

DOSSIÊ

Narrativas de IA: tendências da produção audiovisual



V. 14 – N. 3 - set./dez. 2023

ISSN: 2179-1465 / <https://www.revistageminis.ufscar.br>

DOI: <https://doi.org/10.14244/2179-1465.RG.2023v14i3p87-104>

AUDIOVISUAL E INTELIGÊNCIA ARTIFICIAL: FERRAMENTAS AUTÔNOMAS NA PRODUÇÃO DE CONTEÚDO

AUDIOVISUAL AND ARTIFICIAL INTELLIGENCE: AUTONOMOUS TOOLS IN CONTENT PRODUCTION

AUDIOVISUAL E INTELIGENCIA ARTIFICIAL: HERRAMIENTAS AUTÓNOMAS EN LA PRODUCCIÓN DE CONTENIDOS

Ana Silvia Lopes Davi Médola

Universidade Estadual Paulista (UNESP)

ORCID: <https://orcid.org/0000-0003-2101-3727>

Bauru, SP, Brasil

Vinicius Laureto de Oliveira

Universidade Estadual Paulista (UNESP)

ORCID: <https://orcid.org/0009-0000-5764-7022>

Bauru, SP, Brasil

Henrique da Silva Pereira

Universidade Estadual Paulista (UNESP)

Centro Universitário Nossa Senhora do Patrocínio (CEUNSP)

ORCID: <https://orcid.org/0000-0002-1328-8694>

Bauru, SP, Brasil

Recebido: 14/10/2023 / Aprovado: 03/02/2024

Como citar: MÉDOLA, A. S. L. D.; OLIVEIRA, V. L. de; PEREIRA, H. da S. Audiovisual e Inteligência Artificial: ferramentas autônomas na produção de conteúdo. Revista GEMINIS, v. 14, n. 3, p. 87–104, 2023

Direito autoral: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 3.0 Internacional.



RESUMO

As ferramentas digitais com recursos baseados em análises de dados e inteligência artificial abrem perspectivas de transformações no âmbito do audiovisual. Com o objetivo de refletir sobre experiências na utilização da inteligência artificial na produção de conteúdos audiovisuais o presente trabalho analisa projetos derivados, total ou parcialmente, da geração de textos autônomos, por meio de processamento de linguagem natural e recursos de visão computacional. No estágio das experiências analisadas observou-se limitações importantes no processo de tratamento de dados não estruturados e identificação de padrões por ferramentas de análise de texto e imagem. Com base na semiótica discursiva, o artigo aponta que embora a produção de sentido nesses conteúdos apresente lacunas não compromete o contrato de comunicação.

Palavras-chave: Produção audiovisual; Inteligência artificial; Produção de sentido.

ABSTRACT

Digital tools with resources based on data analysis and artificial intelligence open up the prospect of transformations in the audiovisual field. With the aim of reflecting on experiences in the use of artificial intelligence in the production of audiovisual content, this paper analyses projects derived, in whole or in part, from the generation of autonomous texts using natural language processing and computer vision resources. In the experiments analysed, important limitations were observed in the process of processing unstructured data and identifying patterns using text and image analysis tools. Based on discursive semiotics, the article points out that although there are gaps in the production of meaning in these contents, this does not jeopardise the communication contract.

Keywords: Audiovisual production; Artificial intelligence; Production of meaning.

RESUMEN

Las herramientas digitales con recursos basados en el análisis de datos y la inteligencia artificial abren la perspectiva de transformaciones en el ámbito audiovisual. Con el objetivo de reflexionar sobre experiencias en el uso de inteligencia artificial en la producción de contenidos audiovisuales, este trabajo analiza proyectos derivados, total o parcialmente, de la generación de textos autónomos utilizando recursos de procesamiento de lenguaje natural y visión por computador. En los experimentos analizados, se observaron importantes limitaciones en el proceso de procesamiento de datos no estructurados y en la identificación de patrones mediante herramientas de análisis de texto e imágenes. Basándose en la semiótica discursiva, el artículo señala que, aunque existan lagunas en la producción de sentido de estos contenidos, ello no pone en riesgo el contrato de comunicación.

Palabras Clave: Producción audiovisual; Inteligencia artificial; Producción de sentido.

INTRODUÇÃO

Os avanços tecnológicos relacionados ao aumento do poder de processamento dos computadores, a velocidade das conexões em rede e a geração de grandes quantidades de dados permitiram a expansão da inteligência artificial (IA) impactando diversas frentes de pesquisa e mercado, dentre as quais a comunicação. Esta evolução das tecnologias da computação e análise de dados em grande escala são a base dos avanços na área de inteligência artificial presentes também na produção audiovisual, ponto de interesse do presente artigo, que descreve e analisa aplicações em sistemas de Processamento de Linguagem Natural e Visão Computacional realizadas no final da segunda década do século XXI. Após as análises são apresentados dois experimentos de produção audiovisual, com o propósito de demonstrar o potencial de uso de ferramentas semelhantes disponíveis para usuários até então, sendo o primeiro experimento realizado na *Google Vision* e o segundo na *Google Vision autoML*.

Embora as referências à inteligência artificial sejam encontradas com maior frequência apenas após o início da década de 2010, a área de IA está entre os tópicos recorrentes das pesquisas relacionadas à informática desde a metade do século XX. As consequências geradas com a inserção da IA em diversos sistemas produtivos como os de tecnologia, finanças, gestão de negócios e comunicação, impõem a necessidade de discussões de caráter ético, econômico, cultural, social sobre as implicações da adoção dos sistemas inteligentes na vida das pessoas.

Como exemplo, podemos citar as situações comunicacionais que inicialmente pareciam não produzir grandes impactos, como o uso de *bots* (robôs), e passaram a ter relevância enquanto mecanismos geradores de diferentes implicações na dinâmica das disputas políticas, como na realização dos ataques a perfis de candidatos no *Twitter*, atualmente X, durante os debates televisionados para a eleição presidencial do Brasil em 2018¹. Da mesma forma, cabe registrar o impulsionamento massivo de notícias falsas, as chamadas *Fake News*, que permearam também a campanha eleitoral à presidência da República do Brasil em 2022. Nesse sentido, a necessidade de compreensão sobre a utilização de dados em estratégias comunicacionais está presente em áreas como o jornalismo, a publicidade e o marketing, auxiliando nas análises sobre formação da opinião pública ou criação de projeções para a tomada de decisões.

Cabe salientar que a utilização de ferramentas com inteligência artificial para a comunicação está estruturada no *insight* a partir dos dados gerados e sua posterior automação. Tais sistemas são treinados para realizar funções semelhantes à cognição humana como escrever mensagens, que

¹ “No primeiro debate presidencial, 10% dos tuítes foram de robôs ou fakes” Disponível em: <https://oglobo.globo.com/brasil/no-primeiro-debate-presidencial-10-dos-tuites-foram-de-robos-oufakes-22970273> Acesso em 14 out. 2023

podem ser de ataque ou defesa a candidatos durante os debates eleitorais televisionados, como citado acima. E os impactos destes recursos na comunicação são concretos, conforme observou-se na história recente da cena política brasileira.

Grande parte das técnicas empregadas no desenvolvimento de processos inteligentes baseiam-se na reprodução de padrões, de modo que um computador programado a partir destas técnicas pode identificar e copiar, ou “criar” uma saída adequada aos dados de entrada, como no caso dos avanços atualmente em curso demonstrados nas redes generativas desenvolvidas pela OpenAI, empresa de pesquisa em inteligência artificial, financiada majoritariamente pela Microsoft que busca o aprimoramento de tecnologias de IA. A OpenAI é conhecida pelos produtos DALL-E 3 e ChatGPT. O DALL-E 3 é uma versão avançada do DALL-E, um modelo de IA capaz de gerar imagens novas a partir de descrições textuais, utilizando técnicas de aprendizado de máquina para criar textos imagéticos, possibilitando a geração de imagens com base em textos verbais escritos fornecidos pelos usuários. Já o ChatGPT é um modelo de conversação que permite a interação com o usuário por meio de texto verbal escrito. Treinado em uma ampla gama de dados, o ChatGPT é capaz de responder perguntas, fornecer informações, desenvolver textos de diferentes gêneros discursivos e até mesmo participar de conversas em diversos tópicos.

Sabemos que as primeiras aplicações com IA presentes no cotidiano dos cidadãos tomaram as mesas de trabalho, bolsos e nuvens de usuários em volta de todo o planeta. A grande quantidade de dados gerados diariamente por usuários na internet produz o escopo para o treinamento de diversos sistemas inteligentes que podem atuar, no caso da área de comunicação, como os de sugestão e encadeamento de conteúdo, criação de notícias, personalização de marketing, *chatterbots*² e outros.

Atualmente, entre os relevantes avanços na área da produção audiovisual podemos citar ferramentas de produção de vídeos que incorporam a IA como o Doodly que foi lançado em 2016 pela Doodly LLC. Trata-se de um recurso de criação audiovisual, cuja principal característica é a capacidade de criar vídeos que parecem ser desenhados à mão. Oferecendo uma variedade de elementos de desenhos e animação, o Doodly, pelo seu caráter intuitivo, permite que os usuários consigam criar seus próprios vídeos animados de maneira rápida e fácil sem nenhum conhecimento prévio. No ano seguinte, 2017, foi lançada a Kapwing, uma plataforma de edição de vídeo online com uso de inteligência artificial. A ferramenta utiliza o método de criação de vídeos a partir de textos escritos pelo usuário. Suas características incluem a adição de legendas, música, sobreposições e efeitos especiais. Com a IA da Kapwing Inc. simplificando e acelerando o processo de edição, torna-

² Softwares capazes de conversar com usuários por meio de chats simulando a conversação com um humano através de aplicativos de redes sociais, como o Facebook e o WhatsApp.

se possível criar diferentes tipos de conteúdos, como *vlogs*, tutoriais e memes, tornando a edição de vídeo acessível e rápida. Outro exemplo é a Synthesia, plataforma de geração de vídeo voltada à criação de conteúdos personalizados por meio de avatares digitais realistas. Lançada em 2020 pela Synthesia Technologies Ltd., a ferramenta tem a capacidade de produzir vídeos em vários idiomas e estilos, sendo que os avatares da Synthesia “falam” e “agem” de acordo com o roteiro fornecido.

Assim, com essa gama de tecnologias comunicacionais sendo criadas a partir da lógica dos sistemas inteligentes, novos desafios se apresentam para a pesquisa. Autores como Gunkel (2017), apontam que no contexto da inteligência artificial, os algoritmos não serão meros instrumentos de ações comunicativas de humanos ou meio pelo qual humanos se comunicam entre si. O autor acredita que ocuparão, em vez disso, a posição de outro ator social com quem alguém se comunica ou interage (GUNKEL, 2017, p. 16). Observe-se que tal ponderação demonstra preocupação em relação aos caminhos dos sistemas inteligentes e a forma como estes passam a atuar nas relações de comunicação constituindo-se em um ator neste processo. Nesse sentido, verificamos um alerta para a necessidade de pesquisas que relacionem o campo das ciências sociais aplicadas com os sistemas autônomos, sendo que à comunicação cabe responder a questionamentos relacionados aos impactos sociais advindos dos processos e fluxos de interação entre destinatários e destinatários dos conteúdos produzidos pelos dispositivos que incorporam a IA.

INTELIGÊNCIA ARTIFICIAL ENTRE IMAGINÁRIOS E NARRATIVAS

Da literatura ao cinema, a possibilidade da máquina simular de algum modo a inteligência humana alimenta diferentes visões e formas de percepção sobre a questão. Isaac Asimov definiu em seu conto *Círculo Vicioso*, presente na coletânea *Eu, Robô* (1950), as três leis da robótica: (1) um robô não pode ferir um ser humano ou, por não ação, permitir que um ser humano sofra algum mal; (2) um robô deve obedecer às ordens dadas por seres humanos exceto nos casos em que tais ordens entrem em conflito com a primeira lei e (3) um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a primeira ou a segunda lei. Para Asimov, os robôs evoluíam à medida que seus *cérebros positrônicos* se tornavam mais completos, permitindo às máquinas a interpretação e interação com o mundo à sua volta. A identificação de padrões que permitem a tomada de decisão por uma máquina é o fator de disruptividade que, alinhado à distopia tecnológica, guiou o enfoque de narrativas ficcionais acerca do tema. Cabe lembrar também que em um viés histórico, diversas narrativas que abordam um modo opressivo de dominação das máquinas são desenvolvidas em uma sociedade do pós-guerra, com menos de uma década após o fim da segunda guerra mundial

e durante o período da guerra fria, no qual a ameaça nuclear e a corrida espacial são uma tônica nos discursos.

No tocante à produção simbólica audiovisual, cabe citar o filme *O Dia em que a Terra Parou* (1951, EUA) que foi lançado apenas um ano após a publicação da coletânea *Eu, Robô*. Na trama do longa-metragem, o alienígena Klaatu aterrissa na terra com seu robô, Gort, uma grande figura antropomorfizada de metal, em busca de uma conciliação entre os líderes políticos do planeta, visando parar a corrida armamentista. Durante o filme, o invencível robô Gort é acionado para conter os humanos que veem em Klaatu uma ameaça para a humanidade. Em seu desfecho, Klaatu dá um ultimato à Terra: se os humanos não pararem com extensivo armamento, Gort e outros robôs destruirão o planeta. *O Dia em que a Terra Parou* foi o primeiro longa-metragem de ficção científica feito por um estúdio de renome, a *20th Century Fox*, e com um grande orçamento (US\$ 1.2 milhões), tornando-se um dos elementos do cânone. Outros títulos do mesmo gênero, como *2001 - Uma Odisseia no Espaço* (1968, EUA), *O Exterminador do Futuro* (1984, EUA) e *Matrix* (1999, EUA), também apresentam realidades arruinadas pelas máquinas inteligentes.

As narrativas ficcionais que corroboram para a visão temerária de que as máquinas possam voltar-se contra a humanidade e iniciar algum tipo de conflito, assim como retratado nos filmes, é ratificado no espaço midiático e científico em afirmações como “A inteligência artificial ameaça a existência da nossa civilização”³, proferida pelo do CEO da *SpaceX*, Elon Musk, ou comentários do físico Stephen Hawking em 2014, de que “o desenvolvimento da inteligência artificial total poderia significar o fim da raça humana”⁴.

Embora uma grande quantidade de resultados dos experimentos e aplicações com inteligência artificial tenham acontecido apenas nos últimos anos, simultaneamente, com as observações e questionamentos sobre os vieses da tecnologia, a formalização do campo de estudo se deu em meados do século passado, durante a década de 1950. Em 1955, no estado de New Hampshire, EUA, foi realizada a convenção de Dartmouth, reunindo cientistas de universidades americanas e de empresas como a IBM, com o objetivo de submeter um projeto de pesquisa à Fundação Rockefeller. O estudo, pioneiro na área da informática, apresentou a primeira menção ao termo “Inteligência Artificial”. A pesquisa preconizava que “cada aspecto do aprendizado, ou outra forma de inteligência, pode ser descrito de forma tão precisa que uma máquina pode ser criada para simular isso” (MCCARTHY, 1955, p.12).

³ Disponível em https://brasil.elpais.com/brasil/2017/07/17/tecnologia/1500289809_008679.html Acesso em 14 out. 2023

⁴ Disponível em <https://arquivo.canaltech.com.br/entretenimento/a-era-do- robo-sapiens-o-dia-em-que-a-inteligencia-artificial-dominar-a-terra-72905/> Acesso em 14 out. 2023.

Décadas de pesquisa sobre IA após a convenção de Dartmouth demonstram que a engenharia de fazer máquinas inteligentes está fundada em sistemas computacionais que possibilitam entender a inteligência humana, pois a inteligência artificial é desenvolvida a partir da observação de como humanos procuram pela solução de um problema. “A melhor forma é fazer programas de computador que possam resolver problemas e alcançar objetivos deve ser como nos seres humanos” (MCCARTHY, 1955, p.12), de modo que a IA constitui, muitas vezes, na ação de projeção do pensamento humano, da busca por padrões e categorizações de tudo o que nos cerca, simulando, por meio de programação eletrônica, neurônios e sinapses.

No entanto, um cérebro simulado por um computador, ou *positrônico* como cunhado por Asimov, trabalha com dados de forma diferente do humano. Um algoritmo pode ser brevemente definido como uma sequência de instruções finitas com o objetivo de solucionar um problema. No entanto, algoritmos em um computador funcionam com dados estruturados, ou seja, números inteiros ou racionais, verdades ou falsidades, ciclos de energia, condições binárias. Trabalhar com dados exatos, como idade e renda de uma família, é fácil para um computador. Entretanto, para aprimorar a interface humano-máquina, era preciso entender o que os humanos falavam.

Em 1957, Noam Chomsky lança seu livro *Estruturas Sintáticas*, que segundo NORVING e RUSSEL (2013), apresentava uma teoria linguística formal o suficiente para ser programável. Entre as principais postulações do linguista aplicadas à IA está a chamada Hierarquia de Chomsky, definindo propriedades essenciais sobre a área de linguagens formais e naturais. A hierarquia chomskyana, oferece subsídios à informática para o estabelecimento da relação dos compiladores de linguagem de programação voltados à criação de textos de forma autônoma. Esta relação de linguagem natural e geração autônoma é realizada em níveis sendo o nível um, linguagens sensíveis ao contexto, e zero, linguagens recursivamente enumeráveis. Assim, linguística e inteligência artificial formam um campo híbrido chamado Linguística Computacional ou Processamento de Linguagem Natural, abrindo as perspectivas de inserção da IA em diferentes produtos como o ChatGPT e O DALL-E 3, conforme mencionado, e que impactam diretamente setores produtivos da comunicação, entre os quais o audiovisual, conforme veremos.

PROCESSAMENTO DE LINGUAGEM NATURAL E VISÃO COMPUTACIONAL: A TRADUÇÃO DOS ALGORITMOS NA BASE DE APLICAÇÃO NO AUDIOVISUAL

Com o advento da internet, do aumento do processamento dos computadores e da capacidade de armazenamento, bem como da digitalização dos acervos, diversas ferramentas destinadas às análises de conteúdos audiovisuais com o uso de IA tornaram-se acessíveis aos produtores de

conteúdo. Em função disso é amplo o espectro de possibilidades ligadas ao reconhecimento de padrões que podem ser utilizadas na produção audiovisual. Com o propósito de demonstrar formas de apropriação de ferramentas neste tipo de produção, direcionamos as análises para os primeiros recursos de Processamento de Linguagem Natural, referentes ao reconhecimento em textos escritos e também para a Visão Computacional, em que a máquina pode ser treinada para reconhecer padrões e objetos em fotos e vídeos. Essas ferramentas trabalham, na maioria das vezes, a partir da lógica de reconhecimento de padrões e classificação mediante uma escala. Para fazer o reconhecimento, uma ferramenta depende dos diversos fatores para os quais foi programada para analisar. Esses fatores, denominados parâmetros, atuam como uma referência aos dados recebidos por uma função na programação de softwares.

No que diz respeito à Geração de Linguagem Natural aplicadas à área da comunicação, observa-se que as primeiras ferramentas foram apropriadas para a criação de notícias e publicações automáticas em redes sociais. Encontramos disponíveis nesses ambientes alguns serviços de geração automática de texto que funcionam a partir de um corpus de materiais semelhantes. Entre as aplicações com esta capacidade estão o *Newsomatic* e o *Article Generator*, para a produção de notícias automáticas. Ambos os serviços funcionam a partir do escaneamento de informativos com palavras-chave escolhidas pelo usuário. A partir dos resultados obtidos na busca, o algoritmo consegue realizar a conjunção das informações em um único texto. Outro resultado que o sistema apresenta é o de um texto único, utilizando sinônimos dos trechos de textos capturados das notícias. Podemos destacar outras aplicações bem-sucedidas na área de geração linguagem natural como os *chatbots*, que trabalham com o atendimento aos usuários em perfis corporativos nas redes sociais, e os serviços de assistentes pessoais como o *Google Assistant* e a *Siri*.

Como objeto de interesse da utilização do Processamento de Linguagem Natural no audiovisual analisaremos alguns aspectos da adoção da IA na produção do curta-metragem de ficção científica *Sunspring* (2016). De acordo com MCKEOWN (2017, p.19) “fazer um computador contar uma história é realmente uma história”. A dificuldade em fazer uma máquina entender funções cognitivas que sejam capazes de criar conteúdos inéditos é um desafio à lógica computacional. Embora existam outras experiências na área de criação de roteiros utilizando tecnologias com o processamento de linguagem natural, destacamos o *Benjamin*, rede neural de recorrências responsável pela criação do roteiro do curta-metragem de *Sunspring*. O sistema foi treinado a partir dos roteiros de filmes que representam o cânone do gênero. Dessa forma, os roteiros de filmes como *Blade Runner* (1982, EUA) e *Independence Day* (1996, EUA) estavam presentes no banco de dados de *Benjamin*. A criação das falas do roteiro utiliza redes neurais recorrentes de memória de longo prazo (LSTM),

que segundo Oscar Sharp⁵, dá a capacidade para o computador gerar parágrafos inteiros baseados em um banco de dados, concebendo conteúdo original e não apenas frases copiadas do corpus.

O curta metragem *Sunspring* se passa em um futuro distópico em que três pessoas (H, H2 e C) acabam se envolvendo em um triângulo amoroso. O curta foi originalmente produzido em 48 horas para o concurso de criações do Festival de Ficção Científica de Londres (2016) e foi indicado entre os 10 melhores filmes. O concurso definiu que a escolha dos melhores filmes seria por votação online. Ao perceber que os outros curtas estavam ganhando muitos votos, e provavelmente utilizando *bots* para a votação, os autores utilizaram o poder de processamento de *Benjamin* para direcionar 36 mil votos por hora para o curta. O concurso teve seu resultado cancelado.

O roteiro de *Sunspring* possui uma narrativa descontínua, beirando a aleatoriedade, com falas desconexas. No entanto, ao assistirmos o curta produzido podemos afirmar ao menos que é possível depreender sentido em sequências do filme. Sendo o audiovisual uma linguagem sincrética, isto é, um texto-objeto constituído pela utilização de duas ou mais linguagens de manifestação que interagem, formando um todo de significação (FLOCH, 2001), a produção de sentido extrapola o nível de articulação que parte do roteiro, marcado por um certo grau de incoerência no encadeamento dos programas narrativos. Entretanto, as isotopias figurativas manifestadas visualmente nos cenários, nas projeções dos atores do discurso, no figurino, na gestualidade, bem como na manifestação de linguagens sonoras como os diálogos e a trilha musical, compõem o filme de forma a produzir efeitos de sentido ao articular possibilidades de leituras.

Assim, verificamos que entre essas linguagens, a montagem tem importância fundamental, reiterando temas e figuras, por meio de arranjos enunciativos dos quais emerge a produção de sentido para o roteiro. Seja em procedimentos que remetem à estética de uma não sequencialidade narrativa evidente, presente em filmes como os da *Nouvelle Vague*, por exemplo, ou a iminência de depreender um sentido decorrente da união de planos, remetendo ao Efeito Kulechov (XAVIER, 2005. p.47), *Benjamin* apresenta potencial de atuar tanto como ferramenta coadjuvante no processo de roteirização, quanto como o próprio “roteirista” de textos audiovisuais, ainda que em estágio marcado por um baixo nível de coerência na organização semântica verbal. Esta questão está no centro dos debates dos roteiristas de cinema e TV nos Estados Unidos, desdobrando em movimentos grevistas nos anos recentes e ações para regulamentar a utilização da IA na produção audiovisual.

⁵ GOODWIN, Ross; SHARP, Oscar. *Machines Making Movies*. TEDxBoston. Disponível em https://www.youtube.com/watch?v=uPXPQK83Z_Y Acesso em 14 out. 2023.

IMAGENS E SONS EM MÉTRICAS COMPUTACIONAIS

Outra área que merece destaque nas experiências audiovisuais com inteligência artificial é a de reconhecimento de imagens e sons. A partir da entrada de dados visuais e sonoros é possível processar saídas personalizadas. As transmissões esportivas ao vivo estão entre as dinâmicas televisivas que mais trabalham com dados simultâneos, como os das câmeras, *replays*, videografismos, os de som dos microfones do campo e das arquibancadas, além dos dados estatísticos de pontuação e cronometria. A quantidade de dados armazenados, processados e transmitidos em campeonatos ou torneios de grande escala confere um elevado grau de complexidade para a atividade de diversos profissionais envolvidos.

Com o propósito de agilizar a produção de conteúdos a partir dos dados gerados na transmissão de eventos esportivos, em 2017, a *IBM* aplicou a tecnologia de seu supercomputador, o *Watson*, no centenário torneio de tênis de *Wimbledon*. O desafio proposto era o de criar vídeos automáticos com duração de até dois minutos com os melhores momentos da partida logo ao fim do jogo. *Wimbledon* é o mais antigo torneio de tênis do mundo e realiza mais de 250 partidas em duas semanas. De acordo com Rogério Feris, gerente de *computer vision and multimedia research*, este foi um desafio “além da seleção e curadoria de trechos dos jogos para um editor de vídeo”⁶. Para a criação dos vídeos automáticos foi preciso associar e processar as imagens e áudios captados com os dados de telemetria disponíveis.

Entre os dados da telemetria coletados diretamente na quadra estão as estatísticas de pontuação, acertos e erros, anotações do juiz, velocidades de saque e posições da bola e dos jogadores. A frente de reconhecimento de imagem e som é responsável por entender a reação do público presente e de ações dos jogadores que possam ser diferentes do padrão. Para iniciar a produção do vídeo de melhores momentos, o *Watson* foi treinado para reconhecer o aperto de mão que finaliza a partida entre os jogadores. O treinamento da máquina para reconhecer reações da torcida, ações dos jogadores e o aperto de mão final foi feito a partir de vídeos do torneio de anos anteriores.

O aprendizado do *IBM Watson*, bem como de outros computadores inteligentes, para tarefas envolvendo audiovisual, acontece sempre a partir de uma grande base de dados. Destacamos na seleção de momentos do jogo para a composição do vídeo a lógica de organização dos elementos da linguagem. FLOCH (2001, p. 13) afirma que os objetos semióticos estruturados enquanto linguagens podem ser “considerados sob dois aspectos, o do sistema e do processo”. O autor lembra que o sistema

⁶ IBM. Scaling Wimbledon’s video production of highlight reels through AI technology. 2017, disponível em: <https://www.ibm.com/blogs/research/2017/06/scaling-wimbledons-video-production-highlightreels-ai-technology/> acesso em: 20 jan. 2019.

corresponde ao eixo do paradigma enquanto o processo ao eixo do sintagma. Na articulação paradigmática, ou das seleções, são apresentadas as opções de escolha para o objeto semiótico, e no eixo do sintagma são encadeadas as escolhas feitas no paradigma. Portanto, ao escolher imagens da torcida ou de jogadas, a máquina seleciona no eixo do paradigma quais objetos vão compor melhor a sucessão de elementos do sintagma. Sua escolha é ponderada pelo suporte do texto, ou seja, um vídeo de melhores momentos com distribuição online.

Outro exemplo que relaciona o levantamento de pontos altos dentro de um vídeo é o trailer do filme *Morgan* (2016, EUA). *Morgan* é um filme que transita entre os gêneros de ficção científica e o suspense. No filme, a consultora corporativa Lee Weathers (Kate Mara) deve decidir se Morgan (Anya Taylor-Joy), uma criatura inteligente e artificial, deve ser sacrificada. Em uma parceria da *20th Century Fox* e a *IBM*, foi proposto que o *Watson* desenvolvesse o primeiro “trailer de filme cognitivo” do mundo. De acordo com Smith (2016), gerente de *multimedia and vision* da *IBM*, o *Watson* foi treinado a partir de 100 trailers de filmes do mesmo gênero de *Morgan*.

Os pontos de observação para o treinamento do computador podem ser divididos em três áreas: análise visual, sonora e de composição de cena. Na análise visual foram identificadas as pessoas, cenários e objetos, que formam categorias entre 24 emoções e rótulos. A análise sonora partiu de elementos como o tom de voz dos atores, trilhas sonoras e ruídos, buscando dimensionar indícios de passionalidade nas sequências. A análise de composição das cenas buscou compreender o local representado, a forma do ambiente e iluminação, encontrando um padrão entre os trailers de filmes de suspense analisados. Após a análise de todos os trailers, o *Watson* estava pronto para assistir ao filme *Morgan*. Entre os 132 minutos do longa-metragem, foram selecionados seis pelo computador, incluindo entre eles 10 pontos chaves.

Figura 1 - Captura de tela do trailer de *Morgan*. O gráfico apresenta pontos no filme em que existem mudanças nas emoções.



Fonte: Reprodução / Trailer Morgan

Podemos identificar nos elementos encontrados durante a primeira e segunda parte da análise de *Watson*, visual e sonora, a predominância da relação icônica, de referencialidade a partir da semiótica do mundo natural na qual a semiótica do audiovisual pode estabelecer uma relação referencial. Assim, o computador relaciona os elementos visuais e sonoros com um contexto já conhecido. A inferência, ao ser realizada pelo computador, baseia-se na forma da percepção humana, mas sem a valoração semântica, de modo que o resultado é proveniente apenas da repetição de padrões em relação ao *corpus* disponível. O terceiro nível da análise, de composição de cena, completa a matriz audiovisual, englobando as matrizes visual e a sonora. A *Fox* estima que um editor demoraria de 10 a 30 dias para fazer um trailer, ao passo que todo o processo de treinamento do *Watson* e seleção dos trechos foi realizado em 24 horas. Para Smith (2016) a máquina não vai substituir o humano criativamente, mas que exemplos como o trailer do filme *Morgan* possam ajudar a construir novas perspectivas no fazer criativo.

FERRAMENTAS DE PRODUÇÃO AUDIOVISUAL EM IA

Os recursos computacionais de análises visuais, como os utilizados pela *IBM* no longa-metragem *Morgan*, estão disponíveis para outros usuários através da internet. Entre as diversas empresas que oferecem esse tipo de ferramenta, podemos destacar a *Google*, com os serviços *Cloud Vision* e *Cloud Video Intelligence*. Os serviços oferecem ferramentas de análise e indexação de material em fotos e vídeos. Entre os objetivos dos serviços está o de ser simples. Segundo a própria descrição da *Cloud Vision*, desenvolvedores com pouco conhecimento em *machine learning*⁷ podem treinar modelos personalizados de alta qualidade. Tanto a *Cloud Vision*, que trabalha com imagens estáticas, quanto a *Cloud Video Intelligence*, utilizada para imagens em movimento, podem reconhecer elementos e entidades dentro de uma mídia visual. Dessa forma, conjuntos amplos, como animais, brinquedos e construções, podem ser categorizados pelos serviços. Além disso, a partir do treinamento da máquina, é possível criar categorias mais específicas de classificação dos elementos encontrados nas imagens, o que pode ser de grande utilidade para a análise de conteúdos audiovisuais.

Assim, partindo da premissa da possibilidade de desenvolvimento de forma autônoma por sistemas inteligentes realizamos dois experimentos em etapas da produção audiovisual para demonstrar o uso de ferramentas precursoras e de acesso amplo disponibilizadas pela *Google*. Essas tecnologias se caracterizam por permitir uma análise autônoma a partir de recursos de *machine*

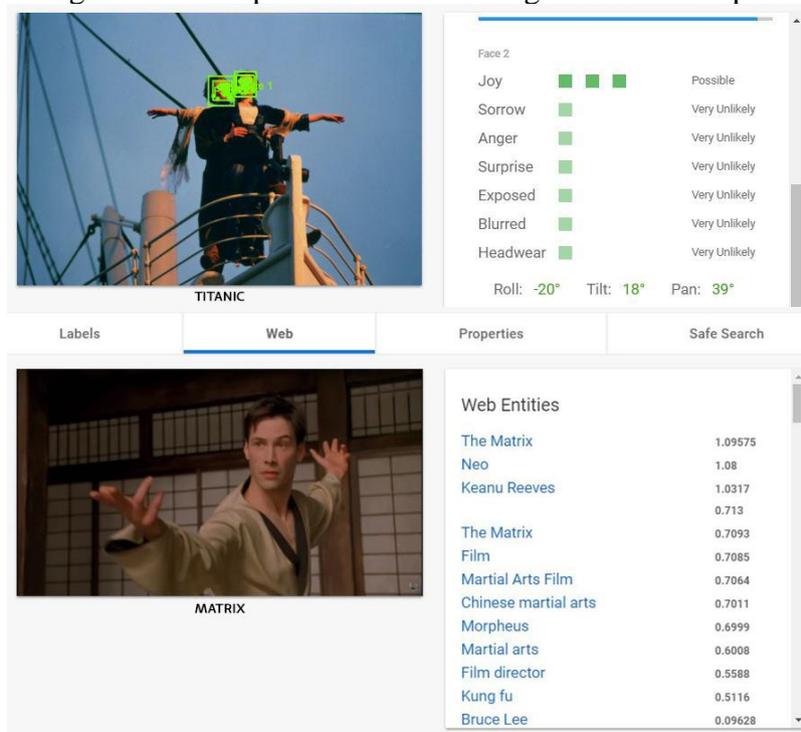
⁷ Machine Learning está no ramo da inteligência artificial responsável pela criação de algoritmos que aprendam a ler e compreender novos dados, analisando-os para determinar respostas dentro de um número finito de possibilidades

learning sem a necessidade de que o usuário tenha conhecimentos avançados de linguagem de programação informática, possibilitando, desse modo, que seja apropriada amplamente por usuários.

Além disso, observa-se que essas ferramentas da Google estão disponíveis de forma acessível, uma vez que o processamento de dados tem valor variável a partir da quantidade de análises realizadas. Em pequena escala, até mil análises mensais, o serviço é gratuito, e acima de cinco mil análises custa um dólar mensal. Outro fator que faz com que as ferramentas se destaquem nesse cenário de processamento de dados é a disponibilidade de suporte e tutoriais destinados aos usuários aprenderem a tecnologia desenvolvida pela Google, permitindo sua utilização de forma “simples”, como propõe o serviço. Cabe salientar que a plataforma de processamento imagético da Google, o Google Vision, lançada em 2015, inaugura uma prática que atualmente está incorporada nos mais diferentes serviços de processamento de imagens e vídeos disponibilizados ao usuário na internet, mais avançados e de acesso universalizado, principalmente na utilização *apps* para *smartphones*.

O primeiro experimento, realizado na *Google Vision*, consistiu em realizar uma análise de duas capturas de tela retiradas dos filmes *Titanic* (1997, EUA) e *Matrix* (1999, EUA), com o objetivo de demonstrar como operam as ferramentas e verificar quais categorias de classificação seriam encontradas.

Figura 2 - Montagem de telas após a análise das imagens dos filmes pela Google Vision



Fonte: Montagem dos autores.

O resultado da análise de cada imagem é dividido pelo sistema em seis partes. (1) Em sua primeira parte, são encontrados os rostos na imagem e a emoção que está associada a cada um. Quatro emoções estão disponíveis para a classificação: alegria, sofrimento, angústia e surpresa, que são graduadas de forma crescente em muito improvável, pouco provável, provável, muito provável. As emoções são apreendidas a partir da comparação de uma face da imagem com a dos bancos de dados da ferramenta. (2) A segunda parte da análise reconhece objetos presentes na cena. Esta é uma etapa mais geral do que será a terceira, destinada à verificação de rótulos. Dessa forma, foram reconhecidos na imagem do filme *Matrix* as figuras de pessoa e homem. (3) A análise se aprofunda na terceira etapa, em que são encontrados elementos metonímicos, como braço, mão e músculo, e outros rótulos que classificam a imagem, como fotografia, templo e kung fu. (4) A quarta etapa de análise utiliza a internet para classificar o conteúdo, encontrando termos de busca relacionados e imagens iguais ou semelhantes na rede. Esta etapa apresenta reiterações do conteúdo. Ainda é possível apreender alguns rótulos, semelhante ao caso da terceira etapa, como ficção científica e artes marciais. (5) A quinta etapa está relacionada a aspectos técnicos da imagem, ou, do ponto de vista semiótico, do formante da expressão cromático como a presença das cores predominantes, sugestões para o corte e a redução. (6) A sexta e última etapa realiza a análise em busca de conteúdos que possam ser classificados como somente para adultos ou conter imagens violentas. A classificação de conteúdos adultos muitas vezes é utilizada por redes sociais nas fotos de seus usuários, para a remoção de imagens postadas que possam violar as políticas de compartilhamento da comunidade.

O segundo experimento, realizado na *Google Vision autoML*, teve como objetivo criar as próprias categorias de análise. Dessa vez, o modelo não utilizou as emoções nos rostos ou a busca de termos relacionados na internet. As categorias desse modelo foram criadas apenas a partir do encontro de padrões das imagens de treino. Para o treinamento foram utilizados os 10 primeiros minutos do primeiro capítulo de três novelas em exibição simultânea da *Rede Globo de Televisão: Espelho da Vida* (09/2018 – 04/2019), *Verão 90* (02/2019 – 07/2019) e *O Sétimo Guardião* (11/2018 – 05/2019). Foram geradas capturas de tela a cada cinco segundos das novelas e aplicadas na plataforma, com um total de 360 imagens, 120 por novela. Dessa forma, foram selecionadas as imagens de cada uma das novelas, tornando-as uma categoria com o título da trama. Após o treinamento do sistema, imagens do capítulo do dia 11 de fevereiro de 2019 foram carregadas para que o computador identificasse a qual novela elas pertenciam. Com a maioria dos resultados positivos, o sistema foi capaz de prever com confiabilidade de até 96% que um quadro pertencia a novela *O Sétimo Guardião*.

Os quadros analisados para formar o banco podem ser divididos em três grupos: (1) os positivos, que serão marcados como o “cânone” dos quadros da categoria, (2) os falsos negativos,

que são quadros pertencentes a novela, mas que não apresentam índices de semelhança elevados com os outros da mesma categoria e, (3) os falsos positivos, que são quadros de outras novelas, mas que se assemelham muito com as daquela categoria. Os resultados falsos negativos e falsos positivos são eliminados da classificação, evitando problemas no aprendizado de máquina. Dessa forma, aprendendo apenas com os quadros positivos o sistema pode classificar de forma mais segura as novas imagens de entrada.

Figura 3 - Quadros da novela *O Sétimo Guardião* analisados identificados pelo modelo.



Fonte: Google Cloud AutoML

As capturas utilizadas para compor o banco da novela *O Sétimo Guardião* foram realizadas no capítulo do dia 12 de novembro de 2018. Já para a análise, foram utilizadas capturas do dia 11 de fevereiro de 2018. Mesmo com a grande distância temporal entre os capítulos, o modelo conseguiu identificar com alta precisão que se tratavam da mesma novela. Em um dos quadros para a análise está a figura do ator Marcelo Novaes, que não possui registro no acervo de quadros retirados do primeiro capítulo. Mesmo com a aparição do ator, o sistema conseguiu identificar a imagem da novela com 93,7% de certeza, em decorrência da reiteração da espacialidade reiterada pelo cenário.

CONSIDERAÇÕES FINAIS

Embora as tecnologias utilizadas para a geração de textos autônomos estejam em um patamar técnico avançado, seus resultados demandam efetivamente maior aprimoramento no tocante à produção de sentido no âmbito comunicacional. Acreditamos que os avanços contínuos em sistemas inteligentes, decorrentes da lógica do aprendizado de máquinas, serão uma constante promovendo

cada vez mais o aprimoramento da tecnologia e, por conseguinte, impondo a necessidade de adaptação para diversos setores produtivos, entre os quais o do audiovisual.

Experimentos realizados durante a segunda década do século XXI como os descritos e analisados neste trabalho, já apontam para os grandes desafios a serem enfrentados por profissionais diante do exercício de atividades relacionadas diretamente à criatividade, como a produção de roteiros e edição com a incorporação de ferramentas autônomas no sistema produtivo. No entanto, a lógica que rege a criação destes conteúdos audiovisuais segue parâmetros próprios, configurados com o propósito de identificação de padrões, que, de forma quantitativa, operam por entre alternativas para uma composição. Esta forma de criação traz implícita a contradição de entregar um resultado programático, regido pelos critérios estabelecidos por um sistema computacional, mas que, até o momento, mostra-se incapaz de dominar as linguagens por meio de um fazer efetivamente criativo.

Diante disso, considerando o avanço das tecnologias da IA no setor, além da perspectiva de alteração nos processos produtivos, é cada vez mais necessário observar com olhar crítico a origem, a qualidade e o uso dos dados que treinam estes sistemas. Além das questões éticas que permeiam as relações de comunicação, os desdobramentos e o alcance ainda pouco dimensionados da incorporação da inteligência artificial na sociedade alimentam debates em torno das perspectivas que o caráter disruptivo desta tecnologia representa. Estando permeados por imaginários simbólicos desde os primórdios das formulações teóricas acerca da inteligência artificial, tais debates impõem-se para que seja possível circunscrever as perspectivas das condições efetivas de apropriação cultural das IAs e seus desdobramentos nos processos de produção audiovisual.

REFERÊNCIAS

ASIMOV, Isaac. Runaround. 1942. In: ASIMOV, Isaac. I, Robot. Nova York: Gnome Press, 1950.

CHOMSKY, Noan. **Syntactic Structures**. 1ª ed. Alemanha: Mouton & Co. 1957.

FLOCH, Jean-Marie. **Alguns conceitos fundamentais em semiótica geral**. Tradução Analice Dutra Pilar. In: Documentos de estudos do centro de pesquisas sociosemióticas. São Paulo. 2001.

GUNKEL, David J. **Comunicação e inteligência artificial: novos desafios e oportunidades para a pesquisa em comunicação**. Galaxia (São Paulo, online), n. 34, jan-abr., 2017.

IBM. **Scaling Wimbledon's video production of highlight reels through AI technology**. 2017, disponível em: < <https://www.ibm.com/blogs/research/2017/06/scaling-wimbledons-video-productionhighlight-reels-ai-technology/> > acesso em: 20 jan. 2019.

MCCARTHY, John. et al. **A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence - 1955**. AI Magazine, v. 27, n. 4, jan/mar, 2006.

MCKEOWN, Lewis. **An evaluation of the Impact of Constraints on the Perceived Creativity of Narrative Generating Software**. University of Kent. Reino Unido. 2017.

MÉDOLA, Ana Silvia Lopes Davi. **Televisão digital brasileira e os novos processos de produção de conteúdos: os desafios para o comunicador**. ECompós, v.12 n.3, set/dez, 2009.

NORVING, Peter; RUSSEL, Stuart. **Inteligência artificial**. Tradução de Regina Célia Simille de Macedo. 3ª ed. Rio de Janeiro: Elsevier, 2013.

SMITH, John. R. **IBM Research Takes Watson to Hollywood with the First “Cognitive Movie Trailer”**. 2016, disponível em: < <https://www.ibm.com/blogs/think/2016/08/cognitive-movie-trailer/>> acesso em: 20 jan. 2019.

WAZLAWICK, Raul Sidnei. **História da computação**. 1 ed. Rio de Janeiro: Elsevier, 2016.

XAVIER, Ismail. **O discurso cinematográfico: opacidade e transparência**. 3 ed. São Paulo: Paz e Terra, 2005.

Filmografia

BLADE Runner. Direção: Ridley Scott. Produção: Michael Deeley. Estados Unidos: Warner Bros. Pictures. 1982. 1 DVD (117 min)

DIA em que a Terra Parou, O. Direção: Robert Wise. Produção: Julian Blaustein. Estados Unidos: 20th Century Fox. 1951. 1 DVD (132 min).

2001: Uma Odisseia no Espaço. Direção: Stanley Kubrick. Produção: Stanley Kubrick. Estados Unidos: Metro-Goldwyn-Mayer. 1951. 1 DVD (148 min)

EXTERMINADOR do Futuro, O. Direção: James Cameron. Produção: Gale Anne Hurd. Estados Unidos: Orion Pictures. 1984. 1 DVD (107 min)

INDEPENDENCE Day. Direção: Roland Emmerich. Produção: Dean Devlin. Estados Unidos: 20th Century Fox. 1996: 1 DVD (145 min)

MATRIX. Direção: Lilly Wachowski e Lana Wachowski. Produção: Joel Silver. Estados Unidos: Warner Bros. 1999. 1 DVD (136 min)

MORGAN. Direção: Luke Scott. Produção: Ridley Scott, Michael Schaefer e Mark Huffam. Estados Unidos: 20th Century Fox. 2016. 1 DVD (92 min)

SUNSPRING. Direção: Oscar Sharp. Produção: Andrew Kortschak, Walter Kortschak, Andrew Swett e Allison Friedman. Reino Unido: End Cue. 2016. 1 vídeo (9 min). Disponível em: <<https://www.youtube.com/watch?v=LY7x2lhqjmc>>. Acesso em 22 out. 2022.

Informações sobre o Artigo

Resultado de projeto de pesquisa, de dissertação, tese: Dissertação de mestrado: “Audiovisual e inteligência artificial: processos de geração do sentido em textos autônomos” de Vinícius Laureto de Oliveira.

Fontes de financiamento: Não se aplica.

Apresentação anterior: Não se aplica.

Agradecimentos/Contribuições adicionais: Não se aplica.

Ana Silvia Lopes Davi Médola

Docente do Programa de Pós-Graduação em Comunicação da Universidade Estadual Paulista – UNESP. Livre-Docente em Comunicação Televisual. Líder do GEA - Grupo de Estudos Audiovisuais.

E-mail: ana.silvia@unesp.br

ORCID: <https://orcid.org/0000-0003-2101-3727>

Vinicius Laureto de Oliveira

Mestre em Comunicação pelo Programa de Pós-Graduação em Comunicação da Universidade Estadual Paulista – UNESP.

E-mail: vinilaureto@gmail.com

ORCID: <https://orcid.org/0009-0000-5764-7022>

Henrique da Silva Pereira

Doutor em Comunicação pelo Programa de Pós-Graduação em Comunicação da Universidade Estadual Paulista. Vice-líder do GEA - Grupo de Estudos Audiovisuais. Docente dos cursos de Comunicação e Cinema do Centro Universitário Nossa Senhora do Patrocínio (CEUNSP).

E-mail: henrique.pereira@unesp.br

ORCID: <https://orcid.org/0000-0002-1328-8694>