ピクチャーハンターⅡ：マルチモーダル LLM によるモンスター生成を用いたリアルタイム参加型ゲーム

中村 裕大[1,a)]　巳波 弘佳[1,b)]　平野砂峰旅[2,c)]　片寄 晴弘[1,d)]

概要：デジタル世界と現実世界をリンクさせることで，コンテンツの「面白さ」を強化する方法がある．デジタル世界のみで構成されたゲームは，プレイヤにとって箱庭的な体験になりがちであるが，現実世界をリンクさせることで，プレイヤの没入感やイリュージョン感覚を高めることができる．本研究では，そのコンセプトを基盤として，新たなゲーム体験を提供する『ピクチャーハンター Ⅱ』を提案する．本コンテンツは，現実世界のオブジェクトをデジタル世界のモンスターにマッピングする際に，プレイヤの意図や工夫が反映される「納得感」と，時には「意外な展開」を生じさせることを目指す．その実現に向けて，「マルチモーダル LLM（Large Language Model）」を活用した手法を開発した．「マルチモーダル LLM」は簡単なプロンプトでプロフェッショナルレベルの絵を生成する技術であるが，必ずしも意図通りの結果が得られるわけではない．発表では，2024 年時点での「マルチモーダル LLM」の長所と短所を「楽しみ」のベースとした多人数参加型のデモンストレーションを行う．

1. はじめに

デジタルエンタテインメントコンテンツにおける「面白さ」の要素として，デジタル世界と現実世界のリンクが挙げられる．デジタル世界は創造された世界であり，そのメリットとして自由な構成が可能である反面，プレイヤにとっての体験は箱庭的なものになりがちである．この問題に対し，現実世界とリンクしたコンテンツは，没入感を高めるだけでなく，現実の環境や出来事が投影されるイリュージョン感覚も楽しめる．

そのようなコンテンツの先駆けとなった重要作品に『バーコードバトラー』（1991）がある [1]．バーコードバトラーは，現実世界のバーコードを読み取り，デジタル世界のモンスターに変換して戦うバトルゲームである．発売当時，人気のあまり店の商品からバーコードが切り取られる社会現象まで引き起こした．このゲームの成功の眶は，「現実世界での行動がデジタル世界と連動している」点にあり，ゲームプレイが単なる仮想の体験ではなく，現実の行動が反映される動的な体験となり，より深い没入感と満足感が提供される．このコンセプトは，『Pokémon GO』[2] や『Ingress』[3] などの AR（拡張現実）コンテンツにも通じている．

典型的な AR ゲームは現実世界をフィールドに見立て，プレイヤが実際に動き回ることで，デジタル世界とのリンクを体験できる．しかし，多くのこれらのコンテンツでは，デジタル世界と現実世界の関連性があらかじめ設定されており，プレイヤは決められたシナリオに沿って行動させられることが多い．一方，バーコードバトラーのようなプレイヤオリエンテッド[*1]なゲームコンテンツは，プレイヤ自身の選択がダイレクトにゲームに反映される．これにより，現実世界とデジタル世界のリンクを通じて，より自由で個別化された体験を提供する．

筆者らはこのコンセプトにヒントを得て，マルチモーダル LLM のプロンプトエンジニアリングの枠組みより，プレイヤが用意した写真からデジタル世界のステータスへとマッピングする『ピクチャーハンター』（2024）を開発した [4]．本稿では，前作のピクチャーハンターの EDA（Entertainment Design Asset）[5]，システムデザイン，特に，「適度な納得感」のあるステータスマッピングの方法について述べ，さらに，多くのプレイヤがリアルタイムに参加できるデモンストレーションの視点からゲームデザインを改善した『**ピクチャーハンター Ⅱ**』について述べる．

[1] 関西学院大学
Kwansei Gakuin University, Sanda, Hyogo 669–1330, Japan
[2] 京都精華大学
Kyoto Seika University, Kyoto City, Kyoto 606–0016, Japan
[a)] gyw82319@kwansei.ac.jp
[b)] miwa@kwansei.ac.jp
[c)] hirano@kyoto-seika.ac.jp
[d)] katayose@kwansei.ac.jp

[*1] プレイヤの行動や選択が主体となる

2. 現実世界とデジタル世界のリンク

AR ゲームをはじめとした「現実世界とデジタル世界をリンクさせる」コンテンツは，プレイヤに新しい体験を提供する技術として発展してきた．初期の AR ゲームである『Ingress』[3] は，プレイヤの位置情報を活用し，**現実世界での移動を通じてデジタル世界と融合する**スタイルを確立した．その後の『Pokémon GO』[2] では，デジタル世界上の存在である**モンスターと現実世界の融合**が実現され，多くのプレイヤを獲得し，高いナラティブ性を実現した．これらのゲームは，現実世界にデジタル世界が投影されるという共通点を持ち，プレイヤのゲームプレイに対する没入感を高める重要な要素となっている．しかし，これらのコンテンツでは，現実世界との関連性があらかじめデジタル世界のシナリオによって決定されており，プレイヤの選択の幅が制限されていた．

『バーコードバトラー』や『モンスターファーム』などのプレイヤオリエンテッドなゲームは，現実世界におけるプレイヤ自身の選択をデジタル世界へと反映できる．バーコードバトラーは，プレイヤ自身が現実世界の商品に付随するバーコードを用意し，その情報を基にデジタル世界のモンスターを生成する独自のシステムを導入した [1]．このシステムは，プレイヤが現実世界の選択を通じてゲームに直接影響を与えることを可能にし，高いインタラクティブ性を提供した．またモンスターファームでは，プレイヤが用意した音楽 CD から，ゲーム内のモンスターが生成される仕組みを取り入れた [6]．この仕組みは，プレイヤが現実世界で収集した音楽 CD を，デジタル世界でモンスターとして育成する楽しみを提供した．結果として，「**現実世界のオブジェクトを読み取ることで，デジタル世界のステータスにプレイヤが関与できる**」コンセプトを実現し，より自由で個別化された体験が提供された．

一方，バーコードバトラーをはじめとした，従来のプレイヤオリエンテッドなゲームでは，現実世界のオブジェクトのプロパティ（そのもの自身が持つ特性）と全く無関係なステータスが生成されていた．その結果，プレイヤの創意や工夫がデジタル世界に反映されず，ゲームに対する没入感や参加意識が低減していた．しかし，完全にプレイヤの意思が反映されても，ゲーム要素として重要な意外性が失われてしまう．よって，現実世界のオブジェクトから生成される結果に対して「適度な納得感」を保ったマッピング手法が必要であった．以下，この問題を解決するためのマルチモーダル LLM を用いたアプローチについて説明する．

3. マルチモーダル LLM を用いたマッピング

近年，ChatGPT などのマルチモーダル LLM は，プレイヤの期待や文脈に沿った「それっぽい」応答を生成する能力がある．マルチモーダル LLM は確率モデルであるが故に，時折ハルシネーションなどの誤った結果が生成される．しかし，筆者らは一見ネガティブなこの特性を活用し，「適度な納得感」を持ったマッピングを実現した．具体的には，プレイヤの意図や工夫を組み込んだ生成に「納得感」を得つつ，時折発生するハルシネーションによって「意外性」を持った結果を取得することができる．

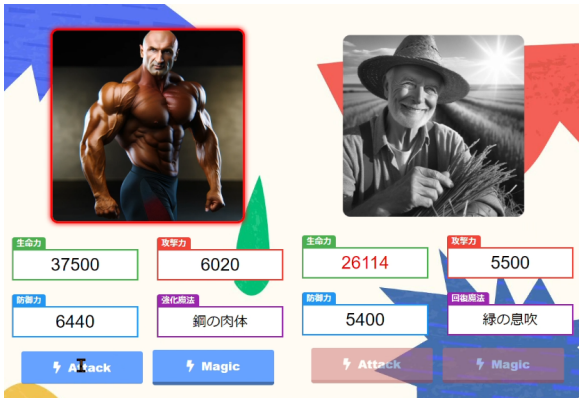

図 1 前作『ピクチャーハンター』のプレイ画像

このアプローチを採用する上で，解決すべき課題として，数値の扱いに関する問題が挙げられる．マルチモーダル LLM は，そのトークナイザーや言語モデルの制約から，数値の扱いに弱いことが知られている [7]．そのため，直接数値ステータス（例えば攻撃力や防御力）を生成すると，時折ゲームバランスを無視した数値ステータスが生成されることがある．

筆者らは，この問題を解決するために，画像キャプショニングと Text-Embedding を用いて数値ステータスを生成するアプローチを提案した．まず，マルチモーダル LLM が得意とする画像キャプショニング（Image2text）の処理を行い，入力画像のキャプションを生成する．このキャプションをプレイヤに提示することで，システムが画像を正しく認識しているという安心感や納得感を高めることができる．次に，生成されたキャプションから Text-Embedding（Text2vec）を用いて意味的な特徴量を保持したベクトル（数値）に変換する．そのようなベクトルに変換することで，「適度な納得感」を保持することができる．加えて，ベクトル同士の類似度を用いることで，ステータスの上限や下限を適切に設定し，マルチモーダル LLM の数値処理の弱点を補うこともできる．

これにより，プレイヤは予測可能性[*2]と意外性のバランスが取れたマッピングを通じて，より深い没入感と満足感を得ることができる．

[*2] 現実世界のオブジェクトを入力する前に，どのようなデジタル世界のステータスが出力されるか予測可能であるという意

4. 前作と本作のゲームデザイン

4.1 前作のピクチャーハンター

現実世界でのプレイヤの選択をデジタル世界へと反映する要素が，数多くのゲームコンテンツの「面白さ」を生み出している．そのようなコンテンツでは「適度な納得感」を持つマッピング手法が重要であった．前作の「ピクチャーハンター」（図 1）ではプレイヤの意図を汲み取る能力を持つマルチモーダル LLM を活用することで，「適度な納得感」のあるマッピングを実現した．しかし，ピクチャーハンターは，1 対 1 のターンベース型対戦ゲームであるため，一度の参加人数に制限があり，進行スピードが遅いという問題があった．

4.2 本作のピクチャーハンター Ⅱ

ゲームコンテンツにおけるデモンストレーションはコンテンツの理解や魅力の伝達のために重要である．したがって本研究では，ピクチャーハンターのゲームデザインをデモンストレーションベースで改善する．

デモンストレーションの効果を最大限に引き出すためには，1 度のデモンストレーションで可能な限り多くの参加者が体験できることが望ましい．デモンストレーションの参加者は，各自興味のあるブースを無作為に訪れるため，訪問時に既にゲームが進行していた場合でも参加できるようなゲーム構成が必要である．そこでピクチャーハンター Ⅱ ではこの問題を解決するために，リアルタイムストラテジー（RTS）型の進行形式を採用する．

また，デモンストレーションにおいてゲームの初期学習コストを低減することも重要である．操作やゲーム構成が複雑であると，参加者がゲームに参加することを躊躇う可能性がある．そのため，ピクチャーハンター Ⅱ では，基本コンセプトである「写真からモンスターを生成する」こと以外，プレイヤの関与を極力減らすことにする．具体的には，モンスター生成後はゲームシステムに一任し，自動的に進行するゲームデザインを目指す．

このようなデザインの参考作品として「にゃんこ大戦争」[8] などのタワーディフェンス型ゲームが挙げられる．タワーディフェンス型ゲームは基本的に 1 対 1，もしくはプレイヤ対コンピューターの対戦を想定した RTS ゲームであるが，いずれも群衆 vs 群衆の戦闘が特徴である．本作ではこれを複数プレイヤ vs 複数プレイヤに置き換えることで，多人数が参加できるゲームデザインを実現する．

これらのゲームデザインを基に，待ち時間なくリアルタイムに多人数が参加できる環境を構築する．詳細なゲームルールに関しては，次項の実装章で示す．

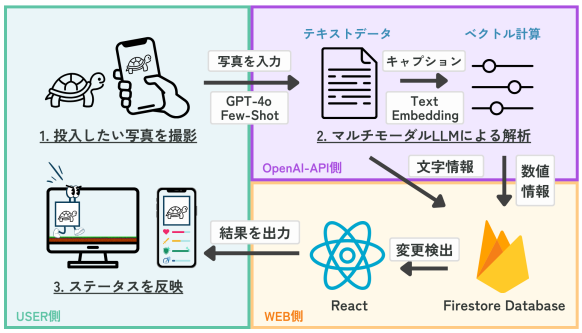

図 2 ピクチャーハンター Ⅱ におけるシステムの仕組み

5. ピクチャーハンター Ⅱ の実装

5.1 実装の概要

ピクチャーハンター Ⅱ は複数プレイヤ vs 複数プレイヤ形式を想定しており，プレイヤから提供された写真を逐次，入力として受け取る．システムは受け取った写真情報をプロンプトテンプレートに埋め込み，OpenAI の API である「GPT-4o」[9] に送信する．

GPT-4o からレスポンス（テキストデータ）を得ると，写真に関するキャプションと，それ以外の文字情報（例えば，モンスターの名前や必殺技の名前など）に分割する．その際，テキストデータをレスポンス上で区別できるように，特定の文字列を間に挟む規則をプロンプトエンジニアリングにより課しており，出力は「Few-Shot Learning」[10] を用いることで安定化させている．

次に，キャプションを OpenAI の埋め込みモデル「text-embedding-3」[11] を使用して解析し，高次元の埋め込み空間内における意味的な特徴量を持ったベクトルを得る．得られたベクトルから類似度を用いてゲーム内での数値ステータスを算出する．数値ステータスはにゃんこ大戦争のゲームデザインに基づき，体力，速度，攻撃力，DPS（Damage Per Second）などのゲーム進行に必要なものを生成する．これらの値はすべて同じ値域を取るが，体力と攻撃力を同じスケールで扱うと，1 度の攻撃で決着がつく可能性があるので，体力を定数倍して，ゲームバランスを調整する．

得られた文字情報と数値情報を Firestore Database に保存し，データベースの更新毎に React による再レンダリングを行う．これにより，プレイヤの入力がリアルタイムで反映される仕組みを構築する．

以上のシステム構成についてまとめたものを図 2 に示す．

5.2 ゲームルールの概要

ピクチャーハンター Ⅱ では，参加者に所縁のあるチーム分けを採用する．例えば，「関西 vs 関東」や「きのこ vs たけのこ」などである．この方法は，参加者が自分に関連する地域や属性を自ら選択することで，チームに対する愛

図 3　ピクチャーハンター Ⅱ の概要を示すパンフレット画像

着が生まれやすくなるという心理的効果を期待している．また，N 対 N とは限らないため，チームメンバーが偏った際は，公平性を保つために NPC（Non Player Character）を用いてバランスをとる．

チームに参加すると，プレイヤは自身のチームメンバーと協力し，戦線を拡大する．プレイヤにより投入されたモンスターは，自陣から敵陣に向かって進行し，相手のモンスターに接触すると攻撃を始める．相手を倒すことで，進行を再開し，自身の戦線を拡大することができる．

本ゲームは，最低 1 枚の写真からいつでも参加可能である．これにより，参加のハードルを低くし，多くの参加者を誘引する狙いがある．また，参加者はアイテム画像を 1 枚追加することで，モンスターのステータスを向上させることができる．これは画像ならではの重ね掛けの利点を活かした，戦略的なプレイを促進する．

投入したモンスターが倒れてしまった場合，次のモンスターを投入できるまでの「待機時間」が発生する．これは前回投入したモンスターのステータス値に基づいて計算される．この仕組みにより，強力な画像を投入する際のリスクと報酬のバランスをとることができる．

最終的に，占領している領域がより広いチームが勝利となる．これにより，時間制限内での戦略的なプレイが求められ，短時間でゲームが終了しないように設計されている．

以上のゲーム概要についてまとめたものを図 3 に示す．

6. おわりに

本稿では，「プレイヤの現実世界での意図や工夫をデジタル世界へと反映できる」という新たなエンタテインメント性をベースに，現実世界の写真を用いてステータスを生成する「ピクチャーハンター Ⅱ」を提案した．マルチモーダル LLM をベースにしたマッピング手法により，プレイヤが「適度な納得感」を得る生成を実現し，多人数がリアルタイムに参加できるゲームデザインを構築した．

現在は単発的なゲーム構成であるため，今後の展望として，モンスターの育成要素を追加したり，SNS を関連付けた新たなゲーム性の構築や，定期的なイベントによるプレイヤ層の拡大を行いたいと考えている．これにより，ゲーム内での協力や競争が活発化し，継続的なユーザーエンゲージメントが実現することを目指す．

参考文献

[1] 株式会社エポック社：バーコードバトラー，（オンライン），入手先 ⟨https://epoch.jp/⟩ (1991)．(Accessed on 07/25/2024).
[2] Inc, N.: Pokémon GO, (online), available from ⟨https://www.pokemongo.jp/⟩ (2016). (Accessed on 07/25/2024).
[3] Inc, N.: Ingress Prime – Ingress Prime, (online), available from ⟨https://ingress.com/⟩ (2013). (Accessed on 07/25/2024).
[4] 中村裕大, 巳波弘佳, 平野砂峰旅, 片寄晴弘：ピクチャーハンター：キャラクター属性の設定にマルチモーダル LLM を利用した令和版バーコードバトラー，研究報告エンタテインメントコンピューティング (EC)，Vol. 2024-EC-71, No. 25, pp. 1–4 (2024)．発行年：2024-03-10.
[5] 小笠　航，片寄晴弘：自己実現理論を起点とした Entertainment Design Asset の提案とその分析事例報告，研究報告エンタテインメントコンピューティング (EC)，Vol. 2017, No. 1, pp. 1–8 (2017).
[6] 株式会社コーエーテクモゲームス：モンスターファーム，（オンライン），入手先 ⟨https://www.gamecity.ne.jp/mf1/⟩ (1997). (Accessed on 07/25/2024).
[7] Yang, J.: Rethinking Tokenization: Crafting Better Tokenizers for Large Language Models, *arXiv preprint arXiv:2403.00417*, (online), available from ⟨https://doi.org/10.48550/arXiv.2403.00417⟩ (2024).
[8] ポノス株式会社：にゃんこ大戦争，（オンライン），入手先 ⟨https://battlecats.club/⟩ (2012)．(Accessed on 07/25/2024).
[9] OpenAI: API Platform | OpenAI (2024).
[10] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. et al.: Language models are few-shot learners, *arXiv preprint arXiv:2005.14165* (2020).
[11] OpenAI: text-embedding-3 (2024). Accessed on 07/25/2024.