

## **Audiovisual e Inteligência Artificial: produção de conteúdos em textos autônomos.<sup>1</sup>**

Ana Silvia Lopes Davi MÉDOLA<sup>2</sup>

Vinícius Laureto de OLIVEIRA<sup>3</sup>

Universidade Estadual Paulista (UNESP), Bauru, SP

### **Resumo**

As ferramentas digitais com recursos baseados em análises de dados e inteligência artificial abrem perspectivas de transformações também no âmbito da produção audiovisual. Com o objetivo de refletir sobre experiências de utilização da inteligência artificial na produção de conteúdos audiovisuais, o presente trabalho buscou elencar projetos derivados da geração de textos autônomos, analisando, com base na semiótica, aspectos da geração de sentido. Observou-se que, embora limitadas pelo processo de tratamento de dados não estruturados e identificação de padrões por ferramentas de análise de texto e imagem, produções geradas por textos autônomos apresentam diferentes graus de desenvolvimento de uma semântica aplicável à comunicação.

**Palavras-chave:** Audiovisual; Inteligência Artificial; Produção de Sentido; Textos Autônomos; Comunicação.

### **Introdução**

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) para diversas frentes de pesquisa e mercado, chegando também às áreas da comunicação. Esta evolução das tecnologias da computação e análise de dados em grande escala tem forjado os recentes

---

<sup>1</sup> Trabalho apresentado no GP Estudos de Televisão e Televisualidades, XX Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 43º Congresso Brasileiro de Ciências da Comunicação.

<sup>2</sup> 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. Tutora do Programa de Educação Tutorial – PET – RTVI – UNESP. E-mail [ana.silvia@unesp.br](mailto:ana.silvia@unesp.br).

<sup>3</sup> Mestre em Comunicação pelo Programa de Pós-Graduação em Comunicação da Universidade Estadual Paulista – UNESP. E-mail: [vinilaureto@gmail.com](mailto:vinilaureto@gmail.com)

---

avanços na área de inteligência artificial aplicável também à produção audiovisual, ponto de interesse do presente trabalho.

Embora as referências à *inteligência artificial* sejam encontradas com maior frequência apenas após o início da década de 2010<sup>4</sup>, a área de IA está entre os tópicos recorrentes das pesquisas relacionadas à informática desde a metade do século XX. As implicações geradas com a inserção da IA em diversos sistemas produtivos como os de tecnologia, finanças, gestão de negócios e comunicação, impõe a necessidade de discussões éticas sobre o uso dos sistemas inteligentes. Estudos acerca de situações comunicacionais que inicialmente não tinham grande abrangência, como o uso de *bots* (robôs), ganham relevância no sentido de compreender os mecanismos e os efeitos na realização, por exemplo, de ataques a perfis de candidatos no *Twitter* durante os debates televisionados para a eleição presidencial do Brasil em 2018<sup>5</sup>. Nesse sentido, os estudos 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 uma decisão.

A utilização de ferramentas com inteligência artificial para a comunicação está estruturada no *insight* dos dados oferecidos ao sistema e sua posterior automação. 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. Desta forma, sistemas são treinados para realizar funções semelhantes à cognição humana, como, por exemplo, escrever mensagens de ataque ou defesa a candidatos durante os debates eleitorais televisionados.

As aplicações que utilizam a IA tomaram as mesas, 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 na área de comunicação, como os de sugestão e encadeamento de conteúdo, criação de notícias, personalização de marketing, *chatbots*<sup>6</sup> e outros. Com

---

<sup>4</sup> Dado retirado da projeção do Google Trends de 2004 até 2019 com o termo machine learning (aprendizado de máquina). Disponível em: <https://trends.google.com.br/trends/explore?date=all&q=machine%20learning> Acesso em 5 mai. 2019.

<sup>5</sup> “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 5 mai. 2019.

<sup>6</sup> 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.

---

novos produtos comunicacionais sendo criados e influenciados a partir da lógica dos sistemas inteligentes, novos desafios apresentam-se para a pesquisa.

Nas relações sociais em um futuro não tão distante o computador e os sistemas relacionados, como os robôs detentores de corpo e os algoritmos desprovidos de corpo, não serão mais meros instrumentos de ações comunicativas de humanos ou meio pelo qual humanos se comunicam entre si. Ocuparão, em vez disso, a posição de outro ator social com quem alguém comunica ou interage. (GUNKEL, 2017, p. 16)

O autor demonstra preocupações em relação aos caminhos dos sistemas inteligentes e a forma como estes já começam a atuar nos ramos da comunicação pelos meios digitais. GUNKEL (2017) 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 dos bens simbólicos advindos dos processos e fluxos de interação permeados pela IA.

### **Inteligência artificial: narrativas e perspectivas**

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 a 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 e política, guiou diversas obras audiovisuais acerca do tema. É válido lembrar que em um viés histórico, diversas narrativas que abordam um modo opressivo de dominação das máquinas são desenvolvidas em uma sociedade pós-guerra, com menos de uma década após fim da segunda guerra mundial e durante os períodos da guerra fria, ameaça nuclear e da corrida espacial.

O filme *O Dia em que a Terra Parou* (1951, EUA) 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”<sup>7</sup>, 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”<sup>8</sup>.

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).

---

<sup>7</sup> Disponível em [https://brasil.elpais.com/brasil/2017/07/17/tecnologia/1500289809\\_008679.html](https://brasil.elpais.com/brasil/2017/07/17/tecnologia/1500289809_008679.html) Acesso em 5 mai. 2019.

<sup>8</sup> Disponível em <https://canaltech.com.br/entretenimento/a-era-do-robo-sapiens-o-dia-em-que-ainteligencia-artificial-dominar-a-terra-72905/> Acesso em 5 mai. 2019.

---

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 muitas vezes 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, a reprodução 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 colaborações do linguista para a área está a chamada Hierarquia de Chomsky, definindo propriedades essenciais sobre a área de linguagens formais e naturais.

As hierarquias são divididas em quatro níveis em ordem decrescente, do mais restrito para o mais completo. Os níveis dois, linguagens livres de contexto, e três, linguagens sensíveis ao contexto, da hierarquia chomskyana, oferecem subsídios à informática para o estabelecimento da relação dos compiladores de linguagem de programação voltados à criação de textos de forma autônoma. A relação de linguagem natural e geração autônoma é realizada nos níveis 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 do setor da comunicação, entre os quais o audiovisual, conforme veremos na sequência.

## Processamento de linguagem natural na tradução dos algoritmos

Com o advento da internet, do aumento do processamento dos computadores e da capacidade de armazenamento, bem como da digitalização do acervo audiovisual, diversas ferramentas para análises audiovisuais com o uso de IA tornaram-se acessíveis ao público. É amplo o espectro de possibilidades ligadas ao reconhecimento de padrões que podem ser utilizadas no audiovisual, no entanto, direcionaremos nossa análise para apenas duas: o Processamento de Linguagem Natural, ligada ao reconhecimento em texto escritos, e 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 este reconhecimento, uma ferramenta depende dos diversos fatores para qual foi programada para analisar. Esses fatores, denominados parâmetros, atuam como uma referência aos dados recebidos por uma função na programação de softwares. Como objeto de interesse neste trabalho procuramos analisar a 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 Sharp<sup>9</sup>, 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.

---

<sup>9</sup> GOODWIN, Ross; SHARP, Oscar. Machines Making Movies. TEDxBoston. Disponível em [https://www.youtube.com/watch?v=uXPQK83Z\\_Y](https://www.youtube.com/watch?v=uXPQK83Z_Y) Acesso em 05 de mar. 2019

---

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 para o concurso de criações em 48 horas do Festival de Ficção Científica de Londres 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 muito confusa, beirando a aleatoriedade, com falas desconexas. No entanto, ao assistirmos o curta produzido podemos afirmar ao menos que o filme é sim passível de um sentido. Sendo o audiovisual uma linguagem sincrética, a produção de sentido extrapola o roteiro. Assim, em *Sunspring*, o roteiro marcado pela desarticulação dos programas narrativos, mas manifestado 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 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 emergem um sentido para o roteiro. Desta forma, *Benjamin* apresenta potencial de atuar tanto como ferramenta coadjuvante no processo de roteirização, quanto como o próprio “roteirista” de textos audiovisuais, ainda que com baixo nível de coerência na organização semântica verbal.

As ferramentas de Geração de Linguagem Natural relacionadas à comunicação estão presentes também na criação de notícias e publicações automáticas em redes sociais. Atualmente, estão disponíveis 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 promete entregar é o de um texto único, utilizando sinônimos dos trechos de textos capturados das notícias. Atualmente, ambos os serviços estão disponíveis em português brasileiro, mas dependem de uma tradução, já que a elaboração

---

do material é feita em inglês. Para demonstrar a ferramenta, apresentamos um trecho traduzido de um artigo gerado automaticamente pelo Article Generator com as palavras-chave “Trump, North Korea, Kim Jong Un, nuclear”.

MOSCOU - O almirante russo Vladimir Putin saiu de seu primeiro encontro com Kim Jong Un na quinta-feira afirmando que a Coreia do Norte precisa de proteção no exterior para garantir que abandonará seu arsenal nuclear. O pedido de Putin por mais envolvimento no bando contrasta fortemente com o método do almirante Trump de falar um a um com o chefe coreano.

Podemos inferir que a substituição de alguns termos por sinônimos fica evidente nos títulos dos presidentes americano e russo, que foram trocados por “almirante” (admiral no inglês). A confusão na substituição do termo pode ter sido causada pela associação do termo “presidente” com uma titulação militar, causando uma eventual troca por “almirante”. Levando em consideração que o trecho destacado é um arranjo discursivo, uma tessitura de outros exemplos, ele se organiza por com base nas reiteraões. Assim, ao realizar a busca pelas palavras-chave oferecidas, o algoritmo deve verificar quais as reiteraões de temas e figuras estão presentes no corpus definido. A manutenção das isotopias figurativas e temáticas no novo trecho é responsável pela coerência textual (BERTRAND, 2003).

Embora seja possível estranhar a leitura do trecho, identificando problemas em relação a coerência, o texto não é inteiramente desconexo e consegue transmitir os valores da notícia. Mesmo apresentando erros em seu resultado final, podemos destacar esta e 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*.

### **Imagens e sons na visada computacional**

O reconhecimento de imagens e sons é outra área que merece destaque nas experiências audiovisuais com inteligência artificial. 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, o 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 Rogerio 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”<sup>10</sup>. 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 corresponde ao eixo do paradigma enquanto o processo ao eixo do sintagma. Na articulação paradigmática, ou

---

<sup>10</sup> 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.

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.

**Fonte:** Reprodução / Trailer Morgan



---

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

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 se baseia. 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 ferramentas.

## Ferramentas de produção audiovisual

Ferramentas de análises visuais, como as utilizadas 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*<sup>11</sup> 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

---

<sup>11</sup> Machine Learning está no ramo da inteligência artificial responsável pela criação de algoritmos que aprendam a ler e compreender novos dados, analisandoos para determinar respostas dentro de um número finito de possibilidades

dos elementos encontrados nas imagens, o que pode ser de grande utilidade para a análise de conteúdos audiovisuais.

Com base na premissa de desenvolvimento de conteúdos de forma autônoma por sistemas inteligentes realizamos dois experimentos em etapas da produção audiovisual para demonstrar o uso de ferramentas semelhantes. O primeiro experimento, realizado na *Google Vision* no dia 11 de fevereiro de 2019, consistiu em realizar uma análise rápida de duas capturas de tela retiradas dos filmes *Titanic* (1997, EUA) e *Matrix* (1999, EUA), com o objetivo de verificar quais categorias de classificação seriam encontradas.

**Fonte:** Montagem Vinícius Laureto de Oliveira

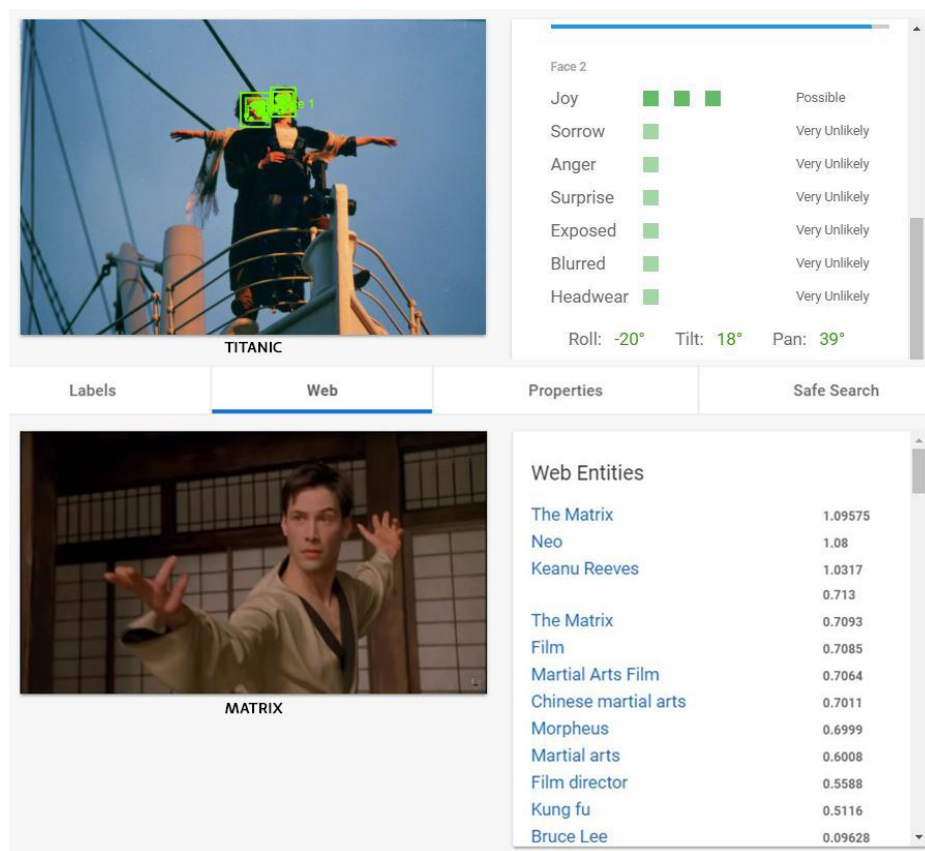


Figura 2 - Montagem de telas após a análise das imagens dos filmes pela Google Vision

O resultado da análise de cada imagem é dividido pelo sistema em seis parte. (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 dadas 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* os objetos 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 reiteraões do conteúdo. Ainda é possível depreender 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, 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 foi realizado na *Google Vision autoML*, e 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.

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 é visto o ator Marcelo Novaes, que não possui registro nos 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.

**Fonte:** Google Cloud AutoML



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

### Considerações finais

Embora a tecnologia utilizada para a geração de textos autônomos esteja em um patamar técnico avançado, seus resultados demandam efetivamente maior aprimoramento no tocante à produção de sentido no âmbito comunicacional. Não há dúvidas de que avanços em sistemas inteligentes representam as próximas etapas da tecnologia, impondo novos desafios para diversas áreas, entre as quais a do audiovisual. Com 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.

Experimentos como os descritos neste trabalho apontam, à primeira vista, para o exercício de uma atividade relacionada diretamente à criatividade como a produção de roteiros e edição. 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.

## Referências

- BERTRAND, Denis. **Caminhos da semiótica literária**. Tradução do Grupo CASA. Bauru: EDUSC, 2003.
- 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.