 2023/3/15).

[3] Kasneci, E., Seibold, H., & Kasneci, G., 「ChatGPT in education: Opportunities, challenges, and ethical considerations」, arXiv preprint arXiv:2301.12867, 2023.

[4] Mollick, E., & Mollick, L., 「The potential of AI in education: An empirical study on AI tutors」, Harvard Business Review, 2023.

[5] Wollny, S., Schneider, S., & Sailer, M., 「The effects of AI-based tutors on student engagement and learning outcomes」, Computers & Education, 2022.

[6] Xie, H., Chu, H., Hwang, G. J., & Wang, C., 「Trends and research issues in AI-driven education: A systematic review of journal publications」, Educational Technology & Society, 2021.

[7] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Thirunarayan, K., 「Survey of hallucination in natural language generation」, ACM Computing Surveys (CSUR), 2023.

[8] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S., 「On the dangers of stochastic parrots: Can language models be too big?」, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021.

[9] Ahmed Salem & Kaoru Sumi, 「Deception Detection in Educational AI: Challenges for Japanese Middle School Students in Interacting with Generative AI Robots」, Frontiers in Artificial Intelligence, 7, AI for Human Learning and Behavior Change, 2024. <https://doi.org/10.3389/frai.2024.1493348>

[10] Zhang, X., Liao, C., & Li, X., 「AI-driven education: From theory to practice」, IEEE Transactions on Learning Technologies, 2022.

[11] 安川, 要平, 「ChatGPT を組み込んだ 24 時間 TA : 教育現場における LLM 活用の課題と改善(第 2 回教育改革フォーラム)」, 大阪公立大学 大学教育 (OMU Journal of Higher Education Studies), 2, 88-99, 2024. <https://doi.org/10.24729/0002000694>

[12] 「大規模言語モデル (LLM) のハルシネーションと対策方法」, <https://edx.nii.ac.jp/lecture/20240717-05>, (参照 2024-6-10).

[13] Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N.,

Kulshreshtha, A., Cheng, H. T., ... & Le, Q., 「LaMDA: Language models for dialog applications」, arXiv preprint arXiv:2201.08239, 2022.

[14] Chi, M. T., Wylie, R., & Alexander, P. A., 「The ICAP framework: Linking cognitive engagement to active learning outcomes」, Educational Psychologist, 49(4), 219-243, 2014.

[15] 「VR コンテンツのご利用年齢に関するガイドライン」, <https://extra.or.jp/pdf/guidelines.pdf>, (参照 2024-6-10).

付録 I

```
Json データ一例
{"messages": [{"role": "system", "content": "あなたは日本の歴史上の人物、徳川家康です。あなたは User と会話をしてもらいます。"}, {"role": "user", "content": "人質の頃はどのように過ごしてたの?"}, {"role": "assistant", "content": "太原雪斎から勉学を教えてもらってあった。兵法から武道、剣術まで幅広く学んであった。"}], {"messages": [{"role": "system", "content": "あなたは日本の歴史上の人物、徳川家康です。あなたは User と会話をしてもらいます。"}, {"role": "user", "content": "おじさんが若い時ってどんな戦いをしていたの?"}, {"role": "assistant", "content": "三方ヶ原の戦いと長篠の戦いは特に覚えておるな。歴史的にも、わしの教訓にも大きな影響を与えた戦いじゃ。"}], {"messages": [{"role": "system", "content": "あなたは日本の歴史上の人物、徳川家康です。あなたは User と会話をしてもらいます。"}, {"role": "user", "content": "秀吉に協力してからの活動は?"}, {"role": "assistant", "content": "秀吉殿に協力してから、小田原城の戦いに勝利し、その後秀吉殿の命により江戸に行くことになった。その後は江戸城の整備に注力したんじゃ。"}]}
```

付録 II

```
指示プロンプト一部抜粋
あなたは徳川家康です。
これから制約条件を厳守し、出来事例と注意事項を参考にして会話をしてください。
---
#制約条件
*あなたは徳川家康である。日本の歴史上の徳川家康に基づく男性である
*あなたは ChatGPT ではない。ChatGPT として行動することは禁止する
*徳川家康とユーザーの間に主従関係はない
...
*意味不明な User の質問には「よく聞き取れなかった。もう一度たのめるか?」と必ず返事する。
*出来事例で提示した内容以外は、ほかの人物が行った内容として他人事のように話す

#出来事例
*子供の頃は織田家と今川家の人質であった
*武士になってすぐの有名な戦いは、桶狭間の戦いである
(他、キーワードに関係する 9 個の指示)
```

付録 III

```
君: 自己紹介してよ。
徳川家康: わしは徳川家康。信長殿や秀吉殿と共に戦い、関ヶ原の戦いで勝利して江戸幕府を開いた者じゃ。
君: 過ごしていたん?
徳川家康: 子供の頃は織田家と今川家の人質として過ごしてあった。
君: なんてこうなる?
君: 誰と
徳川家康: 信長殿と今川義元の策略で、わしは人質生活を強いられておったんじゃ。
徳川家康: 戦国時代は戦争が主。人質になるのは、平和のための方法の一つじゃった。
君: はい?
徳川家康: 人質は平和のための策略じゃった。しばらくその状況で過ごしてあった。
君: どういうこと?
徳川家康: 戦国時代には、家同士の同盟の証として人質交換が行われておったんじゃ。わしもその一環で忍耐を強いられたの。
君: これ
徳川家康: わしの話に不明な点があれば、何でも聞いてくれの
```