周辺視野で情景が揺らぐ リアルタイム動画生成による読書空間演出システム

中島 次郎[1,a)]　波多野 亮平[1,b)]

概要：本研究では，アナログ書籍の読書体験を拡張し，読者が現実世界を離れ，物語の世界へと深く入り込む「物語への移入（Narrative Transportation）」を支援する環境演出システムを提案する．既存手法は中心視野への情報重畳により，読者の内面的な世界構築（Mental Modeling）を阻害する課題があった．本システムは，読書中の周辺視野に対し，物語の文脈に即した「情景動画」をリアルタイムに生成・投影することで，読書空間そのものを物語の雰囲気に浸れる環境へと変容させる．システムの設計指針として，「曖昧さ（Ambiguity）」の有用性を取り入れ，読者の注意を奪う阻害要因（特定のキャラクターや文字情報など）を意図的に排除する「引き算のデザイン」を採用した．また，周辺視野の特性（動き・明暗への感度）を利用した抽象的な映像表現により，環境としての気配を提示する．評価実験では，生成コンテンツの適性（阻害要因の有無）およびシステム応答性の検証を行った．その結果，提案手法が人物等の阻害要因を排除した映像を生成可能であること，および読書ペースに追従可能な頻度（約 3.8 秒に 1 回）で動作することを確認し，システムの技術的妥当性が示された．

1. はじめに

読書とは，単なる情報の受容ではなく，読者が自身の記憶や想像力を動員して物語世界を構築する能動的なプロセスである．読者はテキストの行間にある「空白」を自身のメンタルモデルで埋めることによって，はじめて深い没入感を得ることができる．本研究では，この没入感を，単なる文字を読む行為への集中ではなく，読者が現実世界からの心理的な距離をおき，物語の世界観に深く入り込む「物語への移入（Narrative Transportation）」[1] と定義する．Green らによれば，この状態はイメージ（Imagery），感情（Affect），および注意の集中（Attentional focus）が統合されたものであり，読者が物語の世界に「移動」する体験であるとされる．

しかし，既存の読書体験拡張（Augmented Reading）研究の多くは，AR グラスやタブレット画面を通じて，挿絵や注釈といった具体的かつ鮮明な情報を中心視野に提示するアプローチをとってきた [2], [3]．これらの手法は，情報の「空白」をシステムが強制的に埋めてしまうことで，読者固有の想像を阻害（Imagination Conflict）したり，デバイスによる物理的な負荷を与えたりする課題があった [4]．また，中心視野への過度な情報提示は，物語への移入に必要な「イメージの構築」を妨げ，読者を現実のデバイス操作へと引き戻してしまう恐れがある．

ここで，Gaver らは，一意に定まらない曖昧（Ambiguity）な情報こそがユーザーの解釈を促し，システムへの深い関与（エンゲージメント）を高めるデザインリソースになりうると論じている [5]．特に，効率よりも体験の質が重視される文化的実践（読書や芸術鑑賞など）においては，鮮明すぎる情報は受動的な消費を招きやすい一方，解釈の余地が残された情報は，ユーザー自身がその「隙間」を埋めようとする能動的な想像力を誘発する．

そこで本研究では，この「曖昧さ」の効用を読書支援に応用し，情報の具体性をあえて低減させることで読者の想像力を引き出し，Narrative Transportation を促進するシステムの実現を目的とする．具体的には，アナログ書籍の周辺視野に対し，読者の注意を奪う阻害要因（人物や文字など）を排した「情景動画」をリアルタイムに投影する．Calm Technology の思想に基づき，情報を認知の中心ではなく周辺に配置することで，読書という高負荷な認知タスクを妨げない環境提示を行う．さらに，曖昧な情景表現と時間的に連続した映像提示を組み合わせることで，読書空間全体を物語の感情的文脈として経験させる手法を提案する．

[1] TOPPAN デジタル株式会社
[a)] jiro.nakajima@toppan.co.jp
[b)] ryohei.hatano@toppan.co.jp

2. システムの設計指針

2.1 周辺視野と認知資源の最適化

人間の視覚情報処理は，中心視野と周辺視野で機能が異なる．中心視野は色彩や詳細な形状の認識に優れるが，視野角は狭い．一方，周辺視野は解像度が低いものの，運動と明暗の変化に極めて敏感であり，空間的な雰囲気の把握に特化している [6]．読書中においては，読者の中心視野はテキスト処理に占有されている状態にある．Bahna ら [7] は，読書タスク中の周辺視野に対して関連画像を提示する実験を行い，ユーザが読書速度や内容理解度を低下させることなく，追加情報を獲得できることを実証した．

本研究ではこの知見を発展させ，周辺視野に対して静止画ではなく「情景動画」を提示する．周辺視野は詳細な形状識別には向かないが，動きの変化には敏感であるため，常に揺らぐ環境映像を用いることで，読者に意識的な注視を強いることなく，「環境の気配」を持続的に伝達できると考えられる．これにより，認知資源の競合を回避しつつ没入感を高める（設計指針 A）．

2.2 引き算のデザイン

Gaver らの「曖昧さが関与を高める」という理論を本研究の文脈に適用すれば，没入的な読書体験において AI が提示すべきは，読者のイメージを固定してしまう「正解（具体的なキャラクターの姿など）」ではなく，独自の解釈を誘発する「きっかけ（情景や雰囲気）」であると考えられる．そこで本システムでは，生成 AI をリッチなコンテンツ生成ではなく，読者の注意を奪いやすい対象（Distractor）を意図的に排除する制御に用いる．具体的には，人物，文字情報といった，視線を惹きつけやすい要素を描画させない「引き算のデザイン」を採用する．さらに，生成された映像にガウシアンブラーによるぼかし処理を加えることで具体性を低減し，読者の想像が入り込む余地を技術的に担保する（設計指針 B）．

2.3 時間的連続性

Campos ら [8] は，拡張読書において，読書の流れが途切れない「連続的な読書体験」を保てるような，シームレスで読者の邪魔にならないインタラクションを重視している．一方で，ページめくりごとの静止画が切り替わるような体験は，読書フローを損なう恐れがある．そこで，本研究では，常に揺らぎ続ける「動画」を生成し，無限ループ処理によって時間的な途切れを排除する．これにより，読書という連続的な行為に対して，環境もまた途切れることなく推移する体験を構築する（設計指針 C）．

3. 提案システム

本システムは，ユーザが読むアナログ書籍をカメラでセンシングし，その文脈（情景・感情）を抽出して，周辺視野（壁面など）にアンビエントな動画を投影する．システム構成図を図 1 に示す．

3.1 システム構成

システムは，入力（Sensing），文脈理解（VLM），動画生成（Video Gen），提示（Projection）のパイプラインで構成される．

(1) **入力の安定化**: Web カメラを用いて手持ち書籍を撮影し，VLM へと送信する．はじめに，画像フレームから最大の明領域を抽出することで，画像中における書籍のページ領域を特定する．次に，抽出したページ領域においてラプラシアンフィルタを用いた分散値計算により画像の鮮鋭度を評価する．これにより，手ブレを最小限に抑えた「ベストショット」画像のみを選別して VLM へ送信する．

(2) **文脈理解と阻害要因の排除（VLM）**：大規模視覚言語モデル（VLM）である Qwen3-Omni[9] を用い，画像の情景解釈を行い，動画生成のためのプロンプトを生成する．後段で使用する動画生成モデル（LongLive）はアーキテクチャの特性上，Negative Prompt による生成制御（何を描かないかの指定）を受け付けない．そのため，この VLM の段階でシステムプロンプトにより「人物，キャラクター，顔，文字を描写しない」という強い制約を与え，ページ領域画像から抽出された「場所・時間・天気・雰囲気」などの情景情報のみを記述する視覚生成用プロンプトに変換することで，阻害要因（Distractor）の排除を実現している．また，生成品質安定のため入力のベストショット画像を過去数フレーム分まとめて入力することで，不鮮明な画像入力が差し込まれた場合も頑健に情景情報を抽出可能とする．

(3) **アンビエント動画生成**: 動画生成モデルである LongLive[10] を用い，VLM が生成したプロンプトに基づき情景動画を生成する．動画はリアルタイムで生成し続け，インタラクティブにプロンプトを更新できるようにする．また，構造上の最大フレーム数まで生成したら自動的に初期フレームから生成し直すようにする．これにより，一定間隔（およそ 1000 フレーム毎）で不連続な切り替えが生じるものの，実質無限に映像を生成可能とする．

(4) **視覚的違和感の低減**: 生成された映像に対し，ぼかし処理を適用して投影する．これは周辺視野への刺激を和らげると同時に，生成 AI 特有の微細な描画崩れや

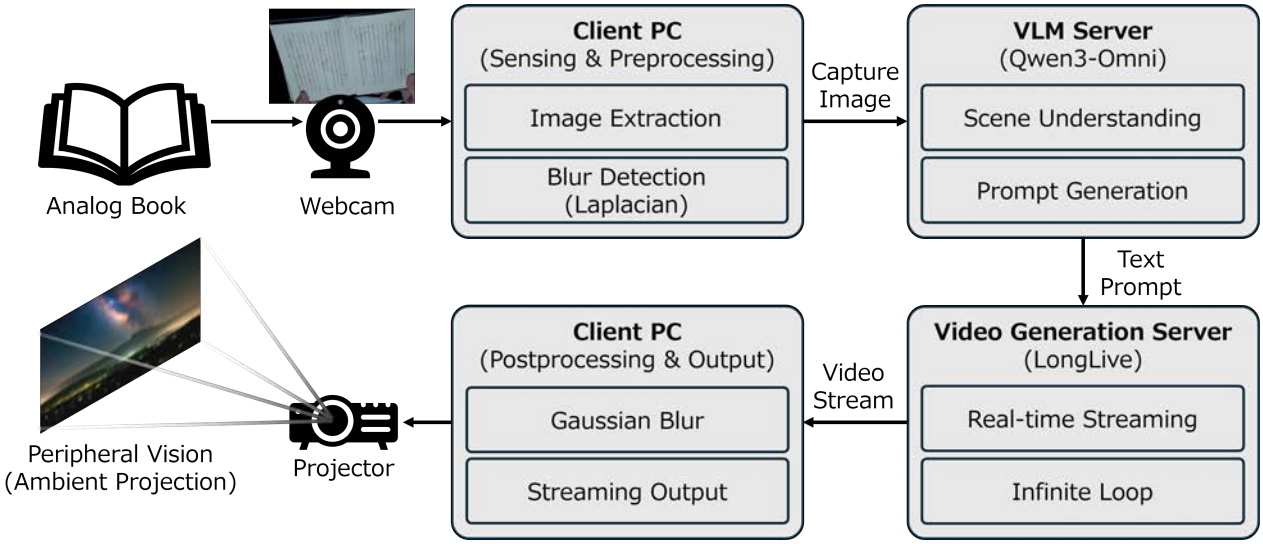

図 1 システム構成図：Web カメラによる書籍画像取得から，VLM によるアンビエントプロンプト生成，LongLive によるループ動画生成，およびぼかし処理を経て投影されるまでのパイプライン．

不自然な動きを隠蔽し，それを「幻想的な揺らぎ」としてポジティブに変換する効果を持つ．

4. 評価実験

本手法によるシステムの評価実験を行った．本章では，提案システムの有効性を検証するための予備的評価を行う．本評価では，没入感そのものの主観的測定を目的とせず，読書を阻害する要因の排除及び，システムの実現可能性に焦点を当てる．

実験は，VLM 推論（RTX A6000）および動画生成（RTX 5090）を行うサーバー群と，クライアント PC をネットワーク接続して実施した．動画生成は，速度と品質のバランスを考慮して解像度 560×320 px，デノイジングステップ数 2 とした．

4.1 実験 1: 生成コンテンツの品質評価

提案手法が，読書を阻害する要因（Distractor）を排除し，周辺視野に適したアンビエントな映像を生成できるかを検証した．

4.1.1 実験設定

青空文庫より，人物の登場や情景描写が特徴的な作品（『銀河鉄道の夜』『羅生門』『武蔵野』等）9 作品を選定した．各作品から固定文字数で切り出したテキストおよびその見開き画像を用い，以下の 2 条件で動画生成を行った．

- **Proposed (With VLM)**: 提案手法．書籍の見開き画像を入力とし，Qwen3-Omni により「Distractor を排除した情景プロンプト」を生成し，LongLive で動画化する．
- **Baseline (No VLM)**: 比較手法．書籍のテキストデータを直接 LongLive に入力し，動画化する．

各作品につき LongLive により連続的に 10 シーン分プロンプトの更新を行い，計 90 シーンに対し，両条件で生成を行い，評価対象として 180 枚の画像を取得した．評価対象画像はいずれもプロンプトの更新から約 2 秒後のフレームを取得した．また，本実験では，生成品質自体の評価のため，周辺視野への提示の際のぼかし処理を行う前の画像を評価対象とした．

4.1.2 評価方法

評価の客観性を担保するため，GPT-5.1-2025-11-13 等を用いた自動評価（VLM-as-a-Judge）を採用した．近年，生成コンテンツの評価において，VLM を用いた自動評価手法が，人間の専門家による評価と高い相関を示すことが報告されている [11]．また，LLM/VLM を評価者として用いる手法の有効性は，対話タスクや画像生成タスクを含む幅広い領域で実証されつつある [12]．本実験ではこれらの知見に基づき，評価プロンプトにおいて評価者を「アンビエントインターフェースの専門家」と定義し，以下の 3 つの指標について 5 段階のリッカート尺度（5 が良い）で採点させた．

(1) **Ambient Suitability（アンビエント適性）**：読者の注意を阻害しないか．特定の人物（主人公など），顔，判読可能な文字などの Distractor が含まれていないことを「良」とする．
(2) **Scenery Fidelity（情景再現性）**：場所・時間・天候が正確か．
(3) **Atmospheric Alignment（雰囲気一致度）**：映像のスタイルがテキストの感情的トーン（Mood）と一致しているか．

4.1.3 結果

各評価指標における平均スコアの比較結果を図 2 に，生

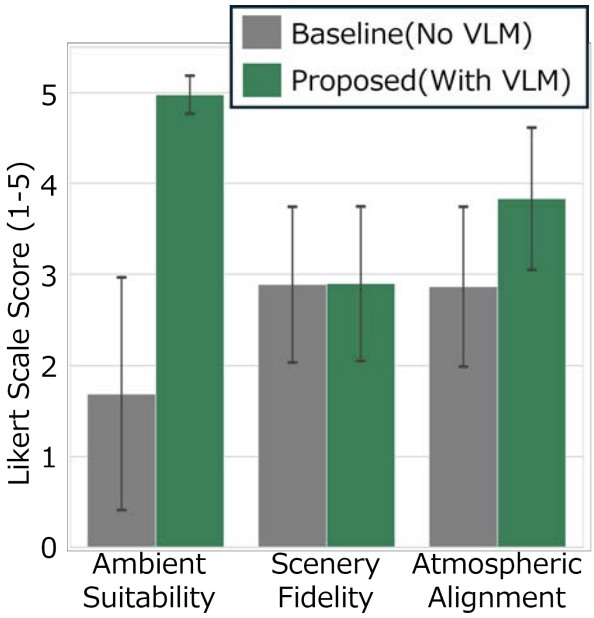

図 2 各評価指標における平均スコアの比較．Proposed (With VLM) は Baseline と比較して Ambient Suitability において有意な向上を示した．

成結果の比較例を図 3 に示す．

4.2 実験 2: システムの応答速度評価

提案システムが，実際の読書ペースに追従してプロンプトを更新し，環境を変化させられるかを確認するため，実機による性能評価を行った．

4.2.1 実験手順

被験者 3 名（20 代，30 代，50 代の男性）に対し，3 つの文学作品の見開き 1 ページを黙読するタスクを課した．その間の「総読書時間」と「プロンプト更新回数」を計測し，システムの更新頻度が読書体験の連続性を担保するのに十分かを検証した．

4.2.2 結果

3 名の被験者が合計 9 回の試行（3 作品× 3 名）を行った結果，総読書時間は 494 秒であり，その間に 129 回のプロンプト更新が行われた．平均更新間隔は約 3.83 秒（494/129 ≈ 3.83）であった．また，見開き 1 ページを読む間に平均で約 14 回（129/9 ≈ 14.3）の更新が発生した．本システムにおいて，最低限見開き 1 ページにつき 1 回以上の情景変化が望まれることを考慮すると，約 3.8 秒ごとの更新頻度は，読者のペースに合わせて細やかに環境を変化させるのに十分な速度であるといえる．

4.3 実験 3: 動作実証とリアルタイム描画性能

提案システムの統合的な動作検証として，実際に異なる 3 つの小説を読書している際のシステム動作風景を取得した．図 4 に，3 つの異なる作品（シーン A:『極楽の池』，シーン B:『雨風の風景』，シーン C:『星空』）を読書中の周辺視野への投影結果を示す．各シーンにおいて，VLM はテキストから情景を読み取り，前述の「引き算のデザイン」に基づき，具体的な人物を描写することなく，作品固有の空気感を反映した映像を生成・投影していることが確認できる．

また，本システムの実環境における動画生成速度を計測した結果，平均して約 15 fps での描画動作を確認した．これは一般的な映像コンテンツ（24-60 fps）と比較して低値ではあるが，周辺視野における環境的な「気配」の提示においては，高フレームレートの滑らかさよりも，ぼかし処理によるフリッカーの低減が重要となるため，十分な品質であるといえる．なお，生成動画の解像度やデノイジングステップ数の調整により，さらなる高速化も可能である．

5. 考察

5.1 アンビエント適性の向上と阻害要因の排除

実験 1 の結果（図 2）が示す通り，Ambient Suitability において Proposed (4.98) は Baseline (1.69) を上回った．Baseline では，動画生成モデルがテキスト内の「登場人物名」や「動作」に反応し，人間等を描写している．一方，Proposed では VLM が Distractor を描かないという制約のもとでプロンプトを生成するため，Ambient Suitability Score が高くなり，読者のメンタルモデル構築を阻害しない「余白」のある映像提示が可能となったと考えられる．

5.2 忠実度と抽象度のバランス

Scenery Fidelity（情景再現性）においては，両条件に有意な差は見られなかった．これは，VLM による抽象化を経ても，物語の主要な舞台設定（場所・時間・天候）は損なわれていないことを示唆する．また，Atmospheric Alignment（雰囲気一致度）においては Proposed が高いスコアを示した．これは，VLM がテキストから単なる物体情報だけでなく，「寂寥感」や「不気味さ」といった感情的トーン（Mood）を明示的に抽出し，それを映像スタイルとしてプロンプトに反映できたためと考えられる．

5.3 リアルタイム生成の意義とスケーラビリティ

体験の質を局所的に最大化するという観点のみに立てば，事前に人間が制作・監修した高品質な映像素材を用意し，それを再生するアプローチの方が，演出の整合性は高まる可能性がある．しかし，そのような作り込みを前提とした手法では，コンテンツ制作に多大なコストと時間を要し，対応できる書籍が極めて限定される．

本システムにおいて生成 AI，および VLM によるリアルタイム生成を採用した最大の理由は，その「即興性」と「スケーラビリティ」にある．生成 AI を用いることで，事前の学習やコンテンツ準備を必要とせず（Zero-shot），ユーザが手に取ったあらゆる書籍（小説，実用書，未知の作品な

図 3 生成結果の比較例．上段「武蔵野」Baseline では地図や文字が出現するが，Proposed では古戦場の情景のみが描画．下段「銀河鉄道の夜」: Baseline では人物が描画されるが，Proposed では教室と星空の情景が描画．なお，生成モデルの学習バイアスにより，生成映像に西洋的な視覚特徴が含まれ，書籍の地理と明らかに異なる映像が生成される場合がある．しかし，本システムではこれらにぼかし処理を適用して提示するため，細部の文化的・様式的な不整合は抽象化され，体験上の課題とはならない．

図 4 システム動作の実証例．左から極楽の池，雨風の風景，星空の表現と，読書進行に伴い，文脈に即した抽象的な情景動画が投影されている．

ど）に対して，即座に環境演出を提供することが可能となる．これは，特定の作品専用のインスタレーションではなく，日常的な読書行為そのものを汎用的に拡張するプラットフォームとしての本質的な要件である．本システムは，生成 AI を「コスト削減」のためだけでなく，未知の文脈に対して動的に適応し続けるための不可欠なエンジンとして位置づけている．

5.4 視覚入力による汎用性の拡張

本システムは，入力情報としてテキストデータではなく「書籍の画像」を採用している．これにより，文字情報のみならず，挿絵やレイアウトを含めたページ全体の情報を VLM が統合的に解釈可能である．この特性は，小説に限らず，漫画や絵本といった視覚情報が主体の媒体への適用可能性を示唆している．VLM は漫画や絵本の描画内容から直接情景や雰囲気を読み取ることができるため，本システムは多様なアナログ媒体に対応する汎用的な読書拡張プラットフォームとして機能しうる．

5.5 著作権リスクの回避と倫理的配慮

インタラクティブな生成 AI 利用において，既存の著作物（キャラクター等）に酷似した画像を生成してしまう著作権侵害リスクは重要な課題である．特に漫画や小説を対象とする場合，特定のキャラクターを AI が再生成することは法的な懸念を招く．本システムは，人物やキャラクターを含む「阻害要因（Distractor）」を描かないという制約をコアに組み込んでいる．これにより，著作権で保護されたキャラクターの姿を直接描画することを構造的に回避している．この「引き算のデザイン」は，読者の想像力を保護するだけでなく，生成 AI をコンテンツ産業で安全に利用するための倫理的・法的なセーフガードとしても機能する．

5.6 限界と今後の展望

本研究の評価は，生成コンテンツの適性およびシステム応答性の技術的検証に焦点を当てたものである．したがって，提案システムが実際の読書体験において，ユーザの主観的な「没入感」を統計的に有意に向上させるかについては，未だ検証されていない．今後の課題として，被験者を

用いた長期的な比較読書実験を行い，主観評価尺度（フロー体験尺度など）や，アイトラッキングなどの生理指標を用いた客観評価を通じて，本システムがもたらす心理的変容を実証する必要がある．

5.7 限界と今後の展望

本研究の評価は，生成コンテンツの適性およびシステム応答性の技術的検証に焦点を当てたものである．したがって，提案システムが実際の読書体験において，ユーザの主観的な「物語への移入（Narrative Transportation）」を統計的に有意に向上させるかについては，未だ検証されていない．今後の課題として，被験者を用いた比較読書実験を行い，Green らが開発した指標 [1] を用いた主観評価等を通じて，本システムがもたらす心理的変容を多角的に実証する必要がある．

6. おわりに

本研究では，アナログ書籍の読書体験を拡張するシステムを開発した．180 サンプルの比較評価と実機性能評価により，提案システムは以下の 2 点を達成していることが示された．第一に，VLM により動画生成のプロンプトを制御し，人物や文字などの読書の阻害要因を排除したこと．第二に，平均約 3.8 秒間隔でのプロンプト更新を実現し，実際の読書ペースに対して十分な追従性を持つこと．また，視覚入力を採用したことで漫画や絵本への拡張性を持ち，かつ人物を描画しないことで著作権リスクを低減する安全性も備えている．「周辺視野への介入」「引き算のデザイン」「時間的連続性の確保」という 3 つの指針に基づいた本システムは，電子書籍にはない「未来のアナログ読書体験」となりうる可能性を秘めている．

参考文献

[1] Green, M. C. and Brock, T. C.: The role of transportation in the persuasiveness of public narratives., *Journal of Personality and Social Psychology*, Vol. 79, No. 5, pp. 701–721 (online), DOI: 10.1037/0022-3514.79.5.701 (2000).
[2] 西ノ原司瑳，河野恭之：小説読書中の視線移動に基づく挿絵自動提示システム，インタラクション 2025 論文集, pp. 972–976 (2025).
[3] Singapore, N. L. B.: Augmented Reading - The National Library of Singapore x LePUB, Available at https://www.youtube.com/watch?v=EysXvf5rmBk (Accessed: 2025-12-15) (2025).
[4] Cai, M.: Enhancing Active Reading: A Human-Machine Co-Creation Journey for Visualized Narratives Reading, Master thesis, TU Delft, Faculty of Industrial Design Engineering (2024).
[5] Gaver, W. W., Beaver, J. and Benford, S.: Ambiguity as a resource for design, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, New York, NY, USA, Association for Computing Machinery, pp. 233–240 (online), DOI: 10.1145/642611.642653 (2003).
[6] Strasburger, H., Rentschler, I. and Jüttner, M.: Peripheral vision and pattern recognition: a review, *Journal of Vision*, Vol. 11, No. 5 (online), DOI: 10.1167/11.5.13 (2011).
[7] Bahna, E. and Jacob, R. J. K.: Augmented reading: presenting additional information without penalty, *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, New York, NY, USA, Association for Computing Machinery, pp. 1909–1912 (online), DOI: 10.1145/1056808.1057054 (2005).
[8] Campos, C., Ducasse, J., Pucihar, K., Geroimenko, V. and Kljun, M.: *Augmented Imagination: Creating Immersive and Playful Reading Experiences*, pp. 57–81 (online), DOI: 10.1007/978-3-030-15620-6_3 (2019).
[9] Xu, J., Guo, Z., Hu, H., Chu, Y., Wang, X., He, J., Wang, Y., Shi, X., He, T., Zhu, X., Lv, Y., Wang, Y., Guo, D., Wang, H., Ma, L., Zhang, P., Zhang, X., Hao, H., Guo, Z., Yang, B., Zhang, B., Ma, Z., Wei, X., Bai, S., Chen, K., Liu, X., Wang, P., Yang, M., Liu, D., Ren, X., Zheng, B., Men, R., Zhou, F., Yu, B., Yang, J., Yu, L., Zhou, J. and Lin, J.: Qwen3-Omni Technical Report (2025).
[10] Yang, S., Huang, W., Chu, R., Xiao, Y., Zhao, Y., Wang, X., Li, M., Xie, E., Chen, Y., Lu, Y., Han, S. and Chen, Y.: LongLive: Real-time Interactive Long Video Generation (2025).
[11] Wu, T., Yang, G., Li, Z., Zhang, K., Liu, Z., Guibas, L., Lin, D. and Wetzstein, G.: Gpt-4v (ision) is a human-aligned evaluator for text-to-3d generation, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22227–22238 (2024).
[12] Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E. and Stoica, I.: Judging LLM-as-a-judge with MT-bench and Chatbot Arena, *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, Curran Associates Inc. (2023).