



PRODUÇÃO DE AUDIOVISUAIS COM MÍDIA DIGITAL E INTELIGÊNCIA ARTIFICIAL

Ana Priscila Cayres de Oliveira¹
Richard Perassi Luiz de Sousa²

***Abstract:** There are disruptive technological-digital transformations resulting from the potentialities and uses of artificial intelligence (AI). The popularization of access to digital resources with AI has also led to the loss of exclusivity for professionals working with audiovisual media. In light of this, this article presents partial results of a qualitative and exploratory-descriptive research on digital AI platforms with online resources for audiovisual production. A significant change in the production of audiovisual material is observed, driven by the increasing accessibility of AI tools, even for amateur users. However, it is essential to explore how AI is influencing the creation and dissemination of audiovisual content, taking into account its implications for the narratives and meanings produced.*

Keywords: audiovisual media; digital technology; narrativity; artificial intelligence.

***Resumen:** Existen transformaciones tecnológicas y digitales disruptivas derivadas de las potencialidades y usos de la inteligencia artificial (IA). La popularización del acceso a los recursos digitales con IA también ha resultado en la pérdida de la exclusividad de los profesionales que trabajan con medios audiovisuales. Ante lo expuesto, en este artículo se presentan resultados parciales de una investigación cualitativa y exploratorio-descriptiva sobre plataformas digitales de IA con recursos en línea para la producción de audiovisuales. Se observa un cambio significativo en la producción de material audiovisual, impulsado por la creciente accesibilidad de las herramientas de IA, incluso para usuarios amateurs. Sin embargo, es esencial explorar cómo la IA está influyendo en la creación y la difusión de contenido audiovisual, teniendo en cuenta sus implicaciones para las narrativas y los significados producidos.*

Palabras clave: medios audiovisuales; tecnología digital; narratividad; inteligencia artificial.

¹Programa de Pós-Graduação em Design - Universidade Federal de Santa Catarina (UFSC) Florianópolis - Brasil. ORCID: <https://orcid.org/0009-0001-7918-2008> email: anapco@gmail.com

²Programa de Pós-Graduação em Design - Universidade Federal de Santa Catarina (UFSC) Florianópolis - Brasil. ORCID: <https://orcid.org/0000-0003-0696-4110> email: richard.perassi@gmail.com



Resumo: Há transformações tecnológico-digitais disruptivas e decorrentes de potencialidades e usos de inteligência artificial (IA). A popularização do acesso aos recursos digitais com IA resultou também na perda da exclusividade de profissionais que trabalham com mídia audiovisual. Diante do exposto, neste artigo são apresentados resultados parciais de uma pesquisa qualitativa e exploratório-descritiva sobre plataformas digitais de IA com recursos online para produção de audiovisuais. Observa-se uma mudança significativa na produção de material audiovisual, impulsionada pela crescente acessibilidade das ferramentas de IA, até mesmo para usuários amadores. No entanto, é essencial explorar como a IA está influenciando a criação e a disseminação de conteúdo audiovisual, levando em consideração suas implicações para as narrativas e os significados produzidos.

Palavras-chave: mídia audiovisual; tecnologia digital; narratividade; inteligência artificial.

1 INTRODUÇÃO

Neste artigo, apresenta-se um estudo sobre narratividades e potencialidades expressivo-comunicativas de sistemas ou plataformas digitais com inteligência artificial (IA) na produção de mensagens audiovisuais digitais. Para isso, foi realizada uma pesquisa qualitativo-descritiva, com etapas exploratória, documental, teórico-bibliográfica e descritivo-interpretativa. Na etapa exploratória, foram identificadas fontes teóricas e documentais adotadas neste estudo. Entre outras coisas, isso resultou na seleção de uma mostra representativa de plataformas ou sistemas digitais, como mídia de acesso online que, com diferentes graus de autonomia, produzem mensagens audiovisuais.

Considera-se aqui que “mídia audiovisual” é qualquer suporte ou sistema de recepção, conservação, produção, edição ou comunicação das informações que são compostas com imagens e sons integrados. A mídia audiovisual, digital e de acesso online, é produzida para comunicar mensagens compostas, com sons, palavras e imagens dinâmicas, às pessoas usuárias dos sistemas digitais de acesso à rede online. Portanto, o objetivo geral é descrever uma mostra dos recursos de produção audiovisual, que é ofertada em plataformas ou sistemas digitais de acesso online.

O interesse que motivou este estudo recaiu sobre plataformas ou sistemas mediadores de produção e comunicação de mensagens audiovisuais que, comumente, são disponibilizadas ou transmitidas em rede online. Aqui são apresentadas características e considerações sobre a



eficiência das plataformas digitais na produção de audiovisuais. Mas, isso é decorrente de sínteses das opiniões de pessoas ou referências técnicas encontradas durante a etapa exploratória da pesquisa. A proposta deste estudo, entretanto, não é avaliar, qualificar ou quantificar as propriedades técnicas ou os graus de eficiência das plataformas ou sistemas digitais aqui descritos.

De maneira geral, entretanto, há mensagens com diferentes linguagens, formatos e conteúdo, incluindo as de curta ou longa duração. Basicamente, há imagens gráficas ou fotográficas em diversos estilos, também, há vários tipos de sons vocais e verbais produzindo diversos efeitos e discursos informativos ou poéticos. Além disso, imagens ou sons são produzidos com diferentes recursos de expressão, edição, movimento, animação e mediação ou remediação.

2 CATEGORIAS E TIPOS DE MENSAGENS AUDIOVISUAIS

A partir de matrizes artesanais ou digitais que, no final, são reunidas e editadas como mensagens audiovisuais, com maior ou menor autonomia, os recursos digitais participam da produção das mais diversas categorias e tipos de mensagens audiovisuais. Com relação às caracterizações das mensagens, entre outras possibilidades, essas podem ser categorizadas como: (1) informativo-noticiosas, (2) instrucionais e didático-pedagógicas ou (3) narrativas documentais ou fantasiosas, decorrentes de diversas linguagens e recursos poético-artísticos. Basicamente, também podem ser indicados cinco tipos de textos ou discursos em diferentes linguagens: descritivo, narrativo, dissertativo, expositivo e injuntivo (Travaglia, 2007).

No texto descritivo, busca-se caracterizar pessoas, objetos, lugares ou fenômenos de maneira detalhada, focando em aspectos sensoriais e qualitativos. Utiliza-se de adjetivos, advérbios e construções que proporcionam ao leitor uma imagem vívida do objeto descrito, explorando seus atributos visuais, auditivos, olfativos, gustativos e táteis. Esse tipo de texto busca transmitir uma impressão completa e rica de detalhes para criar uma experiência sensorial.

No texto narrativo, relata-se eventos e ações ocorridas em uma sequência temporal, geralmente envolvendo personagens em um determinado contexto espacial. A estrutura narrativa inclui elementos como enredo, personagens, ambiente, conflito e clímax, organizados de forma a construir uma história coesa e envolvente. O foco está em contar uma



história, podendo ser real ou fictícia, de maneira que o leitor acompanhe o desenvolvimento dos acontecimentos ao longo do tempo.

No texto dissertativo, propõe-se expor, analisar e discutir ideias ou pontos de vista sobre um determinado tema. Caracteriza-se por uma estrutura lógica e argumentativa, onde o autor apresenta uma tese, desenvolve argumentos que a sustentem e conclui com uma síntese ou reflexão sobre o assunto abordado. Este tipo de texto é frequentemente utilizado em contextos acadêmicos e científicos, visando à persuasão e à construção de um raciocínio crítico.

No texto expositivo, as finalidades são informar e explicar de maneira clara e objetiva um determinado assunto, conceito ou processo. A estrutura desse tipo de texto é organizada em introdução, desenvolvimento e conclusão, onde o autor apresenta informações, exemplos, definições e comparações para elucidar o tema abordado. O foco é fornecer conhecimento ao leitor, sem necessariamente buscar persuadir ou argumentar.

No texto injuntivo, procura-se orientar, instruir ou comandar o leitor a realizar determinadas ações ou seguir procedimentos específicos. Esse tipo de texto é caracterizado por uma linguagem direta e imperativa, frequentemente encontrado em manuais de instrução, receitas, regulamentos e tutoriais. O objetivo é guiar o comportamento do leitor de forma clara e precisa, garantindo que as instruções sejam seguidas corretamente.

Os tipos de textos aqui apresentados podem ser coerentemente relacionados com as categorias anteriormente apresentadas, ou seja, as mensagens podem ser informativo-noticiosas (textos expositivos), instrucionais (textos injuntivos) e didático-pedagógicas (textos dissertativos) e (3) narrativas (textos descritivo-narrativos) documentais (documentários) ou fantasiosas (ficção, poesia, canção, música), decorrentes de diversas linguagens e recursos poético-artísticos.

3 PRODUÇÃO E COMUNICAÇÃO AUDIOVISUAL E NARRATIVIDADE

Cronologicamente, milênios antes do desenvolvimento da tecnologia e das linguagens eletrônico-digitais começaram a ser produzidos e foram constantemente evoluídos os recursos e sistemas de produção e comunicação. Primeiramente, foram mais desenvolvidos os recursos produtores e comunicativos de imagens e, mais recentemente, os sonoros. Portanto, os meios de produção e comunicação audiovisual também foram evoluindo ao longo do tempo, em função dos avanços tecnológicos. Segundo Côrtes (2003, p. 32), “a linguagem audiovisual



constrói continuamente suas características, transformando-se à medida que novas formas de captação e registro de sons e imagens vão sendo descobertos/criados”.

Independentemente de seu tipo, gênero ou categoria, os textos ou discursos podem e devem ser interpretados, porque dispõem de narratividade, como um potencial produtivo de sentidos e significados. Isso permite que, além de espectadoras, as pessoas também assumam o papel de intérpretes ou leitoras. No geral, as interpretações ou leituras podem ser estritamente subjetivas e pessoais. Mas, também podem ser produzidas de maneira coerente com a cultura vigente ou com o contexto sociocultural do tempo em que foram produzidas. Quando isso ocorre, além de servir para a pessoa intérprete, a interpretação também será coerente e válida para outras pessoas com capacidades perceptivas e conhecimentos semelhantes.

Diante disso, quem se interessa em atuar como intérprete ou profissional de produção de conteúdo e mensagens, rapidamente, descobre a necessidade de obter conhecimentos sobre: (I) as características perceptivas físico-psicológicas e (II) a respeito do repertório sociocultural das pessoas que, potencialmente, constituem seu público de interesse. Tais conhecimentos podem ser obtidos: (1) de modo lógico-sistemático, com pesquisa de dados e informações; (2) de maneira vivencial e intuitiva, como auto-observação e percepção atenta (consciente ou não) do contexto sociocultural; também, (3) pode ser obtida com uma mescla eficiente das duas possibilidades anteriores. Assim, a lógica complementa a intuição e vice-versa.

Além de conhecer o perfil perceptivo e o contexto sociocultural do público de interesse, para atuarem como intérpretes ou profissionais de produção de conteúdo e mensagens, as pessoas necessitam dispor e saber usar eficientemente os recursos de produção, mediação e comunicação de conteúdo. Por exemplo, um professor de cursos online ou um influenciador digital deve obter um bom uso dos recursos tecnológicos de produção, edição e comunicação audiovisual, além de desenvolver modos eficientes de comunicação vocal-corporal. Isso implica no uso de tecnologia e também na expressão de sua imagem, voz, gestualidade e outras.

A distinção proposta por Ryan (2005) é que (1) a “narrativa” já é produzida para comunicar uma determinada história ou evento e (2) a “narratividade” decorre da interpretação da pessoa leitora como uma resposta aos estímulos expressivo-conceituais da mensagem. Em suma, a narratividade inspira uma narrativa que é criada pela pessoa leitora em quaisquer tipos de textos, incluindo os textos narrativos. Portanto, compreendendo ou não

o significado proposto na emissão da mensagem, seja narrativa, descritiva, dissertativa, expositiva ou injuntiva, a pessoa receptora e leitora pode obter para si outras interpretações, em função da narratividade do texto.

Em outra publicação, afirmou-se que a narratividade cumpre de maneira específica uma finalidade geral, que é informar e comunicar um conteúdo, como conjunto de ideias ou conceitos (Perassi, 2015). Acrescenta-se ainda que a narrativa é estruturada em diferentes níveis, de acordo com: (1) a expressividade dos elementos da mensagem, (2) sua organização sintática, como palavras, imagens, sons ou sinais reunidos na composição da mensagem, (3) os sentidos e significados do sistema semântico que resultam de superposição e integração de sentidos e significados de cada nível narrativo. Portanto, cada nível da mensagem, (1) expressivo, (2) sintático e (3) semântico, propõe sua própria narratividade que, para a pessoa receptora, influencia ou interfere na significação final da mensagem.

De modo sintético, considera-se (I) o plano estético, que reúne os aspectos expressivos, formais ou sintáticos e produz sentidos, e (II) o plano simbólico, que reúne os aspectos semânticos e produz significados (Figura 2).

Figura 2 – Diferenças estético-simbólicas na produção de imagens



Fonte:firefly.adobe.com

Por exemplo, uma imagem gerada pela ferramenta de IA Adobe Firefly, através do seguinte comando de texto: *“Create a realistic and historically inspired portrait of Helen of Troy, the iconic figure from Greek mythology. Helen should appear as a beautiful woman with golden, wavy hair, dressed in a flowing white or pastel-colored Grecian gown with intricate*



draping, reflecting the fashion of ancient Greek nobility. Her expression should be serene and regal, symbolizing both her divine beauty and tragic fate. Include a backdrop that suggests the ancient city of Troy or a classical Greek setting, with marble columns, vines, or the distant outline of a walled city. The lighting should be soft and golden, reminiscent of the Mediterranean sun, and the style should reflect the techniques of classical painters, emphasizing elegance, grace, and historical detail.”. A partir da  s o geradas diversas possibilidades, ambas as figuras descrevem a Helena de Troia (Figura 2). Mas, a partir do mesmo comando de texto, a imagem gerada pela IA resultou numa figura do g nero fotogr fico (Figura 2/A) e do g nero art stico (Figura 2/B).

No n vel art stico da figura de Helena, houve altera es nos formatos e nas sugest es visuais de textura, volumes e ilumina o, em fun o das mudan as de tonalidades, cores e contrastes visuais, com rela o   figura fotogr fica. No n vel sint tico da composi o com a figura art stica de Helena (Figura 2/B), houve o acr scimo de adornos utilizados por Helena. No n vel sem ntico, associado ao acr scimo de figuras de sentido (arquitetura, natureza), o conjunto das gera es expressivo-sint ticas do campo est tico resultaram na simboliza o de Helena em sua realidade e contexto hist rico.

O exemplo aqui proposto evidencia o potencial dos tratamentos est tico-expressivo e sint tico-figurativos com influentes ou interferentes na significa o final das mensagens visuais. Mas, assinala-se que as mesmas rela es influenciam ou interferem na mensagem, j  que todas as linguagens re nem aspectos est tico-expressivos e sint ticos, mesmo que, em conjunto, n o resultem em figuras mim ticas.

Considerando que se trata de imagens geradas por IA atrav s de um comando de texto,   poss vel afirmar que o mesmo conte do b sico, sendo informado de maneira diferente com varia es na linguagem, pode mudar a narrativa e influenciar na significa o final da mensagem (Perassi, 2015).

4 INTELIG NCIA ARTIFICIAL E PRODU O AUDIOVISUAL

Na atualidade, o uso de programas ou sistemas digitais de Intelig ncia Artificial (IA) na produ o de audiovisuais revolucionou essa pr tica e a ind stria cultural da m dia online. Trata-se de uma tecnologia em cont nuo desenvolvimento que automatiza processos e, al m disso, oferece m ltiplas possibilidades criativas.



Trata-se de uma realidade prática que, imediatamente, foi consolidada, principalmente, porque a estratégia básica das indústrias digitais designadas como *big techs* é oferecer versões gratuitas e com recursos limitados para as pessoas usuárias em geral. Isso populariza o uso da tecnologia, além de motivar ou mesmo obrigar em alguns casos a compra das licenças de uso das versões mais habilitadas ou profissionais.

Alves (2023) já havia destacado a ampla oferta de serviços e recursos digitais para a produção com inteligência artificial (IA) de textos, imagens, áudio, vídeos e audiovisuais com potencial de provocar mudanças na sociedade. Por exemplo, a autora menciona que isso pode melhorar a produtividade e automatização de serviços ou tarefas laborais. Mas, corre-se também o risco da inclusão de conteúdos descontextualizados ou falsificados com técnicas de *deepfake* (Westerlund, 2019). Portanto, Alves (2023) considera que o uso da IA pode melhorar significativamente a produtividade e automatização de processos, levantando também preocupações quanto à qualidade e a veracidade do que é produzido.

Ao analisar o cenário contemporâneo, Manovich (2018) considerou que, cada vez mais, os sistemas computacionais são capacitados para realizarem inúmeras operações com um tipo de super cognição. Aliás, isso permite que os sistemas artificialmente inteligentes desempenhem um papel crucial na transformação da mídia audiovisual, com muitas oportunidades e desafios.

O conceito *deepfake* pode ser referenciado a produtos audiovisuais digitais decorrente da coleta de informações referentes a rostos e vozes humanas em registros diversos. Depois disso, essas informações são usadas para emular uma situação falsa, representando uma performance imaginária com requintes de realismo audiovisual. Em síntese, *deepfake* é um termo baseado na língua inglesa para significar algo como “profundamente falso”, porque chega a simular com perfeição a performance natural de pessoas ou animais em situações que podem não ter ocorrido de fato.

É preciso assinalar que os recursos e os resultados associados com o conceito de *deepfake* podem ser usados de maneira consentida pelas partes envolvidas e legalmente apresentados de maneira esclarecida em produtos institucionais ou comerciais. Inclusive, isso colocou em questão as tradicionais atividades humanas de dublagem porque: além de reproduzir o conteúdo das mensagens nas diversas línguas naturais, a dublagem digital com IA ainda reproduz a frequência de vibração das cordas vocais de uma pessoa, emulando com perfeição o tom e o timbre vocal das pessoas.



Os recursos digitais de IA também configuram a dinâmica facial e os movimentos de boca das imagens de acordo com a fonética articulatória, que está sendo vocalizada na língua dublada. Além de alterar as expressões dos rostos e copiar entonações vocais, os recursos de edição de vídeos, sons e vozes podem substituir, subtrair ou acrescentar partes ou figuras das cenas audiovisuais, com plena verossimilhança, ou seja, com aparência de verdade ou qualidade de ser crível ou realista.

Por fim, há programas digitais abertos e acessíveis na rede online, os quais possibilitam a criação e o uso de personagens dinâmicos, com aparência e vozes em diferentes estilos, sendo ainda possível solicitar ou comandar que os recursos digitais de IA criem audiovisuais inéditos, a partir de imagens, gravações sonoras ou textos verbais que lhes são apresentados. Diante de tudo isso, é evidenciado que há programas digitais de inteligência artificial que, depois de acionados e direcionados, produzem narrativas qualificadas nos três níveis de composição das mensagens: expressivo, sintático e semântico, configurando os planos estético e simbólico multimídia das mensagens audiovisuais. A partir disso, também, possibilitam a produção de mensagens priorizando de maneira distinta ou integrada os cinco tipos de texto: descritivo, narrativo, dissertativo, expositivo e injuntivo.

3 RESULTADOS DA PESQUISA DOCUMENTAL NA REDE ONLINE

Na etapa exploratória da pesquisa desenvolvida para este estudo, foi realizado um levantamento online sobre a disponibilidade de programas ou sistemas digitais de inteligência artificial (IA), cujos recursos são adequados para diferentes fases da produção de audiovisuais. Para Cozman, Plonski e Neri (2021), os recursos digitais de IA podem ser usados nas seguintes fases: (1) pré-produção, (2) produção e (3) pós-produção de mensagens audiovisuais.

Na pesquisa realizada, trataremos de algumas ferramentas com foco na produção audiovisual. Isso é justificado pois nas diferentes fases os sistemas digitais de IA suprimem as atividades de diversas pessoas profissionais que, tradicionalmente, atuam na produção de audiovisuais. Diante disso, dependendo dos objetivos da comunicação, produzir e comunicar uma mensagem audiovisual tornou-se um ato praticamente corriqueiro e automático.

A produção de audiovisuais por sistemas digitais de IA inclui a criação de avatares (personagens) diversos e locuções em diversos idiomas. A seguir apresentamos alguns sistemas online que criam mensagens de áudio ou audiovisuais baseados em comandos de



texto. São sistemas suportados e dinamizados por plataformas digitais de acesso online. Como exemplos, no processo de pesquisa foram selecionadas e destacadas as seguintes plataformas: Synthesia, HeyGen, Eleven Labs e outras plataformas que ainda estão em fase de desenvolvimento, como o Sora da Open AI e VASA-1 da Microsoft.

3.1 SYNTHESIA

A plataforma digital Synthesia é uma plataforma de criação de vídeos que transforma texto em vídeos de alta qualidade utilizando avatares e narrações em mais de 130 idiomas. Suas principais funcionalidades incluem a criação de avatares personalizados, traduções automáticas de vídeos, e geração de roteiros e vídeos a partir de simples ideias ou documentos. Além disso, oferece templates de vídeo pré-projetados e integrações com ferramentas populares, permitindo colaboração e atualizações fáceis de conteúdo. Suas características funcionais e distintivas são as seguintes:

Características Funcionais: (1) conversão automática de textos em vídeos; (2) conteúdos visuais realistas que permitem a produção de audiovisuais eficientes, inclusive por pessoas sem experiência; (3) vasta gama de avatares digitais, que podem ser personalizados de acordo com as necessidades da mensagem; (4) os avatares simulam expressões faciais, labiais e entonações de voz de acordo o texto verbal fornecido; (5) multilinguismo, permitindo a criação de audiovisuais em diversas línguas naturais com precisão de pronúncia e entonação (isso é particularmente útil para pessoas ou organizações que atuam internacionalmente); (6) integração, porque permite a integração com outros recursos digitais de marketing e gestão de mídia ou organização; (7) personalização de cenários, roupas, sons, falas e outros aspectos expressivos.

Características Distintivas: (1) simplicidade; (2) acessibilidade; (3) eficiência de produção; (4) qualidade expressiva; (5) escalabilidade. A plataforma dispõe de uma interface intuitiva, que dispensa conhecimentos técnicos avançados, popularizando a produção de mensagens audiovisuais. A tecnologia digital reduz o tempo e os custos de produção, possibilitando que pessoas e organizações produzam vídeos eficientes com rapidez e economia de trabalho e recursos. Os audiovisuais produzidos apresentam qualidade, com avatares realistas que colaboram na eficácia da mensagem. A plataforma também facilita a criação de conteúdos audiovisuais em massa, permitindo a realização de campanhas de



marketing, treinamentos corporativos e outras atividades com grande volume de mensagens audiovisuais.

3.2 HEYGEN

A plataforma digital HeyGen é outra plataforma que cumpre funções similares e permite criar avatares virtuais personalizados de forma simples e intuitiva. Com ele, os usuários podem escolher características como gênero, idade, etnia, estilo de cabelo, cor dos olhos e muito mais para criar um avatar único. Além disso, oferece opções avançadas de personalização, como expressões faciais, acessórios, roupas e poses. Os avatares gerados podem ser integrados facilmente em diversos aplicativos e plataformas, como jogos, aplicativos de mensagens e redes sociais. A plataforma é projetada para ser acessível a usuários de todos os níveis de habilidade, permitindo criar avatares em pouco tempo. Também oferece recursos adicionais, como animações simples, ferramentas de edição para ajustes finos nos avatares e foi recentemente integrada ao Adobe Express. Há também particularidades e características distintivas.

Características Distintivas: (1) usabilidade e acessibilidade; (2) qualidade expressivo-comunicativa; (3) eficiência; (4) escalabilidade. De acordo com informações de textos técnico-críticos e opiniões de pessoas usuárias coletadas no ambiente online, a tecnologia desta plataforma é mais avançada, há investimentos continuados em inovação, com atualizações regulares para ampliar a funcionalidade e a qualidade dos audiovisuais. Isso aprimora ainda mais o realismo das produções e propõe melhorias para o acesso e o uso dos recursos, com abordagem simplificada, melhorando a acessibilidade de todas as pessoas usuárias. Oferece mais facilidades e recursos para a personalização de avatares, permitindo optar por aspectos particulares de gênero, idade, etnia, estilo de cabelo, cor de olhos ou outros, com a inclusão de mais variações para expressões faciais, acessórios, vestimentas e poses. Os avatares são facilmente integrados a outras plataformas de jogos ou aplicativos de mensagens e redes sociais. Há ainda recursos adicionais de animações simples, edição com ajustes finos. Os tempos de renderização são menores, também os custos são competitivos em comparação com outras plataformas. A plataforma foi integrada ao pacote de produtos Adobe Express. Com relação a escalabilidade, esta plataforma oferece recursos específicos para gerenciar e automatizar a criação de múltiplos audiovisuais simultaneamente, destacando-se também por isso diante das outras plataformas.



3.3 ELEVEN LABS

A plataforma digital Eleven Labs é uma plataforma de inteligência artificial focada na criação de áudios com vozes realistas. Na versão gratuita, possibilita a criação instantânea de vozes em 29 idiomas e oferece recursos de dublagem e clonagem de voz.

Características Funcionais: (1) síntese de voz com alta qualidade, com o uso de redes neurais avançadas, gerando vozes que soam naturais e humanizadas; (2) oferece opções de personalização de tom, velocidade e emoção na voz; (3) multilinguagem, porque reproduz diversas línguas ou dialetos com alta precisão; (4) Interface de programação de aplicativos (API - Application Programming Interface) que pode integrada em diferentes aplicativos móveis, websites ou sistemas de atendimento ao cliente; (5) interface intuitiva e considerada amigável para a criação e edição de vozes sem conhecimentos técnicos avançados; (6) aprendizado contínuo, com melhorias recorrentes usando os feedbacks das pessoas usuárias, atualizando qualitativamente a síntese de vozes.

Características Distintivas: (1) qualidade superior de áudio, com frequente elogios de pessoas usuárias pela qualidade natural das vozes em comparação com plataformas similares; (2) personalização avançada, com mais opções de ajustes que outras plataformas; (3) reprodução com qualidade de diversos idiomas; (4) integração com outros sistemas digitais, por dispor de API bem documentada e interface amigável. Apesar da interface intuitiva, o máximo proveito ou o uso profissional da plataforma Eleven Labs exigem mais tempo de aprendizado.

3.4 SORA

O sistema digital multiplataforma OpenAI é composto por um conjunto de plataformas e recursos baseados em inteligência artificial, entre elas o Sora, que está em fase de teste, mas que já promete ser um modelo de IA que pode criar cenas realistas e imaginativas a partir de instruções de texto. Atualmente está disponível para profissionais avaliarem áreas críticas em busca de danos ou riscos, e para artistas visuais, designers e cineastas com a finalidade de que o modelo seja mais útil para profissionais criativos.

No site da Open AI, é possível acessar um relatório técnico que aponta que o Sora, atualmente, exhibe limitações como simulador. Por exemplo, ele não modela com precisão a



física de muitas interações básicas, como vidro quebrando e outras interações, como comer alimentos, que nem sempre produzem mudanças corretas no estado do objeto. Liu *et al.*, (2024) afirmam que o futuro desenvolvimento de modelos Sora e de geração de vídeo em geral, e como os avanços no campo poderiam permitir novas formas de interação humano-IA, aumentando a produtividade e a criatividade na geração de vídeo.

3.5 VASA-1

A pesquisa da Microsoft denominada VASA-1 (Xu *et al.*, 2024), discorre sobre um framework capaz de gerar rostos falantes realistas de personagens virtuais com habilidades afetivas visuais atraentes, dada uma única imagem estática e um clipe de áudio de fala. VASA-1, é capaz não apenas de produzir movimentos labiais perfeitamente sincronizados com o áudio, mas também de capturar um amplo espectro de nuances faciais e movimentos naturais da cabeça que contribuem para a percepção de autenticidade e vivacidade.

As principais inovações incluem uma dinâmica facial holística e um modelo de geração de movimento da cabeça que funciona em um espaço latente facial, e o desenvolvimento de um espaço latente facial expressivo e desembaraçado usando vídeos. O método não apenas oferece alta qualidade de vídeo com dinâmica facial e de cabeça realista, mas também suporta a geração online de vídeos 512x512 a até 40 FPS com latência inicial insignificante. Essa pesquisa abre caminho para engajamentos em tempo real com avatares realistas que emulam comportamentos de conversação humana.

A pesquisa ainda considera a possibilidade de uso indevido e estão empenhados em desenvolver a IA de forma responsável, com o objetivo de promover o bem-estar humano, por isso, a Microsoft não tem planos de lançar uma demonstração on-line, produto ou qualquer oferta relacionada até que tenham certeza de que a tecnologia será usada de forma responsável e de acordo com os regulamentos adequados.

4 CONSIDERAÇÕES FINAIS

Como foi exemplificado neste artigo, há plataformas ou sistemas digitais de IA que oferecem diversos recursos para a produção de mensagens audiovisuais digitais. Essas são produzidas para serem acessadas ou comunicadas online. A possibilidade de acesso aos recursos digitais de inteligência artificial podem ainda ser direcionados para a produção de



narrativas qualificadas nos três níveis compositores das mensagens: (1) expressivo, (2) sintático e (3) semântico, configurando os planos (I) estético e (II) simbólico na produção de mensagens audiovisuais. Além disso, os recursos digitais de IA também podem ser direcionados para produzirem mensagens audiovisuais com diferentes tipos de texto: (1) descritivo, (2) narrativo, (3) dissertativo, (4) expositivo e (5) injuntivo. Além disso, pelo menos parcialmente, esses tipos de textos podem ser combinados na mesma mensagem audiovisual.

Diferentemente das fotografias que, fisicamente, registram imagens de ocorrências factuais ou “reais”. As habilidades técnicas da IA muito impressionantes e, talvez, as mais relevantes são designadas como “realistas” e diretamente associadas com a “verossimilhança”. Trata-se da conquista continuada de diversos graus miméticos (imitativos), configurando representações imaginárias, que são apresentadas nas telas de dispositivos digitais com sistemas de imagem e som.

As representações imaginárias são ilusoriamente criadas e observadas como registros multimídia de ocorrências da realidade factual. Aparentemente, como algo que já ocorreu ou que, sincronicamente, ocorre em ambientes naturais ou sociais. A pregnante ilusão de verdade diante das mensagens audiovisuais produzidas por IA induziu a utilização do termo deepfake. Assim, parte das imagens ficcionais costumam ser percebidas como reais e personagens irreais agora designados como “avatars” podem ser comumente confundidos com imagens de pessoas reais. As imagens de pessoas reais, vivas ou mortas, também podem ser tecnologicamente e verossimilmente manipuladas para dizerem palavras, tomarem atitudes ou praticarem ações que, até então, só existiam na imaginação de manipuladores da mídia audiovisual.

Diante do exposto, é possível afirmar que as mensagens audiovisuais produzidas com os recursos de plataformas ou sistemas digitais de IA estão transformando significativamente os diversos campos sociais. Observa-se uma mudança significativa na produção de material audiovisual, impulsionada pela crescente acessibilidade das ferramentas de IA, até mesmo para usuários amadores.

No entanto, é essencial explorar como a IA está influenciando a criação e a disseminação de conteúdo audiovisual, levando em consideração suas implicações para as narrativas e os significados produzidos. Além disso, é necessário investigar os impactos na qualidade do conteúdo final e acompanhar as produções audiovisuais em todas as suas fases.



Essa reflexão é fundamental para entendermos como a IA está moldando o cenário da produção audiovisual e como podemos aproveitar seus benefícios de maneira ética e eficaz.

REFERÊNCIAS

- Alves, K. C. Criação de conteúdo audiovisual, Inteligência Artificial e Desinformação: como a TV pode colaborar na popularização dessa discussão. Intercom: PUC-Minas, 2023.
- Cozman, F. G.; PLONSKI, G. A.; NERI, H (Org.). Inteligência artificial: avanços e tendências. São Paulo: Instituto de Estudos Avançados, 2021.
- Côrtes, H. S. O paradigma do pensamento dos professores como contexto do trabalho em Informática na educação. Porto Alegre: EDIPUCRS, 2003.
- Liu, Y., Zhang, K., Li, Y., Yan, Z., Gao, C., Chen, R., Yuan, Z., Huang, Y., Sun, H., Gao, J., He, L., & Sun, L. Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models. ArXiv, abs/2402.17177, 2024.
- Manovich, L. AI Aesthetics. Moscou: Strelka Press, 2018.
- Perassi, R. Do ponto ao pixel: sintaxe gráfica no videodigital. Florianópolis: CCE/UFSC, 2015.
- Ryan, M. L. "Media and Narrative." In: The Routledge Encyclopedia of Narrative Theory. London: Routledge, 2005, p. 288–92.
- Travaglia, L. C. A caracterização de categorias de texto: tipos, gêneros e espécies. São Paulo: Alfa, 51 (1): 39-79, 2007.
- Xu, S., Chen, G., Guo, Y., Yang, J., Li, C., Zang, Z., Zhang, Y., Tong, X., & Guo, B. VASA-1: Lifelike Audio-Driven Talking Faces Generated in Real Time. ArXiv, abs/2404.10667, 2024.
- Westerlund, M. The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9(11), 2019, p. 40-53.