

Lev Manovich

City University of New
York

Emanuele Arielli

Universidade IUAV de
Veneza

Tradução:

Bernardo Girauta



*Este trabalho está licenciado
sob uma licença [Creative
Commons Attribution 4.0
International License](https://creativecommons.org/licenses/by/4.0/).*

Copyright (©):

*Aos autores pertence o direito
exclusivo de utilização ou
reprodução*

ISSN: 2175-8689

Imagens IA e mídias generativas: notas sobre a revolução em curso¹²

*AI image and Generative Media:
Notes on Ongoing Revolution*

Manovich, L., & Arielli, E. Imagens IA e mídias generativas:
notas sobre a revolução em curso. Revista Eco-Pós,
26(2), 16–39. [https://doi.org/10.29146/eco-
ps.v26i2.28175](https://doi.org/10.29146/eco-ps.v26i2.28175)

¹ Capítulo 5 de *Artificial Aesthetics* por Lev Manovich e Emanuele Arielli. <<http://manovich.net/index.php/projects/artificial-aesthetics-book>> . Publicado originalmente em 20 de abril de 2023.

² A versão anterior de uma parte deste capítulo foi publicado em *Diffusions – Taxonomy of Synthetic Imaginations in Architecture* , ed. Matias del Campo, Willey, 2023; e uma versão mais curta de outra parte foi publicada na MoMA Magazine (<https://www.moma.org/magazine/>), 2023. Notas sobre a escrita deste capítulo no idioma original (inglês): ele foi escrito sem o uso de ferramentas de IA; na fase de edição, foi utilizado o quillbot (<https://quillbot.com/>) para oferecer versões parafraseadas das frases dos autores.

Dossiê **Visualidades: estéticas, mídias e contemporaneidade** - <https://revistaecopos.eco.ufrj.br/>

ISSN 2175-8689 – v. 26, n. 2, 2023

DOI: 10.29146/eco-ps.v26i2.28175

RESUMO

O artigo discute a transição da criação de imagens manuais e a partir de lentes para o procedimento preditivo de IA generativa, em que os computadores geram novas mídias ao aprender com grandes conjuntos de dados e prever novos artefatos. Esta mudança da representação para a previsão marca uma alteração significativa no modo como as imagens são produzidas, movendo-se para as “mídias preditivas”. O autor também investiga as características comuns entre os processos de criação humana e de IA, em que ambos aprendem com grandes quantidades de dados culturais para produzir novos artefatos. No entanto, a IA luta atualmente para gerar “átomos culturais” únicos ou raros, que são melhor compreendidos pelos criadores humanos. O artigo também coloca a IA generativa no contexto da arte moderna e da história da mídia. Os artefatos gerados por IA dependem de um grande arquivo de outros artefatos de mídia. Este mecanismo generativo liga os meios generativos a gêneros, métodos e movimentos artísticos anteriores, como a foto-colagem da década de 1920, a pop art da década de 1960 ou a net art da década de 1990.

PALAVRAS-CHAVE: *Inteligência artificial; mídia generativa; novas mídias; era digital;*

ABSTRACT

The article discusses the transition from manual and lens-based image creation to Generative AI's predictive procedure, where computers generate new media by learning from large datasets and predicting new artifacts. This shift from representation to prediction marks a significant change in how images are produced, moving towards 'predictive media.' The author also investigates the common features between human and AI creation processes, where both learn from vast amounts of cultural data to produce new artifacts. However, AI currently struggles with generating highly unique or rare 'cultural atoms,' which are better grasped by human creators. The article also places Generative AI in the context of modern art and media history. Artifacts generated by AI rely on a large archive of other media artifacts. This generative mechanism links generative media to earlier art genres, methods and movements such as photo-collage of the 1920s, pop art of the 1960s or net art of the 1990s.

KEYWORDS: *Artificial intelligence; generative media; new media; digital age;*

Submetido em 15 de maio de 2023

Aceito em 26 de outubro de 2023

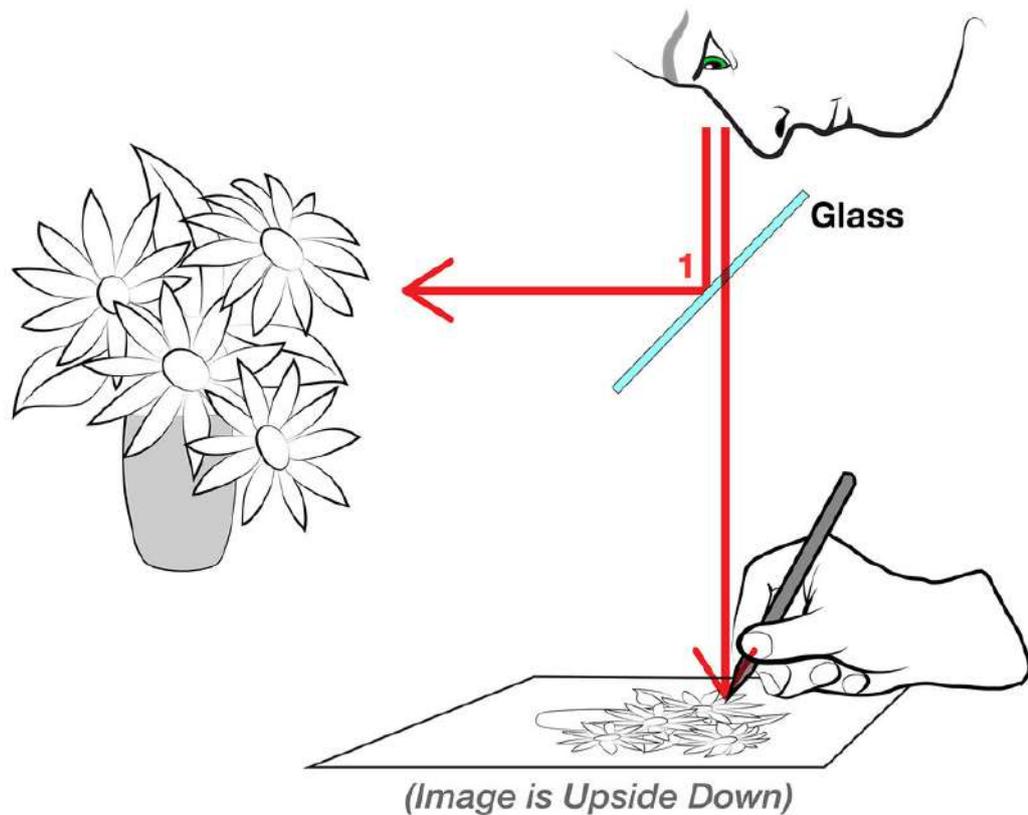


Fig. 1 - Princípio básico da câmera lúcida, aparelho óptico de desenho amplamente utilizado por artistas e estudantes de arte no século XIX. Fonte: <https://neolucida.com/history>.

Parece que nós estamos no início de uma verdadeira revolução na criação de mídia: a ascensão das “mídias generativas”. Eu uso ferramentas de computador para arte e design desde 1984 e vi algumas grandes revoluções de mídias, incluindo a introdução dos computadores Mac e aplicativos de desktop para criação e edição de mídias, o desenvolvimento de gráficos e animações 3D fotorrealistas, a ascensão da web em 1993 e a ascensão das redes sociais a partir de 2006. A nova revolução IA das “mídias generativas” parece ser tão significativa quanto qualquer uma destas. De fato, é possível que seja tão significativa quanto a invenção da fotografia no século XIX ou a adoção da perspectiva linear na arte ocidental no século XVI.

(Se você é novo neste assunto, aqui vai uma breve história. A revolução das mídias generativas esteve em desenvolvimento por mais de 20 anos. Os primeiros artigos sobre IA propondo que o vasto universo não estruturado da web de textos, imagens e outros artefatos culturais poderiam ser usados para treinar computadores para fazer várias tarefas já apareciam em 1999-2001. Em 2015, os métodos de “*deep dreaming*” e “transferência de estilo”, do Google, atraíram muita atenção: de repente, os computadores poderiam criar novas imagens artísticas imitando estilos de muitos artistas famosos. O lançamento do DALL-E em janeiro de 2021 foi outro marco: os computadores podiam, então, sintetizar imagens a partir de descrições de texto. O Midjourney, o Stable Diffusion e o DALL-E 2 contribuíram para a aceleração dessa evolução, em 2022. Desde então, as imagens sintéticas podem ter muitas estéticas que vão desde o fotorrealismo a qualquer tipo de meio físico ou digital, incluindo mosaicos, pinturas a óleo, fotografia de rua ou renderização 3D CG. O código para produzir essas imagens, referido como “modelo” no campo da inteligência artificial, foi tornado público em agosto de 2022, provocando uma enxurrada de experimentos e um desenvolvimento acelerado.)

Neste e nos próximos capítulos, descreverei várias características das *mídias generativas visuais* em suas formas atuais que acredito serem particularmente significativas ou novas. Alguns dos meus argumentos também se aplicam às mídias generativas em geral, mas a maioria se concentra nas *mídias visuais* –, refletindo minha própria experiência de uso de algumas ferramentas populares de imagens IA, como Midjourney e Stable Diffusion (e às vezes também Runway ML), quase todos os dias de meados de 2022 até o início de 2023. Mas, primeiro, vamos definir os termos principais.

Os termos

Neste texto, “artista” ou “criador” refere-se a qualquer pessoa qualificada que cria objetos culturais em quaisquer mídias ou suas combinações. Os termos “mídias generativas”, “mídias IA”, “IA generativa” e “mídias sintéticas” são todos intercambiáveis. Eles se referem ao processo de criação de novos objetos de mídia

com redes neurais profundas, como imagens, animação, vídeo, texto, música, modelos e cenas 3D e outros tipos de mídias. Além desses objetos, as redes neurais também são usadas para gerar elementos e tipos de conteúdo específicos, como rostos humanos fotorrealistas e poses e movimentos humanos. Elas também podem ser usadas na edição de mídias, como para substituir uma parte de uma imagem ou vídeo por outro conteúdo que se ajuste espacialmente.

Essas redes são treinadas em vastas coleções de objetos de mídias já existentes. Tipos populares de redes neurais artificiais para geração de mídias incluem modelos de difusão, modelos de texto-para-imagem, redes adversárias generativas (GAN) e transformadores. Para a geração de imagens estáticas e em movimento usando redes neurais, os termos geração de imagem, imagem sintética e imagens IA podem ser usados de forma intercambiável.

Observe que a palavra “generativo” também pode ser usada de diferentes maneiras para significar a criação de artefatos culturais usando qualquer processo algorítmico (não apenas redes neurais) ou mesmo um processo baseado em regras que não usa computadores. É assim que os termos “arte generativa” e “design generativo” costumam ser usados hoje na cultura popular e nos meios de comunicação. Neste capítulo, eu uso “generativo” em um sentido mais restrito para me referir a métodos de rede profunda e aplicativos de geração de mídia que utilizam tais métodos.

“IA” como uma percepção cultural

Não existe uma tecnologia específica ou um único projeto de pesquisa chamado “IA”. No entanto, podemos acompanhar como evoluiu a nossa percepção cultural deste conceito ao longo do tempo e a que ele se referia em cada período. Nos últimos cinquenta anos, quando uma suposta habilidade exclusivamente humana é automatizada por meio de tecnologias de computação, referimo-nos a ela como “IA”. No entanto, assim que essa automação se torna perfeita e totalmente bem-sucedida, tendemos a parar de nos referir a ela como um “caso de IA”. Em outras palavras, “IA” se refere a tecnologias e metodologias que automatizam as habilidades cognitivas

humanas e estão começando a funcionar, mas ainda não chegaram lá. A “IA” já estava presente nas primeiras ferramentas de mídias computacionais. O primeiro sistema interativo de desenho e design, o *Sketchpad*, de Ivan Sutherland (1961-1962), tinha um recurso que finalizava automaticamente qualquer retângulo ou círculo que você começasse a desenhar. Em outras palavras, ele sabia o que você estava tentando fazer. No entendimento bastante amplo que acabamos de oferecer, isso já era, sem dúvidas, “IA”.

Minha primeira experiência com um programa de pintura de desktop rodando em um Apple II foi em 1984, e foi realmente incrível mover o mouse e ver pinceladas simuladas de tinta aparecerem na tela. No entanto, hoje não consideramos mais isso como 'IA'. Outro exemplo seria a função do Photoshop que seleciona automaticamente o contorno de um objeto. Essa função foi adicionada há muitos anos – isso também é “IA”, em sentido amplo, mas ninguém se referiria a ela como tal hoje. A história dos sistemas e ferramentas de mídias digitais está repleta desses “momentos de IA” – incríveis no início, depois dados como certos e posteriormente esquecidos enquanto “IA”. (Nos livros de história da IA, esse fenômeno é chamado de “efeito IA”.) No momento atual, “IA criativa” refere-se apenas a métodos desenvolvidos recentemente em que os computadores transformam alguns *inputs* (entradas) em novos *outputs* (saídas) de mídias (por exemplo, modelos de texto-para-imagem) e técnicas específicas (por exemplo, certos tipos de redes neurais profundas). No entanto, devemos lembrar que esses métodos não são os primeiros nem os últimos na história e no futuro a simular habilidades artísticas humanas ou a auxiliar os humanos na criação de mídias.

Da representação à previsão

Historicamente, os humanos criaram imagens de cenas existentes ou imaginadas por vários métodos, desde o desenho manual até 3D CG (veja abaixo a explicação dos métodos). Com as mídias generativas IA, surge um método fundamentalmente novo. Computadores usam grandes conjuntos de dados de

representações existentes em várias mídias para prever novas imagens (paradas e animadas).

Pode-se certamente propor diferentes caminhos históricos que levam às mídias generativas visuais de hoje, ou dividir uma linha do tempo histórica em diferentes estágios – aqui está uma dessas possíveis trajetórias:

1. Criar representações manualmente (por exemplo, desenhos com uma variedade de instrumentos, esculturas, etc). Etapas e peças mais mecânicas às vezes eram executadas por assistentes humanos que normalmente treinavam no ateliê de seus professores – então já existia alguma delegação de funções.

2. Criar manualmente, mas usando dispositivos auxiliares (por exemplo, máquinas de perspectiva, *camera lucida*). De *mãos para mãos + dispositivo*. Assim, algumas funções são delegadas a dispositivos mecânicos e ópticos.

3. Fotografia, raio x, vídeo, captura volumétrica, sensoriamento remoto, fotogrametria. De *usar mãos para registrar informações usando máquinas*. De *assistentes humanos para assistentes máquinas*.

4. CG 3D. Você define um modelo 3d em um computador e usa algoritmos que simulam efeitos de fontes de luz, sombras, neblina, transparência, translucidez, texturas naturais, profundidade de campo, *motion blur*, etc. De *gravação à simulação*.

5. IA generativa. Uso de conjuntos de dados de mídia para prever imagens estáticas e em movimento. De *simulação à previsão*.

“Previsão” é, de fato, o termo frequentemente usado por pesquisadores de IA em suas publicações que descrevem métodos de mídias generativas visuais. Portanto, embora esse termo possa ser usado de forma figurada e evocativa, isso também é o que realmente acontece cientificamente quando você usa ferramentas geradoras de imagens. Ao trabalhar com um modelo de IA texto-para-imagem, a rede neural tenta prever as imagens que melhor correspondem ao seu *input* de texto. Certamente, não estou sugerindo que usar todos os outros termos já aceitos, como “mídias generativas”, seja inapropriado. Mas se quisermos entender melhor a

diferença entre os métodos de síntese de mídias visuais IA e outros métodos de representação desenvolvidos na história humana, empregar o conceito de “previsão” e, assim, referir-se a esses sistemas IA como “mídias preditivas” pode capturar bem a diferença.

Traduções de mídias

Existem vários métodos para criar “mídias IA”. Um método transforma o *input* humano de mídia, mantendo o mesmo tipo de mídia. O texto inserido pelo usuário, por exemplo, pode ser resumido, reescrito, expandido e assim por diante. O *output*, como o *input*, é um texto. Alternativamente, no método de geração imagem-para-imagem, uma ou mais imagens de entrada são usadas para gerar novas imagens. No entanto, há um outro caminho igualmente intrigante do ponto de vista histórico e teórico. As “mídias IA” podem ser criadas ao se “traduzir” automaticamente o conteúdo entre tipos distintos de mídias. Como não se trata de uma tradução literal de um para um, coloquei a palavra “tradução” entre aspas. Em vez disso, o *input* de um meio instrui uma rede neural a prever o *output* apropriada de outro meio. Esse *input* também pode ser referida como “mapeada” para determinados *outputs* em outras mídias. Textos são mapeados em novos estilos de textos, imagens, animações, vídeos, modelos 3D e música. Vídeos são convertidos em modelos 3D ou animações. As imagens são “traduzidas” em texto e assim por diante. A tradução com o método texto-para-imagem é atualmente mais avançada do que as outras, mas várias formas irão alcançá-la eventualmente.

A tradução (ou mapeamento) entre uma mídia e outra não é um conceito novo. Tais traduções foram feitas manualmente ao longo da história humana, muitas vezes com intenção artística. Romances foram adaptados para peças de teatro e filmes, histórias em quadrinhos foram adaptadas para séries de televisão, um texto ficcional ou não ficcional foi ilustrado com imagens, etc. Cada uma dessas traduções foi um ato cultural deliberado que requeria habilidades profissionais e conhecimento das mídias apropriadas. Algumas dessas traduções podem agora ser realizadas automaticamente em escala massiva graças às redes neurais artificiais, tornando-se

um novo meio de comunicação e criação de cultura. É claro que a adaptação artística de um romance para um filme por uma equipe humana e a geração automática de recursos visuais a partir de um texto de romance por uma rede não são a mesma coisa, mas para muitos casos mais simples, a tradução automática de mídias pode funcionar bem. O que antes era um ato artístico qualificado, agora é uma capacidade tecnológica disponível para todos. Podemos ficar tristes com tudo o que pode ser perdido como resultado da automação – e democratização – dessa operação cultural crítica: habilidades, algo que se poderia chamar de “originalidade artística profunda” ou “criatividade profunda”, e assim por diante. No entanto, qualquer perda desse tipo pode ser apenas temporária se as habilidades das “culturas IA” forem, por exemplo, ainda mais aprimoradas para gerar mais conteúdo original e entender melhor os contextos.

Como a maioria das pessoas em nossa sociedade pode ler e escrever em pelo menos um idioma, os métodos de mídia texto-para-outro são atualmente os mais populares. Eles incluem modelos texto-para-imagem, texto-para-animação, texto-para-3D e texto-para-música. Essas ferramentas de IA podem ser usadas por qualquer pessoa que saiba escrever ou que utilize um software de tradução prontamente disponível para criar um *prompt* em um idioma que essas ferramentas entendam melhor, como o inglês. No entanto, outros mapeamentos de mídia podem ser igualmente interessantes para criadores profissionais. Ao longo da história cultural humana, várias traduções entre tipos de mídias atraíram atenção. Dentre elas, encontra-se traduções entre vídeo e música (cultura *clubber*); longas narrativas literárias transformadas em filmes e séries de televisão; quaisquer textos ilustrados com imagens em suportes diversos, como gravuras; números transformados em imagens (arte digital); textos que descrevem pinturas (tradição da *ekphrasis*, que começou na Grécia Antiga), mapeamentos entre sons e cores (especialmente populares na arte modernista); etc.

O desenvolvimento contínuo de modelos IA para mapeamentos entre todos os tipos de mídias, sem privilegiar o texto, tem potencial para ser extremamente frutífero, e espero que mais ferramentas consigam fazer isso. Essas ferramentas

poderiam ser usadas sozinhas ou em conjunto com outras ferramentas e técnicas e serão úteis tanto para artistas profissionais quanto para outros criadores. No entanto, sendo eu mesmo um artista, não estou afirmando que a futura “cultura IA” será capaz de igualar, por exemplo, interpretações inovadoras de Hamlet por diretores de teatro de vanguarda como Peter Brook ou filmes abstratos surpreendentes de Oscar Fishinger que exploraram correspondências musicais e visuais. É suficiente que novas ferramentas IA de mapeamento de mídias estimulem nossa imaginação, nos forneçam novas ideias e nos permitam explorar inúmeras variações de designs específicos.

O comum e o único

Tanto o processo moderno de criação humana quanto o processo preditivo das mídias generativas IA parecem funcionar de maneira semelhante. Uma rede neural é treinada usando coleções não estruturadas de conteúdo cultural, como bilhões de imagens e suas descrições ou trilhões de páginas da web e de livros. A rede aprende associações entre as partes constituintes desses artefatos (como quais palavras frequentemente aparecem próximas umas das outras), bem como seus padrões e estruturas comuns. A rede treinada usa, então, essas estruturas, padrões e “átomos de cultura” para criar novos artefatos quando pedimos. Dependendo do que pedimos, esses artefatos criados por IA podem ou não se parecer muito com aquilo que já existe.

De modo semelhante, nossa vida é um processo contínuo de treinamento cultural supervisionado e não supervisionado. Fazemos cursos de arte e história da arte, visualizamos sites, vídeos, revistas e catálogos de exposições, visitamos museus e viajamos para absorver novas informações culturais. E quando nos “incitamos” a criar alguns artefatos culturais novos, nossas próprias redes neurais biológicas (infinitamente mais complexas do que qualquer rede IA até hoje) geram tais artefatos com base no que aprendemos até então: padrões gerais que observamos, modelos para fazer coisas específicas (como desenhar uma cabeça humana com proporções corretas ou editar um vídeo de entrevista) e, muitas vezes,

partes concretas de artefatos já existentes. Em outras palavras, nossas criações podem conter tanto réplicas exatas de artefatos observados anteriormente quanto coisas novas que representamos usando modelos que aprendemos, como combinações de cores e a perspectiva linear. Além disso, os modelos humanos e de IA frequentemente têm um estilo “casa” padrão (o termo usado pelos desenvolvedores do MidJourney). Se um estilo não for especificado explicitamente, a IA irá gerá-lo usando essa estética “padrão”. Uma descrição do meio, o tipo de iluminação, as cores e o sombreamento e/ou uma frase como “no estilo de”, seguida do nome de um artista, ilustrador, fotógrafo, designer de moda ou arquiteto bastante conhecidos, são exemplos de especificações que podem evitar esse padrão.

Por poder simular dezenas de milhares de estéticas e estilos já existentes e interpolar entre eles para criar novos híbridos, a IA é mais capaz do que qualquer criador humano nesse aspecto. No entanto, atualmente, criadores humanos qualificados e altamente experientes também têm uma vantagem significativa. Tanto os humanos quanto a inteligência artificial são capazes de imaginar e representar objetos e cenas inexistentes e existentes. No entanto, ao contrário dos geradores de imagens IA, as imagens feitas pelo homem podem incluir conteúdo muito particular, detalhes minúsculos únicos e uma estética distinta que atualmente está além das capacidades da IA. Em outras palavras, hoje um grande grupo de ilustradores, fotógrafos e designers altamente qualificados e experientes pode representar tudo o que uma rede neural treinada pode fazer (embora demore muito mais), mas também pode visualizar objetos e composições e usar estéticas que a rede neural não pode neste momento (ou que pelo menos tem muita dificuldade em fazer de forma consistente).

Qual é a causa dessa lacuna estética e de conteúdo entre criadores humanos e artificiais? “Átomos culturais”, estruturas e padrões nos dados de treinamento que ocorrem com mais frequência são aprendidos com muito sucesso durante o processo de treinamento de uma rede neural artificial. Na “mente” de uma rede neural, eles ganham mais importância. Por outro lado, “átomos” e estruturas que são raras nos dados de treinamento, ou que aparecem apenas uma vez, dificilmente são

aprendidas ou nem mesmo analisadas. Eles não entram no universo de cultura artificial tal como aprendido pela IA. Conseqüentemente, quando pedimos à IA para sintetizá-los, ela é incapaz de fazê-lo.

Devido a isso, IAs de conversão texto-para-imagem, como o Midjourney, o Stable Diffusion ou o RunwayML, não são capazes de gerar desenhos no meu estilo, expandir meus desenhos adicionando peças recém-geradas ou substituir partes específicas dos meus desenhos por novos conteúdos desenhados em meu estilo (por exemplo, eles não podem executar “*outpainting*” ou “*inpainting*” úteis nas fotos digitais dos meus desenhos.) Em vez disso, essas ferramentas de IA geram objetos mais genéricos do que os que costumo desenhar ou produzem algo que é meramente ambíguo, mas desinteressante.

Certamente não estou afirmando que o estilo e o mundo mostrados em meus desenhos são completamente únicos. Eles também são resultado de encontros culturais específicos que tive, coisas que observei e coisas que notei. Mas, por serem incomuns (e, portanto, imprevisíveis), a IA acha difícil simulá-los, pelo menos sem treinamento adicional usando meus desenhos.

Aqui encontramos o maior obstáculo que enfrentamos como criadores no uso de mídias generativas IA. Frequentemente, a IA gera novos artefatos de mídia que são mais genéricos e estereotipados do que pretendíamos. Isso pode afetar quaisquer dimensões da imagem – elementos de conteúdo, iluminação, hachura, atmosfera, estrutura espacial e detalhes de formas 3D, entre outros. Ocasionalmente, trata-se de algo imediatamente aparente, caso em que você pode tentar realizar correções ou desconsiderar os resultados. Muitas vezes, no entanto, tais “*substituições*” são tão sutis que não podemos detectá-las sem observação extensiva ou, em alguns casos, o uso de um computador para analisar quantitativamente numerosas imagens. Em outras palavras, os novos modelos de mídias generativas IA, bem como a disciplina da estatística desde seu início no século XIX e o campo da ciência de dados desde o final da década de 2010, lidam bem com itens e padrões de ocorrência frequente nos dados, mas não sabem o que fazer com o que é pouco frequente e incomum. Podemos esperar que os pesquisadores de IA sejam capazes de resolver esse

problema no futuro, mas parece fundamental que não antecipemos uma solução imediatamente.

Assunto e estilo

Nas artes, a relação entre “conteúdo” e “forma” é amplamente discutida e teorizada. Esta breve seção não tenta se engajar em todos esses debates ou iniciar discussões com todas as teorias relevantes. Em vez disso, gostaria de considerar como esses conceitos funcionam na “cultura generativa” IA. No entanto, em vez de usar conteúdo e forma, usarei um par diferente de termos que são mais comuns em publicações de pesquisa de IA e conversas online entre usuários: **assunto** e **estilo**.

À primeira vista, as ferramentas de mídias IA parecem capazes de distinguir claramente entre o assunto e o estilo de qualquer representação. Em modelos texto-para-imagem, por exemplo, você pode gerar inúmeras imagens do mesmo assunto. Adicionar os nomes de artistas, mídias, materiais e períodos históricos da arte específicos basta para que o mesmo assunto seja representado de maneira diferente para corresponder a essas referências. Os filtros do Photoshop começaram a diferenciar entre assunto e estilo já na década de 1990, mas as ferramentas de mídias generativas IA têm maior capacidade. Por exemplo, se você especificar “pintura a óleo” em seu *prompt*, as pinceladas simuladas irão variar em tamanho e direção em uma imagem gerada com base nos objetos representados. As ferramentas de mídias IA parecem “entender” a semântica da representação em oposição aos filtros anteriores que simplesmente aplicavam a mesma transformação a cada parte da imagem, independentemente de seu conteúdo. Por exemplo, quando usei “uma pintura de Malevich” e “uma pintura de Bosch” no mesmo *prompt*, Midjourney gerou uma imagem de espaço que continha formas abstratas semelhantes às de Malevich, bem como muitas pequenas figuras humanas e animais, como nas populares pinturas de Bosch, dimensionadas adequadamente para perspectiva.

Além de representar o que solicitei, as ferramentas IA rotineiramente adicionam a uma imagem conteúdos não especificados em meu *prompt* de texto. Isso

ocorre frequentemente quando o texto do *prompt* inclui “no estilo de” ou “por” seguido do nome de um artista visual ou fotógrafo renomado. Em um experimento, usei o mesmo *prompt* com a ferramenta de imagens IA Midjourney 148 vezes, cada vez adicionando o nome de um fotógrafo diferente. O assunto do *prompt* permaneceu praticamente o mesmo – uma paisagem vazia com alguns prédios, uma estrada e postes elétricos com fios se estendendo no horizonte. Às vezes, adicionar o nome de um fotógrafo não tinha efeito algum sobre os elementos de uma imagem gerada que se encaixam em nosso conceito intuitivo de estilo, como contraste, perspectiva e atmosfera. Mas, de vez em quando, o Midjourney também modificava o conteúdo da imagem. Por exemplo, quando trabalhos bem conhecidos de um determinado fotógrafo apresentam figuras humanas em poses específicas, a ferramenta ocasionalmente adiciona essas figuras às minhas fotografias (como Malevich e Bosch, que foram transformados para se adequar à composição espacial da paisagem em vez de duplicados mecanicamente). O Midjourney também muda, às vezes, o conteúdo da minha imagem para corresponder a um período histórico em que um determinado fotógrafo produziu suas fotografias mais conhecidas.

De acordo com as minhas observações, quando pedimos ao Midjourney ou a uma ferramenta semelhante para criar uma imagem no estilo de um artista específico, e o assunto que descrevemos no *prompt* está relacionado aos assuntos típicos do artista, os resultados podem ser muito bem-sucedidos. No entanto, quando o assunto de nosso *prompt* e as imagens desse artista são muito diferentes, a “renderização” do assunto nesse estilo frequentemente falha.

Em resumo, para simular com sucesso um determinado estilo visual usando as ferramentas IA atuais, pode ser necessário alterar o conteúdo que você pretende representar. Nem todo assunto pode ser renderizado com sucesso e satisfatoriamente em qualquer estilo. Essa observação, acredito, complica a oposição binária entre os conceitos de “conteúdo” e “estilo”. Para alguns artistas, a IA pode extrair seu estilo de exemplos de seu trabalho e aplicá-lo a diferentes tipos de conteúdo. Mas para outros artistas, ao que parece, seu estilo e conteúdo não podem ser separados.

Para mim, tais tipos de observações e de pensamentos subsequentes estão entre as razões mais importantes para usar novas tecnologias de mídias, como mídias generativas IA, e aprender como elas funcionam. É claro que, como um teórico das mídias, venho pensando sobre as relações entre assunto e estilo (ou conteúdo e forma) há muito tempo, mas ser capaz de conduzir experimentos sistemáticos como o que descrevi traz novas ideias e nos permite *olhar de volta à história cultural de novas maneiras*.

“Make it New”: IA e modernismo

Durante décadas, pensamos que a IA seria incapaz de simular a mais única das habilidades humanas: a criatividade artística. Traduzir entre idiomas, jogar xadrez e vencer no Go, resumir materiais – essas e muitas outras realizações da IA foram notáveis. Mas um computador capaz de produzir fotografias artisticamente refinadas de pessoas e objetos inexistentes, assim como belas imagens que reproduzem exatamente os efeitos de quaisquer mídias e estilos de milhares de artistas conhecidos? Ou criar uma infinidade de variações estéticas combinando estilos e linguagens artísticas existentes? Tudo isso pareceria impensável apenas cinco anos atrás. No entanto, nos últimos anos, os pesquisadores da IA deram às máquinas esses talentos humanos, supostamente únicos, para criar e imaginar. Se você é um entre os milhões de indivíduos que usam aplicativos como Midjourney, MusicML ou GPT, provavelmente já sentiu uma mistura de inspiração e ansiedade, entusiasmo e perplexidade diante dos recém-descobertos poderes criativos da IA.

Hoje, sabemos muito mais sobre a criatividade humana e como ela funciona do que sobre a “criatividade IA”. Desde a década de 1950, muitas teorias alternativas de criatividade foram estabelecidas na filosofia, psicologia, ciência cognitiva e outros domínios, e vários tipos de criatividade foram descritos. É provável que façamos o mesmo com a criatividade IA ao longo do tempo, mas ainda não chegamos lá.

Depois de treinar em trilhões de páginas de texto ou bilhões de imagens fotográficas e de arte tiradas da web, as redes neurais podem gerar novos textos e imagens visuais no mesmo nível de escritores, artistas, fotógrafos ou ilustradores

profissionais altamente competentes. Essas capacidades das redes de sistemas IA são distribuídas por trilhões de conexões entre bilhões de neurônios artificiais, em vez de serem determinadas por algoritmos padrão. Em outras palavras, desenvolvemos uma tecnologia que, em termos de complexidade, é extremamente semelhante ao cérebro humano. Não compreendemos totalmente como nossa tecnologia IA funciona, assim como não compreendemos totalmente o intelecto e a criatividade humanos.

A geração atual de sistemas generativos IA, como GPT e Stable Diffusion, foi treinada em conjuntos de dados muito grandes e diversos, consistindo em bilhões ou mesmo trilhões de textos individuais ou pares de imagem e texto. É, no entanto, igualmente interessante limitar o conjunto de dados de treinamento a uma área específica da história cultural humana, ou a um conjunto específico de artistas de um período histórico específico. [Unsupervised](#), de Refik Anadol Studio (2022), é um projeto de arte IA que exemplifica essas possibilidades. O projeto usa redes neurais treinadas no conjunto de dados de imagens de dezenas de milhares de obras de arte da coleção do MoMA. Esta coleção, a meu ver, é uma das melhores representações do período mais criativo e experimental da história visual humana – cem anos de arte moderna (1870-1970) –, além de incluir exemplos importantes de explorações artísticas das décadas subsequentes. Ela captura os experimentos febris e implacáveis dos artistas modernistas para criar novas linguagens visuais e de comunicação e “torná-las novas”.



Fig. 2 - [Unsupervised](#), Refik Anadol Studio (2022). Frames selecionados da animação.

Superficialmente, a lógica do modernismo parece ser diametralmente oposta ao processo de treinamento de sistemas generativos IA. Os artistas modernos desejavam afastar-se da arte clássica e de suas características definidoras, como a simetria visual, as composições hierárquicas e o conteúdo narrativo. Em outras palavras, sua arte foi fundada em uma rejeição fundamental de tudo o que veio antes dela (pelo menos em teoria, como expresso em seus manifestos). As redes neurais são treinadas de maneira oposta, aprendendo com a cultura histórica e a arte criada até agora. Uma rede neural é análoga a um artista muito conservador estudando no “meta” “museu sem paredes” que abriga a arte histórica.

Mas todos sabemos que a teoria da arte e a prática da arte não são a mesma coisa. Os artistas modernos não rejeitaram completamente o passado e tudo o que veio antes deles. Em vez disso, a arte moderna se desenvolveu reinterpretando e copiando imagens e formas de antigas tradições artísticas, como gravuras japonesas (van Gogh), esculturas africanas (Picasso) e ícones russos (Malevich). Assim, os artistas apenas rejeitaram os paradigmas dominantes da arte erudita da época, a arte realista e de salão, mas não o resto da história da arte humana. Em outras palavras, eram profundamente historicistas; em vez de inventar tudo do zero,

inovaram ao adaptar certas estéticas mais antigas aos contextos da arte contemporânea. (No caso da arte abstrata geométrica criada na década de 1910, os artistas usaram imagens que já eram amplamente utilizadas na psicologia experimental para estudar a sensação e a percepção visuais humanas. Para uma análise detalhada dessas relações entre a arte moderna e a psicologia experimental, ver Paul Vitz e Arnold Glimcher, [Modern art and Modern Science: The Parallel Analysis of Vision](#), 1983.)

Quando se trata de IAs artísticas, não devemos ficar cegos pela forma como esses sistemas são treinados. Sim, redes neurais artificiais são treinadas em artefatos de arte e cultura humanos criados anteriormente. No entanto, seus *outputs* recém-gerados não são réplicas ou simulações mecânicas do que já foi criado. Na minha opinião, frequentemente são artefatos culturais *genuinamente novos*, com *conteúdos, estéticas ou estilos inéditos*.

É claro que simplesmente ser novo não torna automaticamente algo cultural ou socialmente interessante ou significativo. De fato, muitas definições de “criatividade” convergem neste ponto: é a criação de algo que é tanto [original quanto valioso ou útil](#).

No entanto, estimar qual porcentagem de todos os novos artefatos produzidos por IA generativa são também “úteis” (ou “significativos”) para uma cultura mais ampla não é um projeto viável até o momento. Por um lado, não tenho conhecimento de nenhum esforço sistemático para usar tais sistemas para “preencher”, por assim dizer, uma matriz massiva com todo conteúdo e possibilidades estéticas e que fornecesse milhões de sugestões especificamente projetadas. Em vez disso, é provável que, como em todas as outras áreas da cultura popular, apenas um pequeno número de possibilidades seja realizado repetidamente por milhões de usuários, deixando um longo rastro de outras possibilidades não realizadas. Assim, se apenas uma pequena fração do vasto universo de potenciais artefatos IA está sendo realizada na prática, não podemos fazer afirmações amplas sobre a originalidade ou utilidade do resto do universo.

Mídia generativa e arte de banco de dados

Alguns artistas de IA, como [Anna Ridler](#), [Sarah Meyohas](#) e [Refik Anadol](#), utilizaram em seus trabalhos redes treinadas em conjuntos de dados específicos. Muitos outros artistas, designers, arquitetos e tecnólogos usam redes lançadas por outras empresas ou instituições de pesquisa que já foram treinadas em conjuntos de dados muito grandes (por exemplo, Stable Diffusion) e depois os ajustam com seus próprios dados.

Por exemplo, o artista [Lev Pereulkov](#) ajustou o modelo Stable Diffusion 2.1 usando 40 pinturas de conhecidos artistas “não-conformistas” que trabalharam na URSS a partir dos anos 1960 (Erik Bulatov, Ilya Kabakov, etc). A série de imagens de Pereulkov, [Artificial Experiments 1-10](#) (2023), criada com esta rede personalizada, é uma obra de arte original que captura as características artísticas desses artistas, bem como suas semânticas únicas, surreais e absurdas, sem repetir exatamente nenhuma de suas obras existentes. Em vez disso, seus “DNAs” capturados pela rede possibilitam novos conceitos visuais e significados.

A maioria das milhões de pessoas comuns e dos profissionais criativos que empregam ferramentas de mídias generativas as usam como estão, sem ajustá-las. Isso pode mudar no futuro, pois as redes técnicas que usam nossos próprios dados podem se tornar mais fáceis de usar. Mas, independentemente dessas especificidades, todos os artefatos culturais recém-criados produzidos por redes treinadas têm uma lógica comum.

Ao contrário dos desenhos, esculturas e pinturas tradicionais, os artefatos de mídias generativas não são criados do zero. Eles também não são o resultado da captura de algum tipo de fenômeno sensorial, como fotos, vídeos ou gravações de som. Em vez disso, eles são construídos a partir de um grande arquivo de outros artefatos de mídias. Este mecanismo generativo liga as mídias generativas a gêneros e processos artísticos anteriores. Podemos compará-lo com a edição de filmes, que apareceu pela primeira vez por volta de 1898, ou mesmo com a fotografia composta anterior, popular no século XIX. Também podemos considerar obras de arte específicas que são especialmente relevantes, como o filme de colagem experimental [A Movie](#)

(Bruce Conner, 1958) ou muitas das instalações de Nam June Paik que apresentam fragmentos editados de imagens de TV.

Encarar projetos como *Unsupervised* ou *Artificial Experiments 1-10* no contexto deste método de criação de mídias e suas variações históricas nos ajuda a compreender esta e muitas outras obras de arte IA como objetos engajados em diálogos com a arte do passado, em vez de novidades puramente tecnológicas ou obras de entretenimento.

Vejo muitos momentos e períodos relevantes quando vasculho a história da arte, da cultura visual e das mídias atrás de outros usos proeminentes desse procedimento. Estes são relevantes para as mídias generativas atuais não apenas porque os artistas que trabalhavam nessas épocas o utilizavam, mas também porque o motivo deste uso era consistente em todos os casos. *Uma nova acumulação e uma acessibilidade de massas aos artefatos culturais levaram os artistas a criar novas formas de arte a partir dessas acumulações*. Deixe-me descrever alguns desses exemplos.

Artistas digitais e das redes criaram uma série de trabalhos no final dos anos 1990 e início dos anos 2000 em resposta ao novo universo em rápida expansão da rede mundial de computadores. [readme](#), de Health Bunting (1998), por exemplo, é uma página da web que contém o texto de um artigo sobre o artista, com cada palavra vinculada a um domínio da web existente correspondente a essa palavra. [Shredder 1.0](#), de Mark Napier (também de 1998) apresenta uma montagem dinâmica de elementos que compreendem vários websites – imagens, textos, códigos HTML e links.

Voltando ainda mais no tempo, encontramos um amplo paradigma cultural que também foi uma reação ao acúmulo de artefatos históricos de arte e cultura em coleções de mídias de fácil acesso. Esse paradigma é conhecido como “pós-modernismo”. Artistas e designers pós-modernos frequentemente usavam bricolagem e criavam obras que consistiam em citações e referências à arte do passado, rejeitando o foco do modernismo na novidade e na ruptura com o passado.

Embora existam muitas explicações possíveis para o surgimento do paradigma pós-moderno nas décadas de 1960 e 1980, uma delas é relevante para nossa discussão. O acúmulo de artefatos de arte e de mídias em coleções estruturadas e acessíveis, como bibliotecas de slides, arquivos de filmes, livros de história da arte com muitas fotos das obras de arte e outros formatos – nos quais diferentes períodos históricos, movimentos e criadores foram posicionados juntos – inspirou artistas a começar a criar bricolagens a partir dessas referências, bem como a citá-las extensivamente.

E quanto ao “modernismo” das décadas de 1910 e 1920? Embora a ênfase geral fosse na originalidade e na novidade, um dos procedimentos desenvolvidos em busca da novidade foram as citações diretas do vasto universo das mídias visuais contemporâneas que se expandia rapidamente na época. Grandes manchetes, por exemplo, e a inclusão de fotos e mapas tornaram os jornais visualmente mais impactantes; novas revistas visualmente orientadas, como *Vogue* e *Times*, também foram lançadas em 1913 e 1923, respectivamente; e, claro, a nova mídia do cinema continuou a se desenvolver.

Em resposta a essa intensificação visual da cultura de massa, no início da década de 1910, Georges Braque e Pablo Picasso começaram a incorporar fragmentos reais de jornais, pôsteres, papéis de parede e fragmentos de tecidos em suas pinturas. Alguns anos depois, John Heartfield, George Grosz, Hannah Hoch, Aleksandr Rodchenko e um punhado de outros artistas começaram a desenvolver técnicas de colagem de fotos. A colagem de fotos tornou-se outro método de criação de novos artefatos das mídias a partir de imagens de mídias de massa existentes.

Obras de arte contemporâneas que empregam redes neurais treinadas em bancos de dados culturais, como *Unsupervised* ou *Artificial Experiments 1-10*, continuam uma longa tradição de criação de arte nova a partir de *acumulações de imagens e de outras mídias*. Deste modo, essas obras de arte continuam abrindo novas possibilidades para a arte e suas técnicas, particularmente aquelas às quais eu me referi como precursoras da “arte de banco de dados” (ver meu artigo [Database](#)

as a *Symbolic Form*, 1998). A introdução de novos métodos para *ler bases de dados culturais e criar novas narrativas a partir delas* faz parte dessa expansão.

Assim, *Unsupervised* não cria colagens de imagens existentes, como fizeram os artistas modernistas da década de 1920, nem as cita extensivamente, como fizeram os artistas pós-modernos da década de 1980. Em vez disso, o grupo treina uma rede neural para extrair padrões de dezenas de milhares de obras de arte do MoMA. A rede treinada gera, então, novas imagens que compartilham os mesmos padrões, mas não se parecem com nenhuma pintura específica. Ao longo da animação, viajamos pelo espaço desses padrões (por exemplo, “espaço latente”), explorando várias regiões do universo da arte contemporânea. (Para obter mais detalhes sobre os métodos de treinamento da rede GAN usados pelo Refik Anadol Studio, consulte “[Creating Art with Generative Adversarial Network: Refik Anadol’s Walt Disney Concert Hall Dreams](#)”, 2022).

[Artificial Experiments 1-10](#), de Pereulikov, usa uma técnica diferente para gerar novas imagens a partir de um banco de dados de imagens. Ele escolheu apenas quarenta pinturas de artistas que compartilham características-chave. Eles desenvolveram sua arte oposicionista na sociedade comunista tardia (URSS, anos 1960-1980). Eles também viviam na mesma cultura visual. Nas minhas memórias, essa sociedade era dominada por duas cores: o cinza (representando a monotonia da vida urbana) e o vermelho da propaganda.

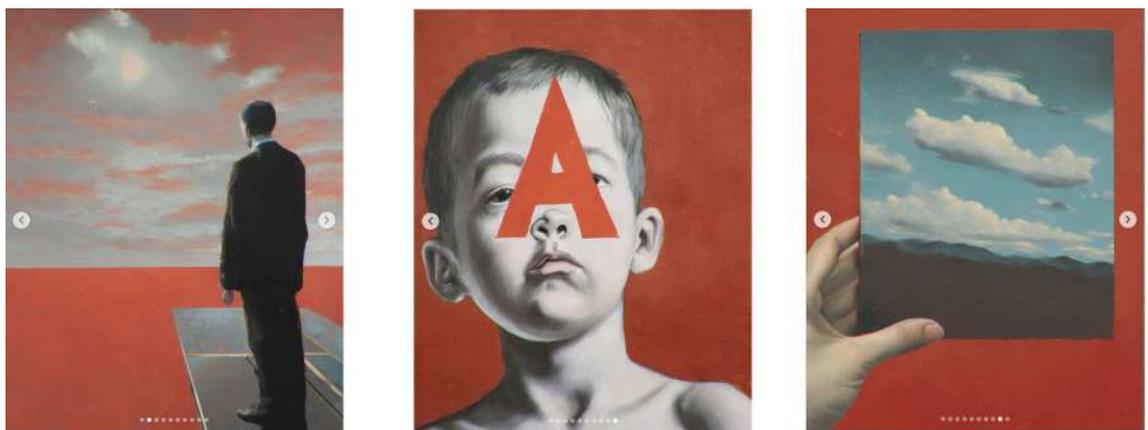


Fig. 3 - Lev Pereulikov, [Artificial Experiments 1-10](#), 2022.
Imagens selecionadas - do Instagram do artista.

Além disso, Pereulkov escolheu pinturas que compartilham algo mais: “Escolhi, via de regra, pinturas que se relacionam conceitualmente de alguma forma com a tela – ou com o espaço nela. Obtive a pintura “Novo Acordeão”, de Kabakov, que apresenta aplicações de papel em cima da tela” (minha comunicação pessoal com Pereulkov, 16/04/2023). Pereulkov também criou descrições de texto personalizadas de cada pintura usada para ajustar o modelo do Stable Diffusion. Para ensinar ao modelo as linguagens visuais específicas dos artistas escolhidos, ele acrescentou termos como “traços grossos”, “iluminação vermelha”, “fundo azul” e “círculos planos” a essas descrições.

Claramente, cada uma dessas etapas representa uma decisão conceitual e estética. Em outras palavras, a chave para o sucesso de *Artificial Experiments 1–10* é a criação de tal banco de dados. Este trabalho demonstra como o ajuste de uma rede neural existente, que foi treinada em bilhões de pares de imagem e texto (como o Stable Diffusion), pode fazer com que essa rede siga as ideias dos artistas; os vieses e ruídos de uma rede tão massiva podem ser superados e minimizados, e não precisam dominar nossa própria imaginação.

Lev Manovich - City University of New York

É um artista, escritor e um dos teóricos mais influentes da cultura digital em todo o mundo. Depois de estudar pintura, arquitetura e cinema, Manovich começou a usar computadores para criar arte digital em 1984. Seus projetos foram exibidos em 12 exposições individuais e 120 exposições coletivas em muitas instituições de prestígio, como o Institute of Contemporary Art (Londres), o Centre Pompidou, a Bienal de Xangai e o ZKM | Center for Art and Media. Atualmente, Manovich é professor de ciência da computação no Centro de Pós-Graduação da City University of New York e diretor do Cultural Analytics Lab.

Emanuele Arielli - Universidade IUAV de Veneza

É Professor Associado de Estética na Universidade IUAV de Veneza. Desde 2011, é pesquisador permanente no Innovationszentrum Wissensforschung da Universidade Técnica de Berlim. É membro da International Association of Empirical Aesthetics - IAEA e do conselho editorial da revista Aisthesis. Realiza pesquisas entre filosofia e ciência cognitiva.

Tradução:**Bernardo Girauta** – Universidade Federal do Rio de Janeiro

É músico, técnico de som, tradutor e doutorando no Programa de Pós-Graduação em Comunicação e Cultura da UFRJ. Pesquisa relações entre música e tecnologia através da comparação entre os conceitos de ruído e harmonia. Possui trabalhos musicais lançados pelos selos QTV e Seminal Records.