『江戸川大学紀要』第 34 号（2024 年 3 月）pp.41-69

AI で学生は創造的になれる（動画・3D）

植田　康孝*・石川　妃菜**・新井　　心***・市川　　栞***・伊藤　颯真***・
今関　真央***・木南　璃遥***・花田　美織***・三好　　葵***・柳生　晴香***

要　旨

コロナ禍の 3 年間，日本だけがマスクをしてぼ〜っとしている間に世界は劇的に動いていた。パンデミックにおいて，人を 1 か所に集めることのリスクが膨れ上がり，クラウド化が急速に進み，「全員集合」がオワコン化した。映像制作の世界では，コロナ以前から「リモート・プロダクション」が試行されていたが，人は移動せず自身の本拠地から作業する「クラウド化」が進んだ。2023 年 4 月にラスベガスで行われた「NAB ショー2023」ではクラウドと AI がメインテーマとなった。2021 年までの映像生成技術の主流であった「GAN」から，2022 年以降に「**拡散モデル**」へと AI 技術が進化したことで，高品質の映像生成が可能となり，2025 年から 2027 年に掛け**人工合成（シンセティック）**動画が動画コンテンツ全体の 90％を占める。更に映画作品でも，2022 年時点では 0％だったが，2029 年には 90％を占めるようになる。生成 AI は作業効率を高めるだけでなく，作品の質を高める。撮影した動画をアニメーションにしたり，アバターが話す言葉を多言語化したりすることが容易である。既存のテンプレート素材を使わずにオリジナルの背景画像を生成することで表現の幅を広げられる。様々な**動画生成** AI ツールが登場したことにより，完パケ（リニア）ではない「ノンリニアコンテンツ」が急増している。AIによる映像編集の最前線では，完パケ（リニア）からオブジェクトベースにシフトするように，AI 技術が映像制作のワークフローを変革した。完パケで育ったプロと置き換わるように，ゲームや音楽でオブジェクトベースに慣れ親しんだ若者世代の台頭が期待される。「完パケ」に慣れたベテランの脚本家や俳優と，「OBM」「OBB」(1)に向かう配信事業者との意識の差は，ハリウッドの**大規模ストライキ**という形で顕出した。生成 AI は映像だけでなく 3D オブジェクトにも向かう。テキストから動画を生成するだけでなく，3D モーションを生成する AI も登場した。「動画配信」から「空間配信」へのシフトは歴史に残る不可逆的な変化である。ゼミナールでは，LiDAR スキャンを用いて現実空間を 3DCG にオブジェクト化してバーチャル空間（ワールド）を製作すると共に，モーションキャプチャーしたアバターと位置を一致させ，3D モーションを生成した。「いつの時代の話だ」と語られる，コロナ禍前のような拘束時間が長い「全員集合」にもう戻ることはない。生成 AI を核とした構造転換により働き方改革につなげたい。2023 年 12 月 19 日，Google は従来の「拡散モデル（Diffusion Model）」とは異なる「大規模言語モデル（LLM）」に基づく動画生成 AI「Video Poet」を発表し動画生成 AI は新時代を迎えた。

キーワード：空間モデリング　GAN から拡散モデルへ　フルトラッキング　モーションキャプチャー　ニューラル場（NeRF）　OBM　OBB　完パケからノンリニアへ　オブジェクト　メタバース　Video Poet

1. 映像生成 AI をめぐる動向

1.1.　映像制作の構造変化

AI による映像編集の最前線では，AI 技術によって映像制作のワークフローが変革する。ハリウッドのストライキで組合が掲げた大きなテーマが制作における「**生成 AI の利用制限**」である。生

2023 年 11 月 30 日受付
*　江戸川大学　マス・コミュニケーション学科教授　理学博士（国際情報通信学）
**　江戸川大学　植田ゼミ第 15 期　令和 5 年度卒業生　総代
***　江戸川大学　植田ゼミ第 16 期

成AIの能力は，人の指示によって脚本を創作できるまでになった。動画を生成する点では既に実用段階に入った。絶望的にカメラや編集装置の操作センスがない機械音痴の学生でも動画生成AI「**Runway**」を使えば，動画を作れる。しかも1回出してくれたものに対して，線で直して欲しい箇所を囲って，イメージが合っているが，この部分だけもっとこうして，とリクエストすると修正してくれる。つまり，自分の思考と成果物（動画）が直結する。昔から映像制作は不毛な作業だと思われており，知的労働者がやる仕事ではないとされた。頭の中ではどのような映像を作りたいかイメージは出来ているのに，膨大な時間と労力を掛けて企画書や撮影台本に書き起こすなど無駄な作業に労力を費やさないといけない。作業しない人が多数集まる全員集合型のワークスタイルは少し頭の良い人であれば，無駄な待機時間が多く割に合わない非効率な仕事（タイパが極めて悪い）に映る。それが今や動画生成AIにプロンプト（指示文）を突っ込めば，自動で動画を生成してくれる。結果として知的労働者が興味を持つ業務になりつつある。生成AIの普及により映像の世界は二極化して行く可能性が高い。AIの映像生成能力は優れており，AIを活用する知的技術者と，代替される肉体技術者に二分される。AIが取って代わるのは肉体技術者の業務であり，知的労働者は生成AIを使うことでアウトプットが5倍，10倍になる。

1.2. GAN（敵対的生成ネットワーク）から拡散モデルへ

1.2.1. GAN（敵対的生成ネットワーク）

https://www.youtube.com/watch?v=2rC2_-HtpsQ

近年，映像制作で注目されていた人工技術が「敵対的生成ネットワーク」（Genera tive Adversarial Networks）であった。**GAN**は生成モデルの一種で，データから特徴を学習することにより，実在しないデータを生成して存在するデータの特徴に沿って変換できる。GANは，正解データを与えることなく特徴を学習する「教師なし学習」の一手法として注目された。教師データを作成する手間が省けるGANは，2つのニューラルネットワークで構成される。1つはGeneratorで，データを生成する。Generatorはランダムノイズを入力することで，ノイズを所望のデータに近付けるようにマッピングする。もう1つはDiscriminatorで，Generatorが生成した偽物のデータと本物のデータが与えられ，その真偽を判定する（**図1**）。**GAN（敵対的生成ネットワーク）**はAIに機械学習させるため大量のデータを必要とした映像生成分野で，少ないデータからでも効果的に様々な映像パターンを繰り返し，学習できるようにした点で画期的であった。

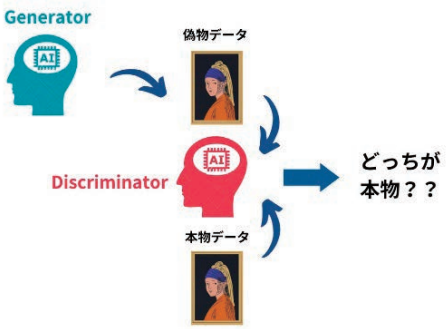


図1　GAN（敵対的生成ネットワーク）

1.2.2. 拡散（ディフュージョン）モデル

2021年以前はGANが注目されていたが，2022年に生成AIが出現した後は「拡散モデル」で議論されるようになった[(2)]。2021年度や2022年度のゼミや演習実習ではGANを用いた映像生成の実習を行なったが，現在から見ると余りクオリティの高いものではなかった。2022年後半に「**拡散モデル**」というAI技術が登場すると，動画生成AIは飛躍的な発展を遂げた。**図2**のように，画像データに少しずつノイズを加えると，画像データはノイズによって少しずつ元の情報を失って行き，最終的には元の情報を完全に失ったノイズそのものになる。徐々に加えて行くノイズの大きさを上手く調節すると，画像データは最終的に正規分布と等しい確率密度関数を持つノイズ（ガウシアンノイズ）に収束する。ランダムにノイズを加えて行くプロセスを逆向きに辿り巻き戻すこと

で，完全なノイズの状態からノイズを少しずつ除去して行けば，ノイズのない元の美しい画像データを生成できる。犬の画像を少しずつ壊して行くと，全くなくなる。それを逆回しする，何もない状態から画像を生成することが出来る。データを拡散して破壊した時に戻り復元できる経路を求めるため，「**拡散モデル**」と呼ばれる。動画や3Dを理解しているため，簡単に条件付け出来るため，数多くのタスクに用いられる。2022年後半，Stable Diffusion（ステーブル・ディフュージョン）など「**ディフュージョン**」（拡散モデル）を使った生成AIがオープンソース化されて世界中の人々に使用されるようになった。筆者（植田）は，2023年度のゼミや演習実習では，動画生成において使用するAIをGANから拡散モデルへ内容を変更した。

図2　拡散モデル

1.2.3.「マルチモーダル」へ

2023年前半まで1種類の情報しか処理できないシングルモーダルであったが，2023年後半に登場した**マルチモーダル**では複数の情報を一度に処理できるようになった。テキスト，画像，音声，動画などを一緒に入力して映画もアニメーションも製作できる。多くの人はいまだチャットGPTを消費者向けの「質問ができるツール」として捉えているが，本稿が出る2024年3月時点では，それはかなり過小評価である。2023年までの個別の機会学習モデルを作る時代から，2024年からは基板モデルに使う時代に移る。膨大なテキストデータで学習した「LLM（大規模言語モデル）」のように大量のデータで事前学習したAIは「**基盤（ファンデーション）モデル**」と呼ばれ，基盤モデルを解明・改良してその能力をもう一段高める試みである。**マルチモーダル基盤モデル**は，映像，音声，画像，テキストなど異なるモーダル間の情報も自由につなげることが出来，総合的に使えるようになる。**表1**に見られる通り，2022年から2023年に掛けて映像生成AIにより膨大な動画素材（オブジェクト）が生成され，「OBM」「OBB」を進展させた。人間の映像スタッフがAIの後塵を拝するのは2025年頃である。2030年頃には，スタジオという空間があらゆる面でAIとリンクし，今とは全く違うものになる。AIを駆使する2030年のスタジオでは，人間の五感に頼った制作では到底入手できないレベルの質を実現し，人間よりもはるかに高質な3D作品を低コスト・短時間で提供できるようになる（**表1**）。**マルチモーダル・モデル**は，2023年後半から2024年に掛けて本格化するとされ，2023年度は間に合わなかったが，2024年度のゼミや演習実習では導入したい。更に工学的に優れた，拡散モデルの良い部分を抽出したモデルである「**プロマッチング・モデル**」が登場しており，既にメタ（Meta）が発表した音声合成で使った。

表1　生成系AIの発展

生成系AIの発展

達成年は更に早まる見込み

	2020	2022	2023	2025	2030
テキスト	コピーライティング アイデア出し	長文生成 文章の編集	各業界の用途に最適化した文章作成	平均的な人間の文章能力を上回る	プロのライターの文章能力を上回る
コード	複数行のコード生成	大規模なコード生成 高精度なコード編集	専門的な開発要件に対応	テキストからプロダクトを生成する	フルスタックエンジニアの能力を上回る
画像		アート ロゴ 写真	プロダクトのモックアップの生成	プロダクトデザインの完成版を作成	プロのデザイナーの能力を上回る
動画		簡単な動画素材の生成	汎用的な動画素材の作成	動画の生成	ゲームの生成 映画の生成

出所：株式会社デジタルレシピ（2023）(3)

1.2.4.　2030年問題と生成AI

続々と改良されたモデルが開発され，**表1**より達成年は早まっているとの見方が大勢である。伴ってテキスト，画像，映像などの産業で急速に置き換えが起きる。境（2023）は「メディアに間違いなく起こる**2030年問題**とは，テレビや新聞などはレガシーメディアと呼ばれ，いつかなくなってしまう存在として扱われて来たが，それがいよいよ，本当になくなる日が近づいて来た。タイム

リミットは**2030年**である」「新聞は2030年までに激しい荒波に揉まれた上に，それ以降は更に部数が物凄い勢いで減少する。新聞購読者の核は現在の70代以上つまり団塊世代とそのさらに上の世代だ。その人々の数が急激に減って行く。新聞社はドタバタと倒れて行く可能性が高い」「テレビも激しく疲弊する。2021年度からテレビ局の放送収入は下降の一途だ。ゴールデンタイムの視聴率が2021年度，2022年度と3%程度ずつ減少しており，その傾向は止まる気配がない。民放キー局5局（日本テレビ，テレビ朝日，TBS，テレビ東京，フジテレビ）の本業である放送収入の合計の推移を見ると，減少傾向である。

2017年度　約8,891億円
2019年度　約8,461億円
2022年度　約8,000億円

5年間で約10%，900億円近く減少している。コロナ禍での巣ごもり生活でテレビをネットに繋ぎ配信サービスを楽しむ人が増え，テレビ放送がどんどん見られなくなっている。テレビ離れが若者だけでなく，中高年や高齢層にも波及しつつある。朝ドラや大河ドラマの視聴率が以前より低いのは，出来の良し悪しより高齢者でさえテレビ離れを起こしている証である。もはやテレビ広告市場が上向くことはない。2030年，マスメディアは本当に多くの事業者が消失する危機を迎える」と指摘する[(4)]。東京商工リサーチによると，既に2023年，過去最高となる14社のテレビ制作会社が倒産した。前年（2022年，6社倒産）の2.3倍であった。新聞社やテレビ局などマスメディアが生き残る方策は，生成AIを活用して生産性を高める他ない。既に映画業界では，脚本家や俳優らによる**ストライキ**が起きたが，事業者も簡単に引けない経営状況にある。映画業界ストライキと生成AIの関係については，**6項**で詳述する。

1.3.　テキストから動画を生成するAI

生成AIには，テキスト（文章）の他，音声，画像，動画を生成するAIが存在する。2022年までは最も難しいとされていたのが動画生成であったが，2023年になり，動画を生成する生成AIツールが続々リリースされた。**動画生成AI**は，画像生成に使われる手法を時間軸に沿って展開することで，連続する画像フレームを生成する技術で，短いクリップや長い動画を作成することが出来る。**表2**のようなAIツールが登場しており，2023年度のゼミや演習実習で使用した。

表2　最新動画生成AI

企業名	動画生成AI	
Runway	Gen2	https://research.runwayml.com/gen2
Seyhan Lee	Cuebric	https://seyhanlee.com/
Flawless	TrueSync	https://www.flawlessai.com/product
Wonder Dynamics	Wonder Studio	https://wonderdynamics.com/

2. 動画生成AIの実習

2.1.　AIリテラシーの必要性

既に画像だけでなく動画もAIで製作，編集することが当たり前になっている。結果，コンテンツ製作費用を大幅に削減できる。第一線で活躍する写真家や映像クリエイターへの影響は小さいが，そこそこの品質の作品を量産して来た粗製乱造・薄利多売の人には，受注が減ったり単価が低下したりする影響は免れない。つまりクリエイターは二極化が加速する。実習で動画生成に取り組んだ学生には分かるが，動画のイメージをテキストで伝えることは非常に難しく，新たなスキルとして確立されるため，動画も生成AIを上手く使いこなせるクリエイターと時代遅れのクリエイターに二極化される。

2.2.　AIによる動画生成

清水（2023）は「YouTubeやTikTokでおすすめされる動画は今後，ユーザーの好みや閲覧履歴を踏まえてAIが自動生成したものになる。早ければ2024年にはそのような世界が訪れる」と言う。パーソナライズされたAIの未来では，自

分の趣味・嗜好に関する情報をAIが学習し，それに合わせた様々なキャラクター動画を出力したり，気分に合わせメンタル面を増幅・修復するような映像を出力したりする。動画生成では2018年設立の米ランウェイ（Runway）のAIが既に映画製作の現場で使われ始めた。RunwayはStable Diffusion（画像生成AI）を共同開発していたが，2022年10月以降，動画生成に集中しサービスは全てクラウドで展開している。

2.3. Runway「Gen2」

アプリ名称	URL
Runway「Gen2」	https://research.runwayml.com/gen2

URL：https://runwayml.com/
参考解説URL：https://find-a.jp/seotimes/runway/

Runwayは，Webブラウザから簡単にアクセスして試すことが出来，テキスト（プロンプト），映像，画像の3通りの指示方法により質の高い動画を生成することが可能である。筆者（植田，石川）が使った「無料プラン」では125クレジット（使い切り）が付与され，**Gen-1**の機能（動画から動画を作る）は1秒あたり14クレジット，**Gen-2**の機能（テキストまたは静止画から動画を作る）は1秒あたり5クレジットを消費する。また，生成1回あたりの動画の長さは最大4秒となっており，筆者（植田，石川）が生成した動画も4秒である。動画スタイルには，SF風，メタル風，スペース風，未来風，クレイメーション（粘土細工によるコマ撮りアニメーション）風など24種が用意され，動画生成前に選んだスタイルに基づく静止画のプレビューが4パターン提示される。その一つを適宜選択して動画生成を行う。上級編として，まったく架空の情景の動画を，テキスト（プロンプト）のみで生成することも出来る。筆者（石川，植田）はこの方法で試したので，以下に紹介する。

【概要説明】

クレジットの制限を意識しながら，筆者（植田，石川）は「Gen2」を試行した。

（操作手順）

操 作1　Runwayサイトで「TRY RUNWAY FOR FREE（ランウェイを無料で試す）」を選択。
操作2　アカウント作成でメールアドレスを入力。
操作3　ユーザー名，パスワード入力。
操作4　名前のみ入力し，認証コードを入力し，次の画面で「skip」を選択して，画面に沿って「try for free」を選択する。
操作5　黒い画面の左videoの欄にある「Generate videos」を選択し，「Gen-2:Text to Video」を確認する。

動画生成AI「**Gen2**」は前モデル「Gen1」のアップグレード版で，テキストからフォトリアルな動画を生成できる。生成動画の長さは4秒であるが，クリエイター界隈での注目度は非常に高く，ブルームバーグなど大手メディアがこぞって同ツールに関する情報を伝えたため，利用者が急増した。米国大学の映像教育では当たり前となり，ネットには学生から数多くの映像作品が投稿される。映画では，歴史上の人物と俳優が話す架空のシーンが良くあるが，2023年時点でそのレベルの動画を誰もが生成できる。

動画生成AIは，動画として創り出したい物語を文章で入力することにより，現実にありそうな写実的な日常や，現実には存在しない空想の世界を動画として生成してくれる。例えば，AIに「A young couple walking in a heavy rain（大雨の中を歩く若いカップル）」とテキストで入力すると，写実的な日常の動画が生成される。写実的な日常のみならず，現実には存在しない空想の世界も動画生成してくれる。例えば，「A panda bear driving a car（車を運転するパンダ）」と文章入力すると，パンダが車を運転する動画が生成される。

2023年度時点では，ランウェイ（Runway）が動画生成のトップランナーであったため，ゼミや演習実習でも優先的に扱った。筆者（植田）がdiscordで「Japan's Most Popular Idol Singer Runs the Beach」と入れると1分くらい待つと4秒の動画が生成された（**図3**）。短尺での対応で

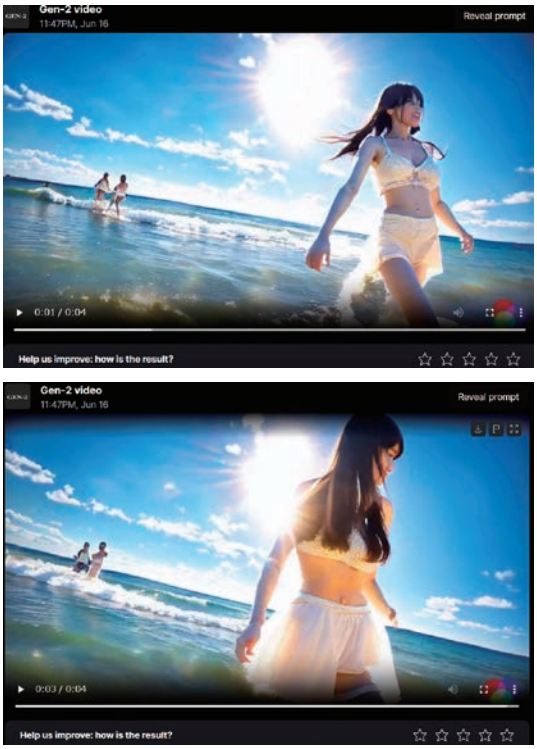


図 3　Runway「Gen2」（植田生成）

あり，画像生成 AI に比べると，生成されるキャラクターや衣装，背景のクオリティは劣るが，走る動作についてはカクカクすることなくスムーズに動いた。アドビソフトを使わなくても，テキストを入力するだけで誰もが簡単に動画を作成できる[(5)]。視聴ユーザーに合わせた膨大な映像素材（オブジェクト）の製作が可能となったため，企業は動画製作を内製化することで動画のオリジナル化，量産化が可能となる。広告として A/B/N テスト用メッセージを自動生成できる。長さ，言語，口調も自由自在であり，目的に合わせ瞬間に生成できる。皆が動画製作できる時代である。

【筆者（石川）の感想】

Runway は，生成の仕方が Midjourney に似ていると思います。人間の語彙力が試される部分や生成結果を 4 つ提示してくる部分など，特徴が共通していると感じました。プログラム AI「教えて，MENTA くん」に，動画生成 AI について質問したら Runway をオススメされました。植田先生のおっしゃる通り，Runway が動画生成 AI でトップを走っていることは間違いないようです。筆者（石川）も動画生成 AI に取り組み，**図 4** の結果を得ました。

図 4　Runway「Gen2」（石川生成）

動画の出来具合や「プロンプトは詳しい方が良い」という植田先生のアドバイスから，ユーザーがどれだけイメージした情景を文章に可視化できるかが一番の鍵であることを良く理解できた。具体的な指示を出せるかが人間に求められる能力であると再確認した。画像をアプロードできたり，生成前に静止画でプレビューできたりする仕様は，上手く説明できなくても「こんなイメージ」と伝えられる補助的仕様として非常に役立った。テキストだけで指示するよりも圧倒的にイメージ画像と共に指示した方が，生成動画に納得が得られた。イメージが違ったと後悔してクレジットを使うよりも，生成前に確認して納得して作れることが良いポイントだと思う。無料プランでは，1 動画 4 秒で生成されるため，約 26 回動画を生成できるが，少なくとも 1 動画 15 秒を生成できるようになると TikTok で活用し易いと思った。

生成動画から，風景映像は得意だが，人の動き

や人の表情が苦手という特徴も分かった。**図4**は，美女と野獣の実写をモチーフにダンスをする男女の映像を指示したが，ドレスの動きに足の動きが追い付かない，ドレスと足が同化，男性の様子がおかしいなど違和感があった。動画で人を違和感なく生成するのは2023年時点では画像より不得意であると感じた（**図4**）[6]。

2.4. 動画編集「Synthesia（シンセシア）」

アプリ名称	URL
「Synthesia（シンセシア）」	https://www.synthesia.io/ https://miseruit.com/2023/04/03/post-4489/ https://note.com/yuru2creative/n/n800c1c591d0a

URL：https://www.synthesia.io/home-update

AIビデオ生成プラットフォームであり，ユーザーが用意した文章を話すAIプレゼンターのアバターを選択し映像作成できる。AIプレゼンターは，様々な言語や口調で話すことが出来る。日本語だと違和感はあるが，人間だと見間違えてしまうほどの精度に到達しており，驚く。

（概要説明）

- 登録不要で無料お試しコースがあり，筆者（植田，石川）は利用した。
- 右側にサンプルイメージが表示される。
- 60カ国語以上に対応，テキスト文字数は半角200文字まで。

（操作手順）

操作1　「Create a free AI video（無料のAIビデオを作成する）」を選択。

操作2　「ビデオのテンプレート」を選択し，「ビデオスクリプト」のテキストを編集して，「続く」を選択。

操作3　必要事項を入力して「Generate Free Video」ボタンをクリック。

操作4　クリック直後，「We're reviewing your AI video（お客様のAIビデオを確認中です）」というタイトルのメールが届き，数分後に「Your AI video is ready（お客様のAIビデオが準備できました）」というメールが届く。

操作5　2通目のメールを開いて，「Click Here To Watch Your AI Video」をクリックすると，ブラウザが開いて作成したAI動画を視聴することが可能となる。

筆者（植田）もやってみた。

I am Taylor Swift, a professor at Edogawa University. Today I will give a lecture on how to compose music with artificial intelligence. Please attend with interest.

（私は江戸川大学の教授であるテイラー・スウィフトです。今日は人工知能で音楽を作曲する方法について講義してみましょう。興味をもって受講ください。）と入力して，名前，アドレスを登録すると，数分経つと，指定したメールアドレスに動画が生成され送付されて来た。動画では見事な会話が英語で読み上げられた（**図5**）。完成したビデオに最大10分しかアクセス出来ないが，十分に使えるクオリティに到達していた。テレビ局でも天気予報や株価情報，スポーツの結果などをリアルタイムで発信できる。簡単に自社で動画広告や店頭や商品棚のデジタルサイネージでの広告を作れるようになるため，外注しなくても自社で低コスト，短時間で動画製作できる。メディアでなくても，エントランスの案内や定型的な受付対応に使える。

https://share.synthesia.io/76c48e7a-1ba4-4512-8dcc-97589631bd04

図5　Synthesia（植田）

【筆者（石川）の感想】

就活面でポートフォリオ用作品として活用できる可能性を考え，日本語，英語の2本の動画を生成した。AI生成の手順は簡単で，1通目のメールは3秒以内で届き，2通目のメールも1分以内に届いたため，他のAIと比べて手続きが早い。

手順も面倒なく簡単である。難しい言葉が使われないため億劫にならずに試行できた。但し日本語で作った場合，リップシンクが合っておらず違和感が残った箇所があった。英語の場合もローマ字が苦手か，英語の訛り判定か分からなかったが，Edogawaの発音（エディゴゥに聞こえる）に違和感を持った。日本語の訛り（イントネーション）の調整が出来れば，PR動画作成に十分使えると思った（**図6**）。

図6　Synthesia（石川）

2.5.「Control Video Net」

テキストのプロンプト（指示文）で自分の思い通りの動画を生成することが非常に難しいことは，「Runway」を行った学生の感想でも多く散見された。しかし，「Stable Diffusion」の拡張モデルとしてリリースされた「**Control Net Video**」を使えば，テキストで細かく指示する代わりに，人物などの動く動画を参考にして動画を生成してくれるようになる。キャラクターと動画を入力すると，動画の「テイスト」を抽出して適用することで，思い通りの動画を出力してくれる。「**Control Net Video**」は**図7**で示す通り，実写で撮影した左の女性の動画を参考にして右のキャラクターに同じ動きをさせることが出来る（**図7**）。テキスト（文章）による動画生成は構図やポーズ，動作の再現性が低いため，お手本となる動画を使って指定する方が，生成動画の精度は高くなる。生成動画は多少粗さが多少目立つものの，動きの滑らかさは人間と変わらない。

図7　Control Video Net

2.6.「Make-A-Video」（メーク・ア・ビデオ）

名称	URL
Make-A-Video	https://makeavideo.studio/

Metaが2022年9月29日に発表した。テキストで内容を指定することで動画生成できる。写真に動きを付けたり，2枚の写真の間の動きを補完して動画にしたりすることも可能である。テキストからの動画生成に加え，静止画からの動画，動画から動きの異なる動画を生成できる。テキストから画像を生成する処理と，生成画像に基づいて動画を構成するフレームを生成する処理が行われる。多数の動画を生成する処理では画素数を上げる超解像も実行される。

図8　Make-A-Video（Text to Video）

【ケース 1】テキストから動画を生成する

「A teddy bear painting a portrait（自画像を描いているテディベア）」とテキストで入力すると，動画を生成してくれる（**図 8**）。

【ケース 2】画像から動画を生成する

1 枚の画像を動画に生成し直す（**図 9**）

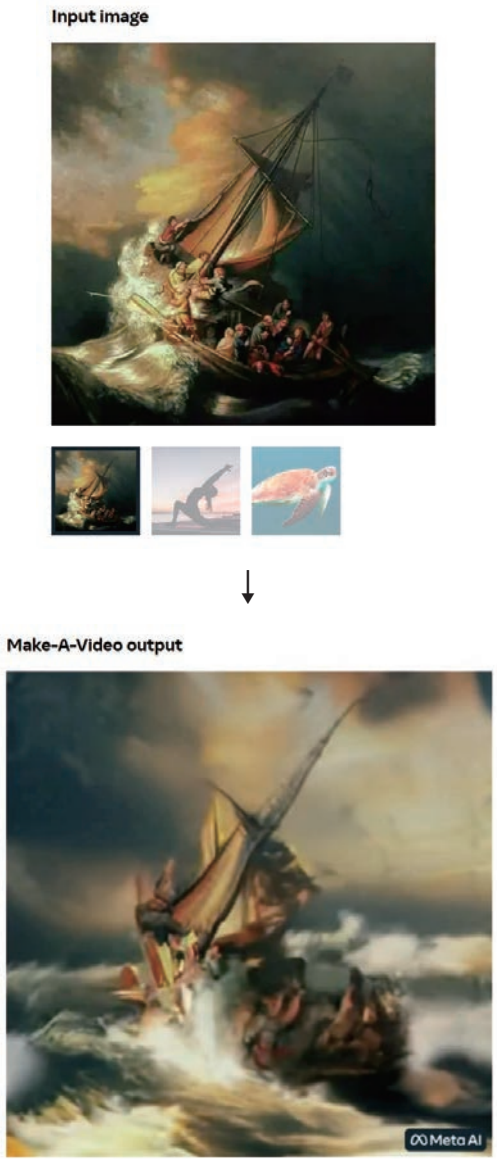


図 9　Make-A-Video（Image to Video）

【ケース 3】動画から様々な動画を生成する

1 枚の動画から動きの異なる色々なバリエーションの動画を生成する（**図 10**）。

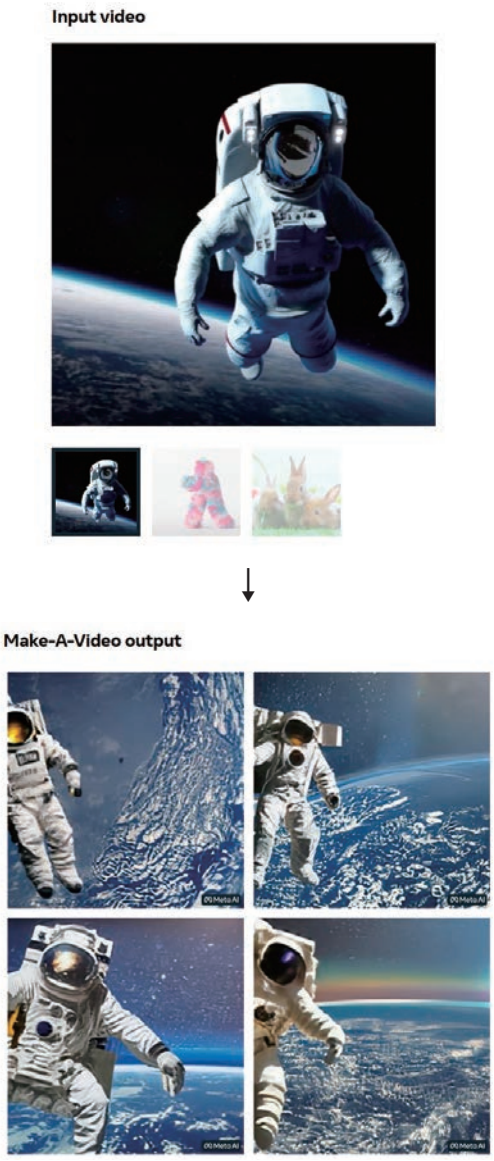


図 10　Make-A-Video（Video to Video）

2.7.　Google「Imagen Video」

https://imagen.research.google/video/

Google が 2022 年 10 月 5 日に発表した。テキストから高画質な動画を生成することが出来る（**図 11**）。1,280 x 768 画素の画質を 24fps のフレーム

レートで，128 フレーム分生成できる（約 5.3 秒）。高画質な動画を生成するため，入力テキストがテキストエンコーダーによってテキスト埋め込みに変換される。変換された埋め込みから動画を生成するが，段階的に画質とフレーム数を増やす。

図 11　Imagen Video

2.8.　Zeroscope Text2Video

文字から動画を生成できる拡散モデルが進化し，2023 年 6 月にオープンソースとしてリリースされた。Stable diffusion の動画生成版である「**zeroscope v2**」での動画クオリティは凄いと 2023 年夏頃，映像製作プロの間で話題になった。text-to-video で生成した動画を video-to-video で高画質にアップスケールして映像を修正する（**図 12**）。

576 x 320 モデル :

https://huggingface.co/cerspense/zeroscope_v2_576w

1,024 x 576　モデル :

https://huggingface.co/cerspense/zeroscope_v2_XL

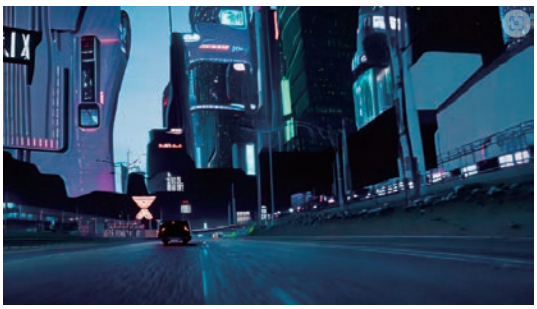


図 12　Zeroscope Text2Video

2.9.　「Wonder Studio」

「**Wonder Studio**」を使うと，撮影した実写の映像から人物を検出して，3D キャラクターに置き換えられる。人物の動きや表情も**モーションキャプチャー**によりアニメーション化され，3D キャラクターに反映される。使用する 3D キャラクターは，デフォルトで用意されるモデルの他，オリジナルのモデルも利用できる（**図 13**）。

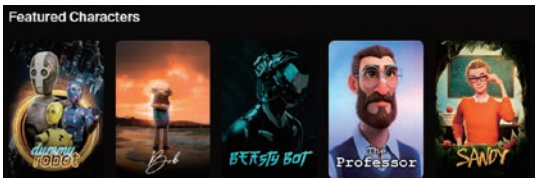


図 13　Wonder Studio

2.10.　Deforum

https://deforum.github.io/

Stable Diffusion の拡張モデルとして，テキストからアニメーションを生成する。ストーリー形式の箇条書き（英語）で与えられたプロンプトから動画生成を行う。生成動画には，チャット GPR で生成したストーリーを設定し，英語に変換したストーリーを Deforum に与える（**図 14**）。

【Deforum チュートリアル動画】

https://www.youtube.com/watch?v=AIjLdiFSEww&t=102s

図 14　Deforum

「Deforum」は，**図 15** のような 10 秒動画に加え，**図 16** のように 15 秒動画を作成できたため，今後は TikTok にも多く生成動画が投稿されることが予想される。筆者（石川）の感想が示す通り作業は難しいため，平易化が実現された 2024 年度以降の段階で一般学生の実習を行いたい。

図 15 「Deforum」10 秒自動生成動画

図 16 「Deforum」15 秒自動生成動画

【筆者（石川）の感想】

やはり 15 秒動画は，動画の変化を楽しみ易い長さで，Runway より変化がはっきりしている点で面白い動画生成 AI であった。しかし，設定や操作は細かく，英語のため，チュートリアル動画がないと操作が難しかった。また，プロンプトが，単語を入力するだけでなく，変化の滑らかさ・度合い（数字で画像の枚数を指定する部分）まで細かく指定できるのは，今までになかった生成の仕方であった。特に，パラパラ漫画みたいな構成・印象を感じた。更に，｛ ｝で文章を締める部分や「,」（カンマ）で文章を終えただけでエラーになる部分などプログラミングのような難しさを強く感じた。

2.11. 動画生成 AI のまとめ

2023 年 10 月時点で動画生成 AI として複数のサービスが公開されているが，一般ユーザーである学生が無料で使えるサービスは「Runway」と「Synthesia（シンセシア）」の 2 つがある。2022 年夏に画像生成 AI がリリースされてから約 1 年で高画質化・高精細化を実現できたことは素晴らしいが，多くは画像生成 AI で得られた知見の応用であり，動画生成 AI 特有の技術的なブレークスルーがあった訳ではない。画像遷移を行っているだけに留まり，映画やドラマのようなストーリー性を持たせることは，2023 年時点ではまだ誰も出来ていない。大規模言語モデル（LLM）を利用した映像への自動的なアノテーションを実現する技術進歩を待っている黎明期にある。しかし，製作側からは生産性向上の方策として期待は大きく，俳優や脚本家からは危惧が大きい。意識差はハリウッドの大規模ストライキにつながった。日本では動画生成 AI の認知がいまだ低く周回遅れ段階にある。大学も理系では研究が進むが，文系では存在すらも知られていない。しかし，「周回遅れ」とされる日本の映像関係者や文系研究者も数年後にはハリウッド・ストライキがなぜ起きたのかを理解できるようになるだろう。そして自分たちが「時代遅れ」であったことを自覚するであろう。

3. バーチャル空間ならではのインタラクション（相互作用）

3.1. フェイシャルキャプチャシステム

Ｖチューバーの顔にリアルタイム描画技術により，表情の細かなニュアンスを反映させる。ライブなど動きのあるシーンでも感情豊かなキャラクターを描き出す。

3.1.1. 「xpression カメラ」

バーチャル空間ならではのインタラクション（相互作用）を学ぶため，「xpression カメラ」を用いた実習を行なった。「**xpression カメラ**」はビデオ会議アプリ上で，自分自身の外見を AI で置き換えることが出来る。写真 1 枚あれば，その人物に変身でき，リアルタイムに表情や身体の動きを反映してコミュニケーションが出来る。例えば，自分のスーツ姿の写真を使えば，寝起きの状態でもオンライン会議に臨むことが出来る。Zoom やグーグル Meet，YouTube など幅広い用途にも対応している。非商用で，既定の 6 種類の顔写真を使用するだけなら無料で利用できる。

名称	URL
xpression カメラ	https://xpressioncamera.com/ https://www.youtube.com/watch?v=HCm0_ZzpBDE https://www.youtube.com/watch?v=_nHB1Z0DGoU&t=50s

【学生の感想①】

xpression カメラを使ってみて口が連動して動くのは予想できたのですが，目線も読み取って見ている方向に動いて想像したより精度が高く驚きました。植田先生が触れられていたように詐欺に使われるなど悪用出来ることも事実であると感じました（**図 17**）。

図 17　xpression カメラ体験（学生①）

【学生の感想②】

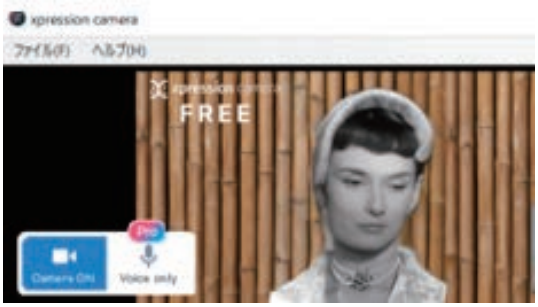


図 18　xpression カメラ体験（学生②）

インストールするだけで簡単に始められ使い易かった。コロナ禍でリモート会議が主流になって来たからこそ登場したアプリであると思いました。時代は進化しているのに，コロナ前に戻そうとする考え方では，ますます時代に取り残されると思いました（**図 18**）。

【学生の感想③】

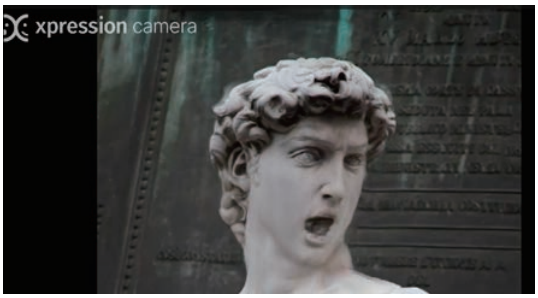


図 19　xpression カメラ体験（学生③）

表情や体の動きを別人に差し替える体験では，別人になった感覚を味わうことが出来た。目線や口の動き，首の傾きを操ることが出来る。自分の顔のままスーツを着たり寝癖を直したりすることが出来るため，直前に起きても，いつもと変わらない状態でリモートワークへの参加が可能になる。芸能人になり情報発信することも可能である（**図 19**）。

【学生の感想④】

モナリザの顔を自分の顔に当てはめました。眉

毛から目，鼻，口など複雑に動かしたのですが，滑らかに動きました。画像をアップロード出来たので，推しの画像をアップロードして自分の顔に当てはめました。推しを身近に感じられました。私のようにジャニーズ推しだけでなく，ゲームのキャラクターの推しにも需要があると思いました（**図 20**）。

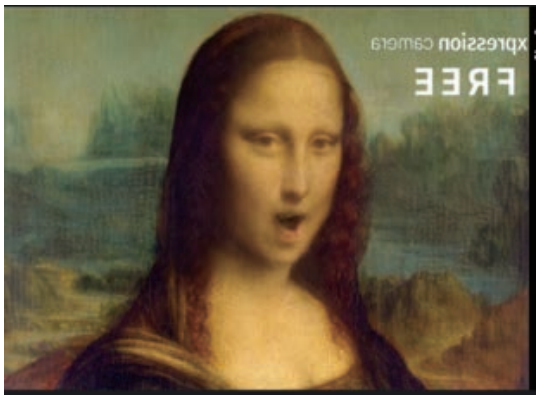


図 20　xpression カメラ体験（学生④）

【学生の感想⑤】

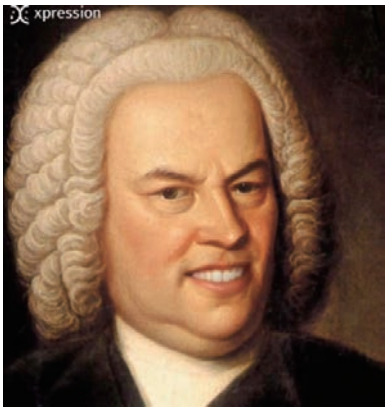


図 21　xpression カメラ体験（学生⑤）

とても凄い技術だと思いました。自分の動作に合わせバッハが生きている動きをしました。偉人だけでなくネットで拾った写真，友達の写真でも使用できました。ボタンを押すと動画を撮影でき，元の画像に自分の声を入れることが出来，面白かった（**図 21**）。石原さとみやスピッツの草野さん，サウシードッグの石原慎也さん，友達の画像を使ってみました。AIを使って友達の誕生日を祝ったり驚かせてみたりと，上手く利用したいです（**図 21**）。

3.2.　リップシンク

3.2.1.　「Diff2Lip」

https://arxiv.org/abs/2308.09716

2023年8月18に提出されたSoumik Mukhopadhyay et.al[(7)]の論文により，「**Diff2Lip**」というAI技術が注目を集めた。俳優の口の動きを，音声に合わせて変化させるAI技術であり，AIリップ・ペインティング・モデルである。映画の外国語への吹き替えがより簡単かつ音声と同期したリアルな口の動きを可能にした。リップ・シンクロナイゼーション Lip Synchronization（lip-sync）のAI技術である。画像生成AIで使われる。Diffusionモデルを活用してノイズを取り除き，新しい口の画像を生成するが，画質，オーディオとの同期，フレーム間の一貫性のために最適化された。Diff2Lipは映画の外国語への吹き替えだけでなく，任意の人間（実人間，キャラクター）や動物がリアルに話す動画を作る作業を単純化した。

4.「動画配信」から「空間配信」〜体験型・空間型クリエイティブ

4.1.　映像コンテンツの革命〜ノンリニア化

2022年まで，放送や映画など映像サービスは「**完パケ**」を基本として来た。プロ用に提供された高額の機材とコンテンツだけでサービスが成立していた。しかし2023年に生成AIが登場したことで状況は一変した。より先端的なサービスを求める階層を吸収する方法を持たないと，視聴者は離れて行く。本項で紹介する最新技術は，オブジェクト（コンテンツ）としてもサービスとしても強力なものであり，重要な資産となり得る。生成AIの応用により，完パケ（リニア）ではない「**ノンリニアコンテンツ**」が急増，映像分野は大変革期を迎えた。「リニア」なコンテンツとは，最初から終わりまで一直線に連続した形で見られ

ることを想定したものであり，映画やテレビドラマが代表例である。「ノンリニア」は「リニア」の逆で，バラバラに断片化しても成り立つコンテンツで，即時性に優れ，検索やリンクであらゆる先に遷移できるネットとは相性が抜群である。生成AIの登場は，従来の映像編集の概念を根本的に変え，映像制作・編集に革命を起こしつつある。従来の大人数でのスタッフによる人力での撮影やスタジオでのローカル処理など「全員集合」型のワークフローは既に「オワコン」になった。

4.2. 「ニューラル場」(NeRF)

「**ニューラル場**」（NeRF）とは，屋外のあるシーンや何かオブジェクトを学習して製作・復元したい場合にモデル化した対象をニューラルネットワークにより「表」を表す技術である（**図22**）。位置を与えることにより，大量の高解像度3Dオブジェクトを生成することが可能となる。3Dオブジェクトを生成できるようになると，リアルなCGを効率的に作ることが出来るため，映像制作，ゲーム，EC，メタバース，機械学習のデータセット作成などの応用が期待される。

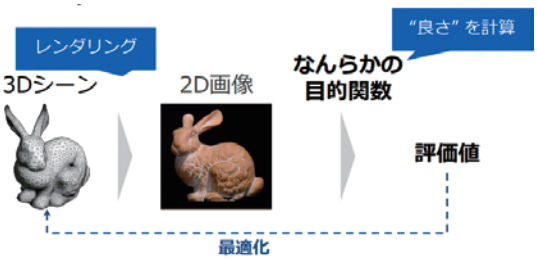


図22　ニューラル場（NeRF）

4.3. 3次元生成AIの実習

平成時代，ユーチューブとスマートフォンが誰にでも動画を作成し配信することを可能とした。生成AIが登場して来てVRと結び付くことにより，誰でも空間を作成し配信することが可能になった。アバターでその場所でしか体験できないコンテンツを共有して楽しめる。

言葉を入れると3Dモデルを生成してくれるAIが登場した。今まで大きな手間が掛かっていたが，手軽になった。画像から3次元モデルを生成する「3次元生成AI」は，ゲームのキャラクターも簡単に作ることが出来るなど活用範囲は幅広い。2022年の画像生成AI（Text-to-Image）モデルの盛り上がりを彷彿させるように，2023年は3次元生成AI（Text-to-3D）モデルが注目された。2022年9月末に発表された「**Dream Fusion**」から急速に発展し，11月には「Magic 3D」や「Latent-NeRF」，12月には「SJC」や「Dream3D」，2023年3月に「**Fantasia 3D**」や「**Text2 Room**」，5月にはOpenAI「**Shap-E**」，「**Text2 NeRF**」，「**Prolific Dreamer**」などが次々と発表され，3次元生成ブームが起こった。本稿執筆時点の2023年10月時点ではどのAIが優勢であるかまだ判断が付かない黎明期にある。私たちは日常生活では文字や音，動画だけでコミュニケーションを取っている訳ではなく，ボディーランゲージ，空間認識，感覚的な部分を用いている。「動画配信」から「空間配信」へニーズがシフトする時代背景を受け，「空間がどうあるべきか」という視点を磨く実習を専門ゼミナールで行った。キャラクターや建物，空間全体のモデリングから，空間内でのキャラクターや物体の動きの定義，VR空間ならではのインタラクションを行った。

4.3.1. Scaniverse（スキャンユニバース）

LiDARセンサーを搭載したiPhoneにScaniverse（スキャンユニバース）アプリをインストールすれば，3Dスキャンは容易である。ゼミナールでは，スマートフォンやタブレットで建物をスキャンし**VR空間**を作成した。3次元（3D）モデリングのスキルを学んだゼミ生は街中の建築物や景色に応用することで，メタバース構築することが出来るようになった。

アプリ名称	URL
Scaniverse	https://scaniverse.com

LiDARスキャンは「部屋の中」を得意とする特徴があるため，ゼミナールでの実習は教室内で実施した。学生が3Dスキャンした画像は**図23,**

図 24 の通りである。

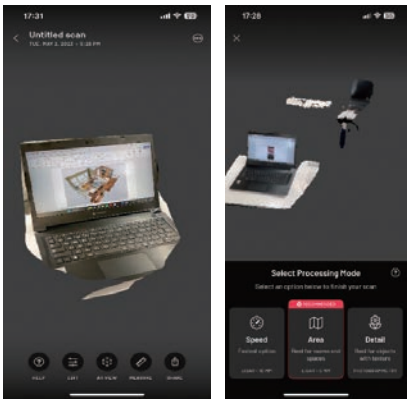


図 23　3D モデルで PC をスキャンした映像を AR 表示

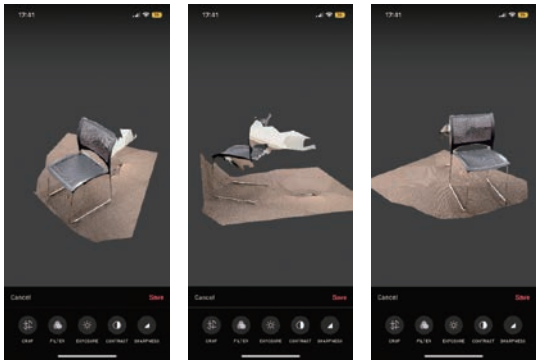


図 24　3D モデルで椅子をスキャンした映像を AR 表示

4.3.2.　VR 空間を作る「WIDAR　3D スキャン」

https://www.widar.io/ja

レーザースキャンする高額な機材を使用することは学生には難しいし，汎用的ではない。スマートフォンの「WIDAR」を使えば，2 次元カメラで複数枚のデジタル画像を撮影して組み合わせ 3D 化できる。360 度取り込むように撮影し，画像を基に 3D のデジタルモデルを作る。スマホのカメラを用意してアプリをインストールすると，チュートリアルで使い方が紹介される。対象物の写真を撮影することにより簡単に AI が 3D データを生成してくれる。20 枚ほど写真を撮ると「OK」の表示が出るため，処理を開始する。データは自動的にサーバーに送信されて 3D 化の処理が行われる。パソコン画面で自在に角度を変えながら細部の動きを確認できる。処理が終わるとデータが完成し，指でぐるぐる回すことが出来る斬新な写真が手に入る。

レーザーを使った場合，光沢があるペットボトルや透明な物体，飲み物の缶などは位置を取り難く 3D スキャンし難いとされるが，**図 25** のように挑戦してもらった。透明な物体については，ぼんやりとして位置を推定して形状を推定する。実習の結果，やはり復元は難しかった。商用の場合，最後は人手でマテリアルを修正して行くことを行う。この他，金属形状の物体，食べ物などは苦手とされる。最初は失敗することもあるが，何度か試しているうちに，うまく撮れるようになる。ゼミナールでは楽しみながら実習してもらい，下記のような感想が得られた。

【学生の感想⑥】

「手元にあったお茶を 3D スキャンしました。40 枚程度の写真を取ることで 3D モデルが簡単に生成できました。多少，ザラつきや背景が混ざってしまいましたが，ここまで簡単に生成できるのは，3D モデルの普及に繋がりそうです。」（**図 25**）

図 25　3D モデルでスキャンした映像を AR 表示

【学生の感想⑦】

「試してみた結果が**図 26** の画像になった。スキャンが雑だったのか，かなりボコボコになった。スペックの高いスマホでスキャンしたら，本格的なものが出来ると感じた」（**図 26**）

図 26　3D モデルでスキャンした映像を AR 表示

レーザーだとなかなか光沢の 3D 復元が難しかったが，「**ニューラル場**」を使うと，オブジェクトを 3D 復元してアセット生成できるようになる。「ニューラル場」を用いた 3D 生成についても「Fantasia3D」（**図 27**），「Text2Room」（**図 28**）のように，いくつかの AI がリリースされている。生成された 3D オブジェクトは形状が分かるようにサイト上で回るように表示される（**図 27**）。

4.3.3.　「Fantasia3D」

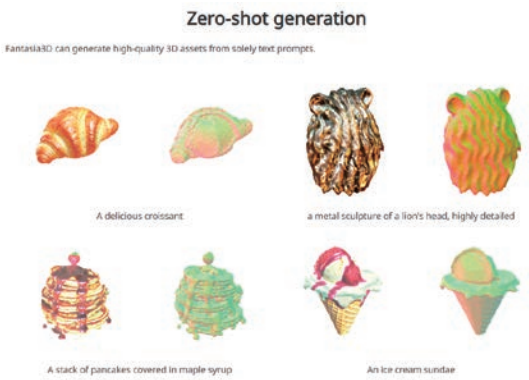


図 27　Fantasia3D

4.3.4.　「Text2Room」

https://lukashoel.github.io/text-to-room/

図 28　Text2Room

4.3.5.　3D モデル自動生成「Point-E」

https://openai.com/research/point-e

OpenAI は Point-E のソースコードを GitHub 上に公開している。**Point-E** も Diffusion モデルを使って**点群**（Point Cloud と呼ばれる 3 次元上の多数の点）から 3D モデルを生成する。従来は時間が掛かった 3D モデル生成を 1〜2 分程度で行う。テキストから画像をで生成，生成画像を使って Diffusion モデルで 3D 点群を生成する。

4.3.6.　3D モデル自動生成「Shap-E」

テキスト（プロンプト）や 2D イメージから「3D データ」を生成する。2D 画像を入れるだけで 3D モデルを生成する。ChatGPT の開発元である OpenAI が公開した「**Shap-E**」は容易に 3D データを生成する。オブジェやおもちゃのようなモノをプロンプトから生成して，3D プリンタで実体化することも日常化できる。

https://huggingface.co/spaces/hysts/Shap-E

https://huggingface.co/papers/2305.02463

Shap-E は，テキストからの生成だけでなく，画像のみから 3D モデルを生成する「Image to 3D」を搭載する。プロンプト入力後に数十秒待つだけで 3D モデルが出力される。OpenAI は色の付いた点を集めて 3D モデルを生成する「**Point-E**」を開発しオープンソースとして提供して来たが，「**Shap-E**」では，多様な角度から撮影した写真から 3D モデルを生成する NeRF（Neural Radiance Fields）が導入され，柔軟な

表現が可能となった。Tex to 3D と Image to 3D があり，実習は両方で 3D モデルを生成した。

【テキスト→画像】

「Plastic bottle containing Japanese tea（日本茶が入ったペットボトル）」というプロンプトを与え 3D 生成した（**図 29**）。

図 29　Shap-E

【2D 画像→ 3D 画像】

犬のぬいぐるみの 2D 画像（左）を与えると 3D（右）生成した（**図 30**）。

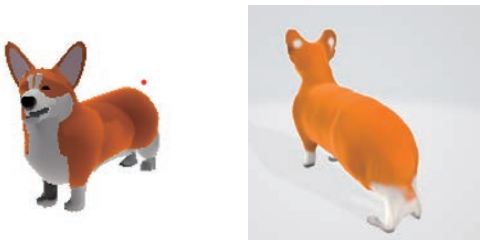


図 30　Shap-E

スニーカーの 2D 画像（左）を与えると 3D（右）生成した（**図 31**）。靴，バッグ，アクセサリーのような立体的な造形があるモノに向くが，人間の身体が入る服は 3D モデリングには向かない。

図 31　Shap-E

生成時間は 15 秒～40 秒である。生成時にオプションを変更できる。シード値（Seed）により，かなり生成にばらつきが出ることが分かった。思ったものが生成されない場合でも，Randomize seed のチェックを ON にしたまま生成を繰り返すと近いものが出て来た。

4.3.7.「Text2 NeRF」

3D 生成については「Text2 NeRF」（**図 32**）の AI ツールがリリースされている。

https://eckertzhang.github.io/Text2NeRF.github.io/

図 32　Text2 NeRF

4.3.8.　LumaAI「Imagine 3D」

https://lumalabs.ai/dashboard/imagine

「NeRF」とは，3 次元シーンを表現するモデルである。微分可能レンダリングにより 3D シーンを最適化する。色々な視点からの 3 次元情報が「**ニューラル場**」で表せる。様々な角度から撮影した画像から視点を自由に動かしたり，3D モデルを作成してくれたりするサービス LumaAI から，文字を入力したら 3D モデルを AI が作成してくれる AI「**Imagine 3D**」が公開された。欄に「chair」と入れると，**図 33** のような 3 次元モデルが生成された。修正や完成品を高める難しさがあるため，納得しないものが生成された場合には，ゼロから作り直した方が良いことが分かった。

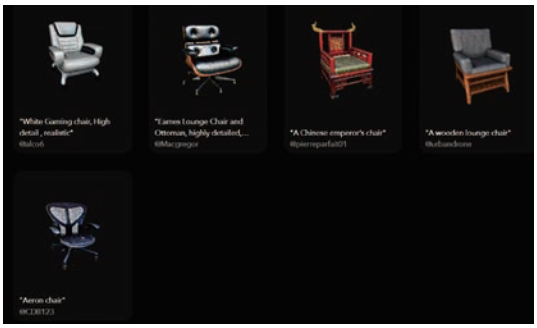


図 33　LumaAI「Imagine 3D」

4.3.9.「Prolific Dreamer」

https://ml.cs.tsinghua.edu.cn/prolificdreamer/

オブジェクトであれば，従来では考えられない高精細 3D モデルが生成される（**図 34**，**図 35**）。

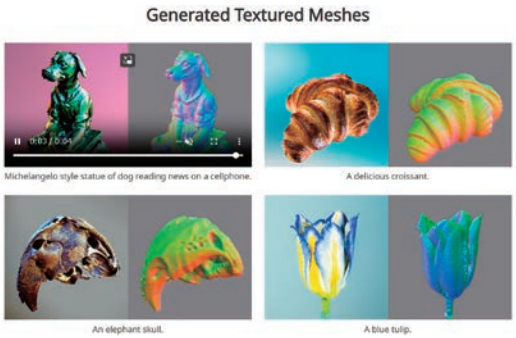


図 34　Prolific Dreamer

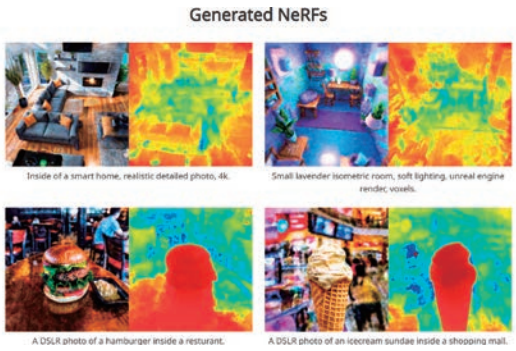


図 35　Prolific Dreamer

4.3.10.　3 次元生成 AI（ニューラル場）のまとめ

画像生成 AI の登場から僅か 1 年で，映像生成と共に 3 次元生成 AI の急速な技術進化を体験する実習となった。3 次元の表現形式である「ポリゴン」を直接扱うことはかなり難しいが，映像生成の「**拡散（diffusion）モデル**」と「NeRF（Neural Radiance Fields）」の技術の組み合わせで急速に発展している。以前は 3D モデラーが 10 時間以上掛けて作業していたが，**3 次元生成 AI（ニューラル場）**の登場により誰もが参加できる世界になった。

靴，バッグ，アクセサリーのような立体造形があるモノに向くことが分かったので，今後，メルカリのようなフリマアプリ市場で応用されると期待される。映像分野でも 3DCG 制作をかなり効率化できることが分かった。3DCG を制作する過程で完成するまでに相当な時間が掛かっていたが，生成 AI を利用することで途中の進捗を素早く他のメンバーと共有することが可能になる。入力画像と 3D 生成モデルでは色味もかなり変わってしまったり，素材感も再現することが難しかったりしたため，2024 年度以降のゼミや演習実習における課題として，調整を上手く出来る方法を模索したい。人間も試したが，アウトラインは再現される一方，顔は怖くなってしまった。人物の 3D 生成についても改善したい。

5.「フルトラッキング」

動画生成 AI の研究分野として注目されているのがモーション生成である。**モーションキャプチャー**[8]とは，人，物の動きを数値化するための計測技術である。3D トラッキング，VR，CG アニメーション制作などの場面におけるモーションデータをリアルタイムに計測する。モーションキャプチャーについては，V チューバーが 6 年前から盛り上がり始め，世界的にはメタバースも注目されている。ハリウッドのアニメや実写では当たり前の技術であったが，AI の技術の進化により低価格化や簡素化が実現し，広く使われる環境が整備された。

一般に，アニメーションやゲームにおけるモーション制作には，複雑な工程を要する。そのため，**モーションキャプチャー**を用いて人間の動きをデータ化し，それをアバターに当てはめる手法

を採る。このようなモーション制作に動画生成AIを活用することで，工程短縮したりコスト削減したり出来る。2Dのイラストを立体的に動かす「Live2D」，ソニーのモバイルキャプチャ装置「mocopi」が普及したため，3Dオブジェクトが急増している。最新の動画生成AIでは，キャラクターの動きを自動生成できるようになっているが，モーションキャプチャーの場合，スマホで自分を撮影しながら体の動きをフルトラッキングする。ショットに応じてそのキャラクターの動きを演じて，それを観察することで，体の動きや勢いがどこに出るのかを知ることが出来る。それを基に1コマ1コマ作ったCGキャラクターを動かす。自分の動きを取り入れているため，自分の魂を吹き込んでいる感覚になる。画面上でCGモデルが動き出した時は嬉しく感じる。

5.1. モーションキャプチャー「mocopi」

ゼミナールでは，同じくソニー「**mocopi**」の実習を行なった。Vチューバーや3Dバーチャルヒューマン，AIアバターが急増する中，実写映像と3DCGを自然に融合してくれる。裸眼でHMD非装着でありながら，**没入感**を得られ，自分の身体を認識できる。身体に6カ所マークを付ければスマホにフルトラッキングしてくれる。わずか6点トラッキングでも，あとはAIが動きを補足してくれるため，自分の身体の動作に合わせ3Dのアバターを動かすことが出来る。加速度センサーを使用する関係上，15分に1回程度の簡易なキャリブレーション（空間補正），30分に1回程度の通常のキャリブレーションをすることが推奨される。特に学生の実習に「mocopi」が適しているのは，**機器の可搬化**である。装置を固定化することは大学のスタジオや教室を占領することになり，大変なコストを要し得策ではない。Vチューバーが音楽ライブを行うためには迅速に展開できる形を考えなければならない。「迅速な展開」は，設置・撤収の容易さだけでなく，キャリブレーション（校正）の容易さも含まれる。

「マシンスペック次第では実写の人物を動かすことが出来るためボカロが音楽を変えたように，フルトラッキングは映像を変えるインパクトを備えた。その割に放送業界の反応が鈍いのがとても気になる」と江口（2023）は指摘する[(9)]。音楽では，AdoやYOASOBIなど，ボカロで曲を作っていた経験を持つアーティストが活躍しており，日頃ボカロ曲を聴く機会が多いZ世代には非常に関心の高い実習となった。フルトラッキングを体験したゼミ生による感想は次の通りである。

【学生の感想⑧】

図36　mocopi体験（学生⑧）

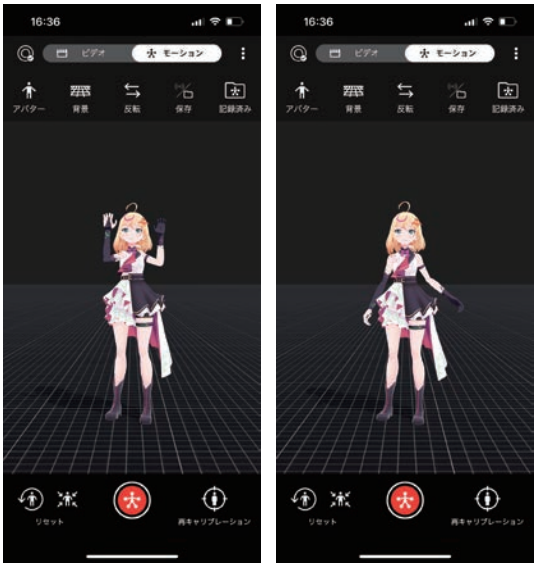


図37　mocopi体験（学生⑧）

後頭部，両腕，腰，両足にセンサーを付ける。小型で装着していても違和感がなかった（**図36**）。セッティングやキャリブレーションも直感

的かつ簡単な印象であった。スマホ上で動いているアバターの姿をビデオ形式で録画したり，モーションをデータとして保存したり出来た。アバターの姿を録画する時も，背景をグリーンバックにして後々の編集加工がやり易くなる機能も搭載されていた。VRM 形式のアバターに対応しており，自分が VRoid で製作したアバターを持ち込むことが出来た。mocopi からスマホまでは遅延が気にならなかった。歩いたり，足を上げたりする動作から，ダンスや正座まで様々な動きをしたが，きちんと動きが取れた（**図 37**）。

【学生の感想⑨】

mocopi を初めて使用したが，6 点のトラッキングをするだけで簡単にキャラクターを動かすことが出来ることに感心した。精密に動かしたいならお金が掛かるが，この値段でここまで出来るのであれば，かなり凄いと思った（**図 38**）。

図 38　mocopi 体験（学生⑨）

【学生の感想⑩】

図 39　mocopi 体験（学生⑩）

スマホとモコピがあれば簡単にアバターを動かすことが出来て驚いた。モコピ自体も 5 万円程で購入できるため，誰でも V チューバーになれる時代になったと感じた（**図 39**）。

【学生の感想⑪】

センサーを付けた部分がキャラクターと連動して同じ動きをしたので驚いた。キャラクターに命を吹き込んだかのように滑らかな動きをした。動きの再現具合がかなり良く，VTuber になった気分になった。身振り手振りを大きくしてみたり，ダンスをしてみたりすることで，キャラクターに躍動感を与えられた。一方，苦手な動きもいくつかあった。横に寝転ぶのが苦手なようであり，寝転ぶと地面にめり込んでしまう現象が起きた（**図 40**）。

図 40 のアバターは，私が VRoid で作成したアバターをそのまま持って来たものである。別アプ

図 40　mocopi 体験（学生⑪）

リで作成したものを，連携させて使えるのは，とても便利である（**図 41**）。


図 41　mocopi 体験（学生⑪）

歩いたり足を上げたりダンスなど基本動作から応用動作まで対応可能だった（**図 41**）。

【学生の感想⑫】

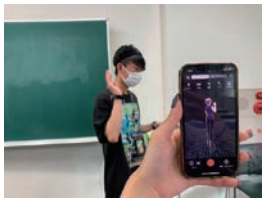

図 42　mocopi 体験（学生⑫）


図 43　mocopi 体験（学生⑩）

機械をスマホと連携することによって人間の動きに合って動くことに驚きました。着けるべき場所に装着して一歩動くだけで認証されるため分かり易かった。私はダンスが出来るので，踊ると，アバターが踊ってくれることが楽しかった。ヘビーユーザーには VR ゴーグルをしたまま睡眠をする「VR 睡眠」という文化があるが，寝転ぶとうまく動きがトラッキングされないのと，そもそも仰向けに寝ると後頭部のセンサーが床に当たり電源がオフになってしまう可能性があるため，工夫を必要とした。バッテリーは 10 時間以上持つため，その点は安心できると思った。植田先生があらかじめ研究室で充電してくれていたため，実習時間にはまったく問題なかった（**図 42，図 43**）。

【学生の感想⑬】

mocopi（モコピ）を体験して設定は簡単に出来た。両手，両足，頭，腰に機器を付け，一歩前に進むだけで設定が完了する。手の動きに機器が付いて来て凄いと思った。早い動きや大胆に大きな動きをしても機器が付いて来てくれた。V チューバーによる 3D 配信は，この機器を使っていると思った。実習を通じて，V チューバーがぐっと身近になった。

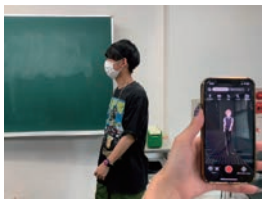


図 44　mocopi 体験（学生⑬）

【学生の感想⑭】

mocopi を付けることによってアバターが自分と全く同じポーズをすることが面白かった。動くとアバターが同じ動きをして細かい動きまで読み取ってくれることが凄いと分かった。頭，腰，手足に着けるだけで，同じように動いて，踊るとアバターも同じ踊りをした。

図 45　mocopi 体験（学生⑭）

5.2.　「Meta Quest 2」と「mocopi」を組み合わせ

Quest 2 と組み合わせて使用する場合，頭と両手のトラッキングは Quest 2 とコントローラーのデータを使用し，腰と両足は mocopi のセンサーのデータを使う。学生が手足を動かすと動きが一度メタバース上を経由して反映されるため，他の学生から見ると動きが遅れる印象であった。Quest 2 とスマートフォンの組み合わせで全身モーションキャプチャー環境が作れてしまう点は非常に魅力的であった。全身の動きが反映されることは，下半身を含めて全身を使って他の人とコミュニケーション出来るということである。

5.2.1.　VR ゴーグル「Meta Quest 2」装着

VR によりリアルタイムで視点を変え映像を見ることが出来る。映像をリアルタイムでレンダリングして VR で視点を変え動画を見ることが出来る。HMD を装着することで，リアルタイムで視点を変えて見ることが出来る。マウスやトラックボールでの観察は，どうしても細かな動きが出来ない。デバイスを顔に当てているからこそ，自然な動きで見る場所を決められる。エンタメのオンライン化を背景に物語や芸術の世界に入り込む「**イマーシブ（没入）**」体験が注目されている。VR で映像を動かすのはカメラではなく，3 つの感覚（視覚，聴覚，触覚）に働き掛け**意識誘導**を行う。AI を利用することでアバターがメタバース上で能力を超える動きをしたり，視覚・聴覚以外の感覚の体験が出来たりするようになった。

【学生の感想⑮】

図 46　Meta Quest 2（学生⑮）

メタクエスト 2 を装着するとゲーム世界に入り込んだ映像が前に広がった。普段ゲームをやらないため操作は慣れなかったが，ワールド内を散策できて良かった（**図 46**）。

【学生の感想⑯】

図 47　Meta Quest 2（学生⑯）

VR でメタバース空間に入って遊んでみた。操作は難しくなかった。自分が製作したメタバース空間に入って自分が空間にいる感覚になり楽しかった。自由に歩き回ったりジャンプしたりと自分の身体を動かさずとも行うことが出来る。今回試した VR のように体を動かさずにコントローラーを用いてアバターを動かす方法は，「歩き回って

障害物にぶつかってけがをする」といったことがないため，安全に楽しむことが出来る（**図 47**）。

【学生の感想⑰】

図 48　Meta Quest 2（学生⑰）

クラスターで作った自分のバーチャル世界に入ることが出来た。身長を設定することが出来，自分の目線の高さで行動することが出来るのが凄いと思った。スマホでワールドを作っていたので，作っている時はとても小さく感じたが，VR ゴーグルを着けてメタバース空間に入ると，自分の見えている世界と同じ大きさになった。メタクエストの設定も慣れてしまえば簡単に設定できる。やる前は難しいと思ったが，それは思い込みでスムーズに進める事が出来た。メタバースと VR の連携を理解できたので，良い実習になった（**図 48**）。

5.2.2.　メタバースでの授業

2023 年 11 月 17 日（金），メタバース「クラスター（cluster）」「ガイアタウン（GAIA TOWN）」で実習を行った（**図 49**，**図 50**，**図 51**）。PC やスマホ操作だけでなく VR ゴーグル操作の学生も多く，学生の新メディアへの順応と習熟の速さを実感した。また VTuber の 2 人「甘野氷さん」「阿和音あまえびちゃん」の 2 人は HTC VIVE のフルトラッキング機能を使っており，学生は滑らかな動きに関心を寄せた。

【学生の感想⑱】

V チューバーの方はやはり動きが滑らかで操作も完璧なのですごいと思いました。トラッカーというものをつければあのような動きになるということを今回の授業で学びました。

【学生の感想⑲】

甘野氷さんの方に参加したのですが，見た目もそうですが，声がとにかくかわいいと思いました。変声機を使って声を変えているとのことなので私も使ってアバターを使った活動を体験してみたいと思いました。

【学生の感想⑳】

プロの方はとても動きがスムーズで移動が早かったです。

【学生の感想㉑】

甘エビさんや甘野氷さんのアバターは凝っていてかわいかったです。VR などを腕や足につけているから動きがなめらかであるのだなと感じました。手招きをしていたら，バイバイをしていたりして凄いなと思い原理を知れて凄いなと感じました。

図 49 「ガイアタウン」講義

図 50 「ガイアタウン」集合写真

図51 「クラスター」集合写真

5.2.3. VRゴーグル装着の実習から得られた知見

VR体験は圧倒的な臨場感と没入感を誇ることが魅力である。しかしVRゴーグルを被って実際にメタバース空間にフルダイブ（完全没入）してこそ得られる感覚で，未体験者に魅力を伝え難い。メタバース空間の魅力を伝えるためには実習してもらうことである。

6. 全米俳優　43年ぶりスト

全米の俳優ら16万人が加入する映画俳優組合が2023年7月14日，43年ぶりに**ストライキ**に踏み切った。ハリウッドでは1万人以上が加盟する全米脚本家組合が2023年5月からストを続けた。AIの活用[10]について，配信企業と俳優・脚本家で意見が分かれた。配信企業が声優の代わりに人工音声を使おうとする動き[11]に対し，組合は仕事を奪われると危機感を強めた[12]。脚本家も生成AIを用いた脚本作成を危惧した[13]。しかし，下手すると自らの首を絞める行為になるリスクもある。全米映画テレビ制作者協会（AMPTP）が譲歩してAI活用に制限を設ける[14]と，非加盟の独立系製作会社がAIを自由に使った作品を作り，結果的にハリウッドが衰退して行く。もう1つは実写作品の「**コスパの悪さ**」である。欧米の人気ドラマの制作費は1話数億円とも言われる一方，日本アニメの制作費は1話数千万円である。日本アニメは1話30分ほどで手軽に見られ，タイパ時代に合っているため，世界中に熱心な支持者を生みファン層を拡大している。配信でファンが増えたため，ネットフリックス，アマゾンPrime，ディズニープラスなど配信サービスはオリジナルアニメを強化している。配信企業はAIを用いて実写作品の制作費を抑えようとする[15]が，アニメへの傾斜を強め実写の衰退を加速させる可能性がある。コンビニエンスストア「セブンイレブン」が百貨店事業「そごう・西武」を切ったように，**ディズニー**はテレビを切って動画とテーマパークに力を入れる。「地上波テレビ」「ケーブル局」「映画」「テーマパーク」を持っていたディズニーはネット動画が主役となった時代に合わせ，米3大ネットワークの「ABC」などテレビ局を売却しようとした。ストライキはテレビ事業への影響が深刻で，世界的な景気後退懸念から広告市場には逆風が吹いたが，ストの影響で広告主に売り込める良質の新コンテンツも少なくなった。

2023年11月10日に賃上げなどで合意し約4か月ぶりにストライキは収束したが，生成AIを使ってバーチャルヒューマンを作った場合，報酬を支払う必要がないため，むしろバーチャルヒューマンの起用を加速する環境を自ら作ってしまった。膨大なオブジェクトが生成され，ユーザーが自由に視点を選択できると，カメラワークが不要となる。結果，カメラマンの仕事が失われるが，VRではユーザー視点がカメラとなるため，テクノロジーの発達による部分である。日本でこのような問題が生じていないのは，動画生成AIの取り組みが遅れているだけであり，同様のことは起きる。映像製作費が高騰する中，生成AIの利活用は避けて通れない。むしろ長期的視野に立ち，生成AIを駆使して効率的に新たなIP（知的財産）を生み出し続けることが大事である。

テクノロジーは敵か味方か。産業革命に沸く19世紀の大英帝国では，失業を恐れた労働者（織工）が機械を打ち壊す「**ラッダイト運動**」が起きた（**図52**）。政府は軍を動員して暴動を鎮め，国力の礎を築いた。米国人アーティストのロス・グッドウィン氏は脚本家たちのストライキを，産業革命期に英国で起きたラッダイト運動に例えた。19世紀初め，産業用機械という新たなテクノロジーに雇用を奪われることを恐れた労働者が，綿織物などの機械を壊して回った。だが，

技術の進展で人々は単純労働ら解放され，より創造的な仕事が生まれた。五十嵐（2023）は「AIは脚本の書き方だけでなく，私たちの考え方も変えていく，私たちの脳を補完するようなものだ。この新しい技術を活用できないことを心配すべきだ」と言う(16)。英経済誌「The Economist」は「このままでは映像労働者はラッダイト運動の参加者のように**歴史に取り残されてしまう**」と指摘する。マサチューセッツ工科大学（MIT）のダロン・アセモグル教授とサイモン・ジョンソン教授「Power and Progress（力と進歩）」は，過去1,000年の歴史を調査し，新技術が生活の向上をもたらすのは，技術がコスト削減だけでなく雇用を生み出すと結論付けた。

図52　ラッダイト運動

作家の**ニーナ・シック**は，2025年までにハリウッド映画の90%においてAIが生成するようになると予測する。**ロバート・デニーロ**，**ブルース・ウィルス**，**ハリソン・フォード**などベテラン有名俳優はAIを使って若い年代の役を演じるようになっている。近年リリースされたハリウッド映画では当たり前に使用されており，過去に出演した映像を基に，若いころの彼らの姿が再現されている。本人の許諾が取れていれば問題はない。**トム・ハンクス**は自らの死後もAIを使って新作映画に出演することに意欲的である。

7. 将来の映像制作現場

AI技術は映像技術の現場で既に大いに活用されており，今後様々な分野で映像業界に劇的な変化を及ぼすのは確実である。映像制作「ザ・シミュレーション」は米国のテレビ番組「サウスパーク」のデータを基に番組の新作エピソードをAIが自動的に作り出す仕組みを公開した。ユーザーから与えられた大まかなストーリーに基づきチャットGPTのキャラクターが架空の街で様々な物語を展開する。

(1) 脚本生成：AIは物語や脚本を生成することが出来る。実現すれば，脚本家の仕事が減少する可能性がある。

(2) 映像編集：AIを使用すると，映像の編集や音楽の選択を最適化することが出来る。編集マンや音響デザイナーの役割が大幅に変わるか，不要になる。

(3) 視覚効果（VFX）：AIは視覚効果を自動で生成することも出来る。これにより，VFXアーティストの必要性が低下する恐れがある。

(4) 映画配役：AIはキャスティングの選択に使用される。キャスティングディレクターの役割を変える可能性がある。日本では忖度がなくなり実力主義になる期待がある。

(5) 音楽作成：AIは映画やテレビ番組のためのオリジナルの音楽を作成することが出来る。作曲家や音楽プロデューサーの仕事の性質や需要が変わる。

(6) 予測分析：AIは映画のヒット予測や観客の反応を予測できる。マーケティングやプロモーションの戦略が変わる。

(7) ディープフェイク技術：実際の俳優を使用せずに，過去の映画のシーンを再現したり，新しいシーンを作成したり出来る。俳優や監督の役割に影響を及ぼす可能性がある。

ハリウッドの労働者たちが感じる不安や懸念は十分理解できるが，AI技術の進化は避けられないものであり，伴って職種の変化を考慮すること

が重要である。「10 年後の映像制作現場を描け」とプロンプトに入力して「Stable DIfffusion」最新版が生成したイメージが**図 53** の通りである。映像制作現場では全作業を AI が担うようになることを示唆する未来図である。

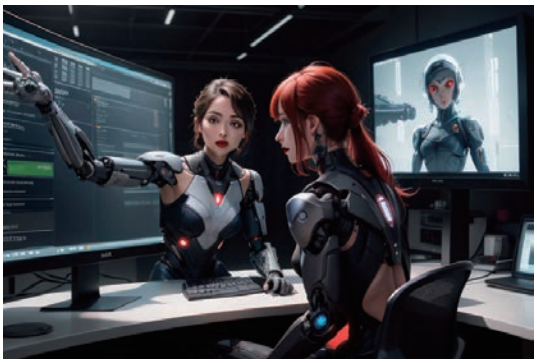


図 53　将来の映像製作現場
出所：相武 AI with Stable Diffusion XL

7.1.　JimakuAI for 日本語

https://jimaku.ai/jp

「**JimakuAI**」は 2023 年 9 月 4 日，AI を活用して日本語字幕が生成できる「**JimakuAI for 日本語**」の提供を開始した。手作業だった動画の文字起こしから翻訳，字幕生成に至るまで自動化でき，コスト削減や省力化に貢献する。日本語音声からの日本語字幕作成のサービスも可能である。**図 54** は，日本語音声から英語字幕を生成した事例である。最近，俳優のアル・パチーノや歌手テイラー・スウィフトが日本語を話す CM が流れるが，AI がそのままの声色で翻訳を可能にした。

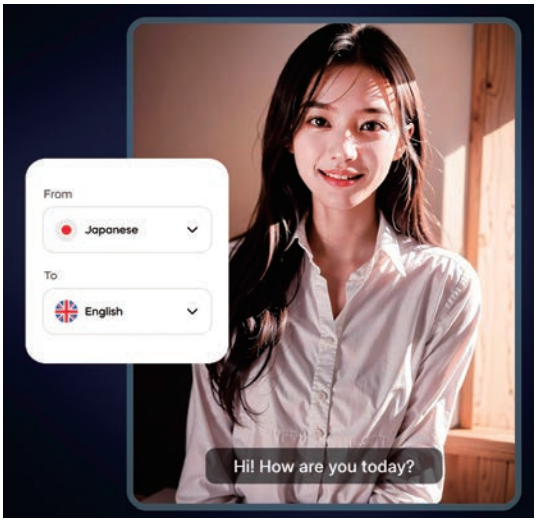


図 54　JimakuAI for 日本語

8. まとめ～「動画配信」から「空間配信」へ

「動画配信」から「空間配信」へと時代ニーズがシフトして，映像制作・編集の手法が大きく変革する中で，学生が特に興味を持って取り組んでくれた「実習」が，自分が製作したアバターを同じく自分で製作した 3 次元空間（ワールド）で，自身の身体の動きをフルトラッキングして動かすことである。実習の結果，オリジナルのアバターを作成し，V チューバーのように人格を持つタレントとして SNS や媒体への露出をプロデュース出来るようになった。V チューバーは，メタバースのようなバーチャル空間を舞台にした参加型の体験を演出する。ゼミナールでは，空間（ワールド）やキャラクター（アバター），それらの動き（フルトラッキング），相互作用（フルダイブ）など空間配信をデザインする能力を培った。ChatGPT が登場した 2022 年は文章を書くためのアシスタントだったが，2023 年は動画や 3D モデルを生成してくれる存在になった。何時間も掛かる大変な作業を自動で出来るようになった。今の映像制作現場は入れ替え時であり，新しい形の制作手法が出来つつある。

ここまで，動画生成を技術や演出という観点で，専門ゼミや演習実習における現場からアイデアや事例を紹介した。AI の進化で加速するデジタル革命の中，AI 教育を最優先とする米大学の存在は圧倒的であり，2020 年代の映像教育の世界市場を席捲しよう。各国と比べると，日本の映像市場の変化はまだ小さく，構造転換は途上との見方が一般的である。ハリウッド・ストライキのように，「昭和」を引きずる日本の映像制作分野も何れ周回遅れで AI 時代に突入する。シュンペーターが唱えた「**創造的破壊**」の波をいつまでも食い止めておくことは不可能である。労働組合が

組合員の生活を守るためには進化する技術に抵抗するのではなく，それを活用して共存して行く必要がある。新技術を導入したディズニーやアマゾンなど企業へ怒りをぶつけるのではなく，労働者の活躍の場を増やし，誰もが新技術の恩恵に預かれるようにすることが望ましい。昭和世代から受け継がれて来た古き悪しき前近代的な業界風土[17]がいよいよ通用しなくなって来た[18]。業界は時代の変化の音に耳を塞いで来たが，社会環境が変わって結果的に時代錯誤に陥った。2022年にはテレビ局（TBS）OBの女子大教授によるゼミ生に対する性的暴行事件が起き，業界内で培った人格（人権意識の低さ）や幼児性（閉鎖的・高圧的）[19]は業界を代えても払拭できず，越境させることを示した。「性加害やセクハラを許さない」という認識を業界全体で共有する契機にしなければならない。今後，「記者クラブ」など時代錯誤的なシステムに依存する恥部も続々とあらわになるだろう。今まで業界が「伝統」として続けて来たこともすべて「ムラ社会型不祥事」にあたる可能性がある。コンプライアンスの意識が高まって来た現代において不適切な環境を黙認・放置していた「過去の悪行」が発覚したら，組織の根底が揺るがされる企業は少なくない。

このような悪循環の中で，**生成AI**は，業界内の性加害やパワハラ・セクハラ・モラハラ[20]を知りながら見過ごしていた業界人をゼロクリアして，自浄できない業界をクリーンに生まれ変わらせられる「解体的出直し」「**イノベーション**」になり得る。「朝三暮四」（見せかけだけの対策）では変われない。変革を望まず古い「全員集合」文化や不透明な慣習を温存して来た昭和世代[21]よりも未経験の令和世代を選ぶ方が業界は健全に「成長」できる。もちろんここで言う「令和世代」とは，現在の業界や「全員集合」ワークスタイルに憧れを抱く社畜タイプ[22]（同じ穴のムジナ）を意味しない。

時代の転換は次の主役を生み育てる土壌になる。コンテンツで若者の心を掴むためには，若い人の感性を重視すべきである。「**エモい**」「**チルい**」という言葉を日常的に使っているかどうかが基準となる。その感覚が分からない，腹落ちしない人は制作に口を出さない方が良い。動画や3Dを生成するAIツールは日本の大学教育にとって前進になり，映像教育も進化して行く必要がある。動画生成AIを活用して映像オブジェクトを製作するスピードが上がれば，コストが下がり，学生がより多くのコンテンツを発表できるようになる。国費が入る大学は教育資源を最新の社会需要に合わせて提供する必要がある。神田（2023）は「生成AIによって学び直しより教え直しが重要になった」と言う[23]。今後のAI教育が行われる上で，本稿で紹介した現場の実践的な知見がヒントになれば幸いである。

謝辞

本稿の作成に際して，2022年度および2023年度「卒業研究」「専門ゼミナール」「演習C」「実習C」受講学生の協力を得た。この場を借りてお礼申し上げたい。但し，本稿に関する誤りは筆者に帰属している。なお**肖像権**の関係から，学生本人の顔が映っている実習画像は掲載しないよう配慮した。また，メタバースでの実習については，パーソルマーケティング株式会社，NPO法人バーチャルライツ國武悠人理事長，VTuberの「甘野氷」さん，「阿和音あまえびちゃん」さんにご講義を得た。この場を借りてお礼を申し上げたい。但し，本稿に関する誤りは筆者に帰属している。

参考文献

［1］ Soumik Mukhopadhyay, Saksham Suri, Ravi Teja Gadde, Abhinav Shrivastava（2023），" Diff2Lip: Audio Conditioned Diffusion Models for Lip-Synchronization"，https://arxiv.org/abs/2308.09716
［2］ デジタルレシピ「Generative AIの企業における活用方法の最新事例ご紹介」（2023年2月21日）
［3］ 境治（2023），日本マーケティング協会主催2023年7月28日「メディアはこれからどうなっちゃうか，どうしたらいいか」
https://www.jma2-jp.org/event/seminar/230728

《注》

（1） **OBM（Object based Media），OBB（Object based Broadcast）**
生成AIの進化により，映像分野で最大の変革となったのが，「OBM」と「OBB」である。「**OBM（Object based Media）**」は英国公共放送「BBC」を中

心にして研究されて来たオブジェクトベースのメディア概念であり世界に急速に広まった。「**オブジェクトベース**」とは，従来の放送のようにコンテンツ全体を単一のアセット，造り手側の意思のみに基づいた「**完パケ**」としてだけ配信するのではなく，個々のオブジェクアセットとして提供し，端末側で視聴者ごとにオブジェクトを再構築するものである。OBMを用いて放送を行う場合は「**OBB（Object based Broadcast）**」と言う。OBMは，ユーザーやデバイスの様々な要求に応えるために，様々な方法で組み立てることが出来る新たな映像時代のコンテンツ製作ワーフロークである。「OBM」「OBB」共に，映像の専門家以外にはまだまだ認知度が低いが，映像制作や編集の分野では大きな革命となっているため，ゼミナールや演習実習で体験してもらった。

（2） ガウシアンモデル，StyleGANモデル，拡散モデルの3種類のAI生成モデルが研究されている。

（3） デジタルレシピ「Generative AIの企業における活用方法の最新事例ご紹介」（2023年2月21日）

（4） https://www.jma2-jp.org/event/seminar/230728
境治（2023），日本マーケティング協会主催2023年7月28日「メディアはこれからどうなっちゃうか，どうしたらいいか」

（5） 2023年2月に発表された第1世代の「Gen-1」は，動画をプロンプトに応じて別の動画へと変換する（video to video）サービスだったが，2023年6月に一般にもリリースされたGen-2からは，描いて欲しい場面をテキストプロンプト（入力命令）として入力すると，動画を生成する「text to video」が実現できるようになった。4秒間の非常に短い動画だが，首尾一貫性を保たせて動画を成立させているところが凄い。

（6） Gen-2が得意なのは，顔のアップや波のような自然物のようである。奥行き（Depth）を推定し，簡単なボーンを付け，どのように動かすのか決めている。

（7） Soumik Mukhopadhyay, Saksham Suri, Ravi Teja Gadde, Abhinav Shrivastava（2023），”Diff2Lip: Audio Conditioned Diffusion Models for Lip-Synchronization”，https://arxiv.org/abs/2308.09716

（8） モーションキャプチャーは赤外線カメラで専用スーツを着た俳優の動きを捕捉して，リアルタイムで3次元のCGを作る技術である。人間らしい自然な動きを演出できる他，ゼロからCGを作成するより時間を短縮できるため，近年はゲーム業界で活用が進んでいる。

（9） 江口靖二（2023）「NAB Show100周年，本物のゲームチェンジャーが現れる前に」，NAB Show 2023現地レポート（2023.5.10）

（10） ディズニープラスが2023年6月21日から配信を始めたマーベルの新作ドラマシリーズ「シークレット・イノベーション」のオープニングクレジットには，生成AIを使ったスクラル人が登場する。肌の緑色が気持ち悪い表現となっており，制作担当者はエイリアンによる擬態を表現したかったと説明した。川上一郎（2023），「月刊フルデジタル・イノベーション（2023.7）」22p.

（11） 「AI俳優」の基になるデータを，人間の俳優から集めるためのオーディションが増えている。俳優が自分の姿を360度スキャンされて，そのデータを使って企業が3Dモデルを作成し，AI俳優として映画に使う。以前からも，著作権がないフリー素材としての写真を集めた「ストックフォト」のためのオーディションが存在し，駆け出しの俳優はそれで稼いでいる。しかしAIデータになると，話の内容も合成音声で生成される。俳優の見た目であるのに，俳優が言っていないことを話しているように使われる。ディープフェイク的な使われ方に対し，製作者が提示したのは半日から1日分のギャラだった。それだけの報酬で，企業は永遠にデータを所持して使い続けられる。2023年8月7日付日経産業新聞13面

（12） 南カルフォルニア大学のジョナサン・タプリン教授は「制作会社がチャットGPTのような技術を独自に開発し，これにマーベルから出版される人気マンガの新作の脚本を書かせるとする。その際，既に保有しているデジタル画像を使えば脚本は6か月ではなく3週間で完成する」と指摘する。2023年8月4日付日本経済新聞7面

（13） 既に一部の契約書には俳優の容姿をコピーしデジタル化したものを数年間使えるとする条項がある。制作会社は「AIはハラスメントで問題を起こさないし，賃上げ要求も病気や死亡もしない」と言う。

（14） AMPTPは「生成AIの作品は文芸的著作物とはみなさない。脚本の一部で生成AIが作る素材を基にしても，脚本家の報酬やクレジットが不利益を被ることはない」と提案した。

（15） AIの活用が大きな利益をもたらす可能性がある。AIは大量の仕事を迅速に低コストでこなすことが出来るため，視聴者は今より良い作品をはるかに安い価格で見ることが出来るようになるかもしれない。事業者の売り上げも伸びるであろう。ただそれが業界全体の底上げにつながるのか，労働者が仕事を失う一方で企業幹部と投資家だけが甘い汁を吸うことになるかは未知数である。「NEWSWEEK（2023.6.20）」44p.

（16） 五十嵐大介（2023），2023年5月27日付朝日新聞2面

（17） ジャニー喜多川氏の性加害で，テレビ局内で行われた可能性のある場としてNHK，テレビ朝日，TBSが挙げられた。当初，ジャニー喜多川氏の「個人犯罪」であったのが，ジャニー氏の性加害をテレビ業界全体で黙認・隠蔽していた「組織犯罪」へと事件の性質が変わった。

（18） 映画監督の西川美和（2023）は「映画業界は怒鳴ったり追い詰めたりと古い体質が残っています。お金も時間もない中でみんなで歯を食いしばって，やって行くという変わらないノスタルジーを大人たちは求めている」と指摘した（2023年8月24日付朝日新聞15面）。厚生労働省は2023年10月，男女640人を対象にしたアンケート結果を公表，俳優・スタントマンを受けた経験では，セクハラ被害を受けた人は20.4%であり，最も多い被害としては，「性的関係を迫られた」で11.1%だった。この他，「仕事の関係者に必要以上に体を触られた」（10.2%），「恥ずかしいと感じるほどの体の露出をさせられた」（9.3%），「羞恥心を感じる性的な実演をしなければならない」（8.3%）と続いた（2023年版厚生労働省「過労死等防止対策白書」，2023年9月14日付朝日新聞9面）。俳優や音楽家らの活動を支援する「日本芸能従事者協会」は2023年2月

～5月，俳優やタレントなど245人を対象にしたアンケートを実施，「仕事中にハラスメントを受けたことがありますか」との質問に「はい」と答えたのは61.2%，「見聞きした」と答えたのは21.5%にのぼった。2022年実施した別のアンケートでは。自由記述の欄に「仕事が欲しいなら従え，とレイプされた」「事務所に所属している際，脱ぐ仕事ばかり勧められた」などと記されていた（2023年9月16日付朝日新聞32面）。

(19) 「先輩には礼儀正しく媚び，後輩にはパワハラ」という典型的な体育会体質とされる。

(20) テレビメディア，新聞や出版社はジャニーズ事務所，宝塚歌劇団，東北楽天ゴールデンイーグルスなど特殊な異境の中での問題を，取材していれば誰もが薄々知っていたのに，伝えるべきを伝えず聖域化したため，おびただしい数の被害者が生み出された。芸能界，マスコミ界，スポーツ界に限らない。世の中にはやっかみもあれば，マウントを取りたがる人もいる。いじめ，いびり，暴力は摘んでも摘んでも油断すると直ぐに芽を出す。

(21) ジャニーズ問題は事件全体の規模だけ考えても，戦後最悪の性加害事件である。この事件が大きく取り沙汰されたのは，外国メディアの報道が発端である。国連人権理事会の動きもあった。事務所，ひいてはマスコミ業界全体の構造的な改革が必要である。事件の詳細を踏まえると，今まで業界に関与していた人たちは改革して行く力に期待できそうにない。全業界人を一掃するくらいの思い切りが必要である。教育界や法曹界を含め他業界に移った人たちが大勢になびく「手のひら返し」をしているが，もちろん，このような人たちが人権を尊重する責任を果たせるに値するのかどうかを見極めることも肝要である。

(22) マスコミの教科書は，マスメディアで働く者は，あらゆる権力や権威から距離を置き，徹底して**在野の存在**であることを求められる，と教える。しかし利己（採用や出世）のために既存権力や権威と馴れる学生が多い。大学の新聞部や放送部は運動部の活躍や行事を紹介する広報と捉えられる。社会が抱える環境問題や格差・差別問題に切り込むジャーナリズムの視座に欠ける。存在するのは業界への就職活動でガクチカとしてアピールする「利己」と大学からお金をもらい易いという「利己」である。マスメディアは受信料，購読料，スポンサー料，そして何よりも，名もない多くの視聴者，読者に支えられている。記事作成や機材の操作以前に真っ先に叩き込むべきは視聴者，読者の存在とコンプライアンスである。一橋大学教授の楠木健（2023）は「仕事は自分以外の誰かのためにするもの。その結果として自分の利得が発生する。この順序が逆になるとヘンなことになる」と指摘する。大人の世界では，東京五輪・談合，ビッグモーターやダイハツ工業による不正，自民党裏金など，多くの組織でヘンなことが起きた。「他者の不在」が問題の本質である。

(23) 神田昌典（2023），2023年3月27日付日経MJ3面