

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO

Bruno Sabino

IA Generativa: o quinto paradigma da imagem

MESTRADO EM TECNOLOGIAS DA INTELIGÊNCIA E DESIGN DIGITAL

São Paulo

2025

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO

Bruno Sabino

IA Generativa: o quinto paradigma da imagem

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência para obtenção do título de MESTRE em Tecnologias da Inteligência e Design Digital sob a orientação da Profa. Lucia Santaella.

MESTRADO EM TECNOLOGIAS DA INTELIGÊNCIA E DESIGN DIGITAL

São Paulo

2025

SABINO, Bruno

IA generativa: o quinto paradigma da imagem

Registro: 2025

Orientadora: Profa. Dra. Maria Lucia Santaella Braga

Dissertação de Conclusão de Mestrado – Pontifícia Universidade Católica de São Paulo

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação por processo de fotocopiadoras ou eletrônicos.

Banca Examinadora

Profa. Dra. Maria Lucia Santaella Braga
PUC-SP (Orientador)

Profa. Dra. Roseli Gimenes
UNIP

Prof. Dr. Winfried Nöth
Universidade Metodista

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001

Agradecimentos

À Professora Lucia, pelo exemplo.

Ao Professor Montalvo, pela arte.

Ao Professor Maceira, pela poesia.

Ao Professor Du e ao Professor Paulo, pela amizade.

E à Zélia, por ter sido minha professora.

Resumo

A presente pesquisa investiga a emergência da inteligência artificial generativa texto-imagem na criação visual contemporânea, analisando se a produção imagética com IA pode ser caracterizada como um novo paradigma da imagem, distinto dos quatro paradigmas descritos por Lucia Santaella. O trabalho parte de uma revisão teórica sobre os paradigmas da imagem de Santaella e as tecnologias que os moldaram, incluindo a produção artesanal, fotográfica, pós-fotográfica e híbrida. Posteriormente, aprofunda-se na análise técnica e conceitual das inteligências artificiais generativas, como DALL-E, MidJourney e Stable Diffusion, examinando seus fundamentos tecnológicos, produtivos e criativos. Utilizando uma metodologia comparativa, a pesquisa avalia as similaridades e rupturas entre os modelos de produção imagética anteriores e a IA generativa, considerando aspectos como o papel do agente produtor, os meios de armazenamento, transmissão e recepção. Conclui-se que a IA generativa inaugura um momento paradigmático na história da imagem, introduzindo um modelo de cocriatividade humano-máquina que reconfigura as práticas imagéticas tradicionais e amplia significativamente as possibilidades de criação e acesso à produção visual. O impacto dessa transformação aponta para uma reformulação ontológica das imagens e suas interações culturais, consolidando a IA generativa como um novo estágio evolutivo na produção visual.

Palavras-chave: inteligência artificial, paradigmas da imagem, IA generativa, Santaella, produção visual.

Abstract

This research investigates the emergence of generative text-to-image artificial intelligence in contemporary visual creation, analyzing whether image production with AI can be characterized as a new paradigm distinct from the four paradigms described by Lucia Santaella. The study begins with a theoretical review of Santaella's paradigms of the image and the technologies that shaped them, including artisanal, photographic, post-photographic, and hybrid production. Subsequently, it delves into the technical and conceptual analysis of generative artificial intelligences, such as DALL-E, MidJourney, and Stable Diffusion, examining their technological, productive, and creative foundations. Using a comparative methodology, the research evaluates the similarities and ruptures between previous image production models and generative AI, considering aspects such as the role of the producer, storage, transmission, and reception mechanisms. The study concludes that generative AI inaugurates a paradigmatic moment in the history of images, introducing a model of human-machine co-creativity that reconfigures traditional image-making practices and significantly expands the possibilities of creation and access to visual production. The impact of this transformation suggests an ontological reformulation of images and their cultural interactions, consolidating generative AI as a new evolutionary stage in visual production.

Keywords: artificial intelligence, paradigms of the image, generative AI, Santaella, visual production.

Sumário

Introdução.....	10
Questão de Pesquisa.....	13
Justificativas.....	13
Objetivos.....	13
Hipótese.....	14
Metodologia.....	14
Capítulo 01: Os Paradigmas da Imagem.....	15
1.1 Primeiro Paradigma: Artesanal ou Pré-Fotográfico.....	18
1.2 Segundo Paradigma: Fotográfico.....	22
1.3 Terceiro Paradigma: Pós-Fotográfico.....	29
1.4 Quarto Paradigma: Híbrido.....	33
Capítulo 02: IA Generativa Como Um Novo Modo De Produzir Imagens.....	40
2.1 IA na Produção de Imagens.....	42
2.2 IA Texto-Imagem.....	52
2.3 Os Sistemas Atuais.....	55
2.4 Propriedades Emergentes Do Uso Das IAs Texto-Imagem.....	57
Capítulo 03: Um Novo Paradigma?.....	65
3.1 Modo de Produção.....	65
3.2 Papel do Agente Produtor.....	67
3.3 Meios de Armazenamento.....	69
3.4 Meios de Transmissão.....	70
3.5 Papel do Receptor.....	71
3.6 O Quinto paradigma.....	72
Considerações Finais.....	75
Referências Bibliográficas.....	78

Introdução

As belas-artes e seus diferentes gêneros datam de um tempo bem diferente do nosso, no qual o poder dos homens sobre as coisas e sobre suas relações era insignificante em face do que possuímos hoje. O extraordinário crescimento que os nossos meios experimentaram em suas habilidades de adaptação e precisão impõe significativas mudanças, em futuro próximo, à antiga indústria do belo. Em todas as artes há um aspecto físico que não pode mais ser considerado ou tratado como no passado, pois não pode mais se furtar aos efeitos da ciência e da práxis moderna.

Matéria, espaço e tempo não são mais o que eram há vinte anos. Inovações tão colossais, que alteram o conjunto das técnicas artísticas, acabam por influenciar a própria invenção e talvez terminem por modificar da forma mais extraordinária o próprio conceito de arte.

(Paul Valéry, 1928 p. 103-104)

Escrita há quase um século, a epígrafe de Paul Valéry não poderia ser mais atual. As palavras do autor, que captam a inquietação diante das transformações tecnológicas do início do século XX e seus efeitos na arte, ecoam com vigor no cenário contemporâneo, em que avanços exponenciais no campo das inteligências artificiais reformulam não apenas as ferramentas, mas os próprios alicerces das práticas criativas.

A emergência da inteligência artificial (IA) como protagonista de transformações na cultura não é um fenômeno recente. Desde os primórdios do campo, na década de 1950, com os primeiros experimentos em inteligência simbólica, até a consolidação do aprendizado de máquina e das redes neurais profundas, o percurso da IA tem sido caracterizado por avanços graduais. No entanto, é com a recente incorporação de uma habilidade específica — o processamento de linguagem natural — que a interação humano-máquina adentra um novo patamar. Essa competência semiótica permite que máquinas não apenas processem e interpretem comandos em linguagem natural, mas também sejam capazes de gerar conteúdos originais, como textos, músicas, vídeos e

imagens com alto grau de sofisticação e conveniência (Santaella e Kaufman, 2024). Tais tecnologias, representadas por modelos como ChatGPT, Stable Diffusion e MidJourney, conferem às máquinas a capacidade de operar como agentes co-criativos no âmbito da produção cultural, desafiando as noções tradicionais de autoria e originalidade (Oppenlaender, 2022).

Especificamente no campo da criação visual, a relevância dessa transformação tem sido amplamente debatida por artistas, pesquisadores e críticos culturais, que frequentemente a caracterizam como um momento de grande ruptura histórica. Para Arielli e Manovich (2025), o surgimento das IAs generativas texto-imagem é comparável à invenção da fotografia no século XIX ou à introdução da perspectiva linear no Renascimento, redefinindo as bases ontológicas da imagem. Oppenlaender (2023) descreve o momento atual como a entrada da humanidade em uma nova era criativa, marcada pela acessibilidade e a criatividade das máquinas. Já Poltronieri (2022) propõe a ideia de uma disrupção simbiótica, na qual humanos e IAs atuam de forma concomitante e inédita no processo criativo.

Embora plausíveis, tais afirmativas carecem de um referencial teórico que as sustente, especialmente no que diz respeito à inserção das IAs generativas no arco histórico da produção de imagens. Com frequência, as análises que celebram essas tecnologias como revolucionárias se limitam a destacar sua capacidade técnica ou seu impacto cultural imediato, sem, contudo, aprofundar-se e posicioná-las de maneira clara no contexto das transformações paradigmáticas que moldaram a história da imagem e muito menos estabelecer critérios comparativos entre elas. As IAs generativas texto-imagem representam de fato uma ruptura na história humana da produção de imagens? Se sim, como e a partir de que critérios elas diferem dos períodos predecessores? É, em face dessas lacunas, que esta tese se propõe a se debruçar.

Para isso, será utilizado como arcabouço científico a teoria dos paradigmas da imagem de Lucia Santaella (2013). Publicada pela primeira vez em 1998 e atualizada apenas em 2013, essa teoria categoriza a evolução da produção visual em quatro

momentos a partir de grandes rupturas ocasionadas por inovações tecnológicas, sendo eles: o paradigma pré-fotográfico, o fotográfico, o pós-fotográfico e o híbrido. Ao optar por uma abordagem materialista baseada nos processos produtivos e estabelecer parâmetros bem definidos para a comparação entre as fases, esse modelo se mostra valioso para avaliar a situação contemporânea das IAs generativas texto-imagem, podendo ser utilizado para postulá-las como parte de um paradigma já existente ou configurando-las efetivamente como um paradigma isolado – o quinto paradigma.

Assim, o primeiro capítulo deste trabalho será dedicado a uma revisão bibliográfica detalhada sobre a teoria paradigmática de Lucia Santaella. A definição original do termo "paradigma", introduzida por Thomas Kuhn, será revisitada e também serão exploradas outras divisões possíveis que analisam a evolução da produção imagética humana. Em seguida, será conduzida uma análise aprofundada dos quatro paradigmas propostos por Santaella, avaliando-os conforme os critérios delineados pela autora: modo de produção, papel do agente produtor, meios de armazenamento, meios de transmissão e papel do receptor.

De forma semelhante, o capítulo dois analisará uma vasta e atual bibliografia a respeito das propriedades produtivas e das características específicas das inteligências artificiais generativas texto-imagem, com foco em sistemas como DALL-E, MidJourney e Stable Diffusion. Para isso, será feita uma revisão das bases tecnológicas evolutivas que sustentam essas ferramentas, incluindo os conceitos de redes adversariais generativas (GANs) até os modelos de difusão, explicando seus processos de produção a partir de comandos textuais. Posteriormente, serão discutidos os fundamentos técnicos e criativos dessas IAs, pavimentando o caminho para uma comparação com os paradigmas de Santaella.

Por último, o terceiro capítulo se apoiará numa metodologia de estudo comparativo para investigar como as inteligências artificiais generativas texto-imagem se inserem no arcabouço teórico dos paradigmas de Santaella. A análise considerará os cinco critérios fundamentais delineados pela autora para comparar os paradigmas já

estabelecidos com a produção contemporânea de imagem com IAs e avaliar se essa representa ou não uma ruptura paradigmática.

Questão de Pesquisa

A produção visual com IA generativa texto-imagem pode ser considerada um novo paradigma da imagem, distinto dos quatro paradigmas anteriores descritos por Santaella?

Justificativas

Além disso, ao atualizar e ampliar o entendimento sobre a evolução dos paradigmas da imagem de Santaella frente às tecnologias recentes, este trabalho também contribui para a continuidade do desenvolvimento científico no campo da história da imagem, fornecendo uma base teórica robusta para futuras investigações sobre as transformações tecnológicas e culturais no âmbito da criação visual.

Objetivos

O objetivo geral da pesquisa é investigar se a IA generativa pode ser considerada um quinto paradigma da imagem, adicionando uma nova fase ao desenvolvimento dos paradigmas visuais descritos por Lucia Santaella. A pesquisa busca explorar como a IA generativa se diferencia dos paradigmas anteriores em termos de autonomia criativa e impacto na produção e recepção de imagens.

- **Objetivos Específicos:**
 1. Examinar os quatro paradigmas da imagem estabelecidos por Santaella (pré-fotográfico, fotográfico, pós-fotográfico e híbrido).
 2. Investigar a origem e o desenvolvimento da IA generativa, com foco em suas aplicações na criação de imagens no campo artístico e tecnológico.
 3. Explorar as similaridades e diferenças entre a IA generativa e os paradigmas anteriores, buscando identificar possíveis rupturas ou continuidades.

4. Analisar se as características da IA generativa sugerem a existência de um novo paradigma da imagem.

Hipótese

A hipótese principal deste trabalho é que a produção visual com IA generativa pode ser considerada um novo e quinto paradigma na produção de imagens, distinguindo-se dos demais em todos os cinco critérios estabelecidos por Santaella.

Metodologia

A abordagem será teórico-conceitual e comparativa, com o seguinte desenho metodológico:

1. Revisão Teórica:

- Análise detalhada dos quatro paradigmas descritos por Lucia Santaella, através de uma revisão crítica de suas obras e de outros autores que discutem a evolução das imagens e das tecnologias visuais.
- Revisão de literatura sobre IA generativa, explorando suas bases tecnológicas (redes neurais, GANs) e sua aplicação no campo das artes visuais.

2. Estudo Comparativo:

- Análise comparativa entre as características dos paradigmas pré-fotográfico, fotográfico, pós-fotográfico e híbrido com a produção de imagens pela IA generativa. Esse estudo visa identificar pontos de convergência e ruptura.
- A comparação será feita a partir de critérios como:
 - Papel do Agente Produtor
 - Meios de Armazenamento
 - Meios de Exposição, Disseminação e Transmissão
 - Papel do Receptor

Capítulo 01: Os Paradigmas da Imagem

A palavra "paradigma" tornou-se célebre desde a publicação de *A Estrutura das Revoluções Científicas* de Thomas Kuhn, em 1962. Neste trabalho, e especialmente em textos posteriores, Kuhn define "paradigma" em dois sentidos principais. No sentido mais amplo, um paradigma representa uma matriz disciplinar, um conjunto de compromissos compartilhados que define as práticas e métodos aceitos por uma comunidade científica específica. Já no sentido mais específico, um paradigma se refere a exemplares modelares, soluções concretas que funcionam como referência para a resolução de problemas específicos dentro daquela disciplina. Originalmente, o conceito foi utilizado por Kuhn para explicar mudanças radicais na ciência, quando uma estrutura de conhecimento é suplantada por outra, por meio de um processo de ruptura paradigmática. Nesse sentido, um novo paradigma emerge quando o antigo já não consegue dar conta de resolver as anomalias observadas num determinado campo de estudo (Kuhn, 1962).

A partir dessa ideia, dada a enorme repercussão produzida pelo termo em meios científicos e extra-científicos, seu sentido foi paulatinamente extrapolando à moldura kuhniana, aplicando-se também a outros campos do conhecimento de maneira mais imprecisa e metafórica, para caracterizar "quaisquer realizações científicas ou não científicas reconhecidas que, definindo os problemas e métodos que uma dada comunidade considera legítimos, fornecem subsídios para a prática científica, artística, acadêmica ou institucional dessa comunidade" (Nöth e Santaella, 1998).

No contexto da criação visual, Lucia Santaella (1998) utiliza o conceito expandido de paradigma a fim de examinar a evolução da imagem em relação as suas técnicas e tecnologias de produção, onde cada paradigma abarca rupturas irreversíveis que reestruturaram a forma como as imagens são produzidas, distribuídas e compreendidas ao longo do tempo. Inicialmente, entre os anos 1980 e 1998, a autora organizou a produção de imagens em três paradigmas: o paradigma pré-fotográfico, o paradigma fotográfico e o paradigma pós-fotográfico. Em 2013, com o avanço das tecnologias digitais e a crescente convergência entre mídias, um quarto paradigma – o híbrido – foi adicionado à divisão.

Em vias gerais, o paradigma pré-fotográfico abrange as imagens artesanais, como a pintura e a escultura, onde a habilidade manual do artista é central e a criação visual depende da relação direta entre o criador e o objeto representado. No paradigma fotográfico, inaugurado com a invenção da fotografia, a imagem é capturada automaticamente, introduzindo a máquina como mediadora do processo e estabelecendo uma conexão direta e supostamente objetiva com o real. O pós-fotográfico, por sua vez, marca a ascensão das imagens digitais e sintéticas, liberando-as da referência física e possibilitando a criação de universos visuais independentes da realidade concreta. Finalmente, o paradigma híbrido surge como resultado da convergência entre mídias tradicionais e digitais, criando novas formas de produção visual que combinam elementos físicos e virtuais, inaugurando um espaço multimidiático e interativo.

Entretanto, a busca por classificar a trajetória evolutiva da produção da imagem em grandes momentos ou períodos não é exclusiva à Santaella. Edmond Couchot (apud Santaella, 2013), por exemplo, propõe uma divisão baseada na transição fundamental entre os períodos de representação e simulação. Para Couchot, a arte e a produção visual desde a antiguidade até a era do vídeo, sempre buscaram representar o mundo visível, de forma que as imagens estavam atreladas a um referente no mundo real, seja ele figurativo ou abstrato. Com o advento da computação gráfica e das tecnologias digitais, contudo, as imagens passam a simular uma realidade própria, completamente desvinculada do mundo físico. Neste contexto, a imagem se torna autônoma, não necessitando mais de uma referência concreta para existir. Assim, a simulação substitui a representação na medida em que a imagem digital pode criar realidades novas e completamente independentes, que se autossustentam. Esta teoria sugere que, com a simulação, as imagens alcançam um novo patamar ontológico, onde a realidade virtual se torna não apenas um reflexo do real, mas uma construção visual com existência própria e independente.

Essa abordagem, embora importante, apresenta limitações, especialmente em relação à sua compreensão restrita da representação. Ao vincular a ideia de representação exclusivamente a objetos físicos, Couchot desconsidera a importância das representações abstratas e simbólicas que, mesmo sem uma referência concreta,

desempenham um papel essencial na experiência visual e na construção de significado. Além disso, a divisão rígida entre representação e simulação ignora a complexidade das interações entre o real e o virtual na prática contemporânea. Em muitos contextos, as imagens digitais coexistem e se sobrepõem às imagens representativas, criando um campo híbrido onde as distinções entre o real e o virtual se tornam fluidas. A teoria de Couchot, portanto, carece de uma análise mais nuançada que considere as dinâmicas e as sobreposições entre os diferentes modos de produção de imagens, limitando-se a um dualismo que não abarca a multiplicidade das práticas visuais contemporâneas.

Paul Virilio (1988), por sua vez, propõe uma classificação em três regimes das máquinas de visão: o regime formal, representado pela pintura e gravura; o regime dialético, que surge com a fotografia e o cinema; e o regime paradoxal, que surge com a videografia e infografia. Para Virilio, a imagem evolui da representação estática para uma dinâmica e, finalmente, para uma forma de telepresença, onde a instantaneidade do tempo real transforma a visualidade. Embora inovadora, essa estrutura também mostra-se limitada, dado que, ao misturar diferentes lógicas sem estabelecer critérios sistemáticos para as rupturas que identifica, dificulta sua aplicação prática para entender a totalidade do universo da produção visual .

Em contraste, a abordagem dos paradigmas de Santaella diferencia-se das supracitadas pelo seu caráter materialista e pela forma como considera as bases tecnológicas concretas na produção de imagens. Para a autora, cada paradigma se define não apenas pela mudança de estilo ou técnica, mas, primordialmente, pelas transformações concretas nas condições de produção, afetando as relações entre o criador, a imagem e o espectador. Esse enfoque materialista é particularmente relevante para esta tese, que busca investigar como a IA generativa, como uma tecnologia de criação de imagem, impacta o campo da produção visual e suas práticas estabelecidas. A metodologia de Santaella, ao focar nos meios de produção e estabelecer critérios sólidos para a divisão de cada paradigma – os meios de armazenamento da imagem, o papel do agente produtor, os meios de transmissão e o papel do receptor – oferece um modelo robusto para entender as implicações dessa nova forma de criação e compará-las com as demais. Assim, o trabalho se baseará

nessa estrutura para avaliar se a IA generativa constitui um novo paradigma, analisando suas consequências para as dinâmicas entre produção e recepção de imagens na contemporaneidade.

Para isso, será necessário examinar cuidadosamente cada um dos quatro paradigmas já estabelecidos, a fim de determinar se a IA representa de fato uma ruptura tão significativa a ponto de justificar um novo paradigma, ou se pode ser compreendida como uma extensão dos modelos já existentes.

1.1 Primeiro Paradigma: Artesanal ou Pré-Fotográfico

O primeiro paradigma da imagem, mais frequentemente referido por Santaella (1998) como pré-fotográfico, tem como característica principal a forte imposição da materialidade em sua produção, ou seja, a ênfase que a fisicalidade dos suportes, materiais e instrumentos empregados confere à obra. Essa presença da matéria é constante desde as pinturas rupestres até o desenho, pintura, gravura e escultura, de modo que, ainda que a natureza imagética da escultura e da arquitetura pudesse gerar uma discussão à parte e que haja distinções claras nos processos de produção de pintura, desenho e da gravura, o elemento central é o fato de todas essas imagens dependem do trabalho manual e material. Em função disso, Santaella elege a pintura como o principal exemplar do paradigma artesanal, sendo suas características também aplicáveis, de certa forma, às demais.

A criação artesanal de imagens exige, portanto, um suporte – quase sempre uma superfície – que sirva como receptáculo para as substâncias, normalmente tintas, que o artista, neste caso o agente criador, utiliza para deixar a marca de seus gestos através de ferramentas apropriadas. O corpo humano é o principal instrumento nesse processo, criando extensões à medida que surgem as necessidades. No caso da pintura, o pincel atua como essa extensão dos dedos e do movimento da mão, permitindo ao artista desenvolver habilidade técnica e maestria no uso dessa ferramenta. A pincelada visível é, de fato, a marca do gesto que a originou (Nöth e Santaella, 1998).

O que se obtém desse processo não é apenas uma imagem, mas um objeto singular, autêntico, carregado de uma certa solenidade e sacralidade, fruto da originalidade daquele momento único em que o pintor capturou sua visão do mundo e a transformou em forma com um gesto irrepitível. Por isso, a produção artesanal possui uma característica essencialmente monádica. Como Dubois (1994) sugere, a tela só pode receber a imagem progressivamente, construindo-se lentamente, toque a toque, linha por linha, com pausas, recuos e aproximações, no controle minucioso da superfície, com esboços, correções, retomadas e retoques constantes. Mesmo com interrupções e a lentidão inerente ao processo, isso não elimina a natureza monádica da imagem. Nesse tipo de criação, o sujeito, o objeto criado e a fonte da criação se fundem em um gesto único e inseparável.

Essa unicidade do processo artesanal é também aquilo que confere à obra o que Walter Benjamin (2012) chamou de aura. Em seu influente texto *A Obra de Arte na Era de sua Reprodutibilidade Técnica*, originalmente publicado em 1936, Benjamin ressalta que a obra de arte artesanal possui uma "aura" única, um valor ligado à sua autenticidade e irrepitibilidade. A ligação intrínseca entre o objeto, o tempo e o lugar em que foi criado confere à obra uma singularidade que não pode ser replicada. Assim como Santaella descreve a inseparabilidade entre o gesto do artista e a obra produzida, Benjamin entende que essa relação cria uma experiência irrepitível para o espectador, em que a obra é percebida como uma extensão direta da experiência sensorial e emocional do artista. O conceito de "aura" reforça, portanto, a ideia de que o paradigma pré-fotográfico está profundamente ancorado na autenticidade e unicidade, algo que será perdido nos paradigmas subsequentes com a introdução da reprodutibilidade técnica.

Por tudo isso, ao analisarmos o primeiro paradigma a partir de seus meios de produção, nota-se que a imagem surge diretamente do trabalho manual, onde a fisicalidade dos materiais e o gesto humano são indissociáveis da criação artística. Nesse contexto, o artista ocupa um papel central e insubstituível, pois sua habilidade, imaginação e controle sobre os materiais definem completamente o resultado final da

obra. A imagem é, portanto, uma extensão do corpo e da visão do criador, refletindo sua subjetividade e experiência individual com o mundo.

Definidas as principais características acerca do modo de produção, torna-se mais fácil compreender suas consequências em outros aspectos, como os meios de armazenamento, o papel do agente produtor, os meios de transmissão e, também, o papel do receptor.

Papel do Agente Produtor

O artista, ou agente produtor, desempenha um papel central e insubstituível nesse processo artesanal. No paradigma pré-fotográfico, a criação está completamente subordinada à habilidade manual e à visão criativa do artista, que age como o principal motor da produção. Ao contrário dos paradigmas que surgem com o avanço da tecnologia, em que o processo criativo pode ser mediado por dispositivos mecânicos ou digitais, aqui o artista é o único responsável pela obra. Sua subjetividade, domínio técnico e controle sobre os materiais são essenciais para o resultado final. Essa ligação pessoal entre o criador e a criação confere à obra uma autenticidade única, na qual cada traço, pincelada ou entalhe é uma marca do gesto humano. Assim, o artista não é apenas executor de uma técnica, mas também um intérprete e mediador entre sua visão de mundo e a materialidade que utiliza para expressá-la (Nöth e Santaella, 1998).

Meios de Armazenamento

No que se refere ao armazenamento, no paradigma pré-fotográfico, ele está intimamente ligado ao suporte físico sobre o qual a imagem é criada. Diferente dos paradigmas posteriores, nos quais a imagem pode ser reproduzida e armazenada em diferentes formatos e mídias, aqui a obra é inseparável de seu suporte. Telas, papéis ou superfícies de pedra não apenas preservam a imagem, mas integram sua própria existência. Esse vínculo com a materialidade impõe desafios, pois a durabilidade das obras depende diretamente da qualidade dos materiais utilizados e dos cuidados necessários para sua conservação. O desgaste natural ao longo do tempo reforça a

ideia de singularidade da obra, que não pode ser replicada exatamente como no momento de sua criação (Nöth e Santaella, 1998).

Meios de Transmissão

A transmissão dessas imagens, no entanto, é limitada pela própria materialidade da obra. Ao contrário das imagens que surgirão com a fotografia e a digitalização, que podem ser facilmente disseminadas por diferentes meios, as obras pré-fotográficas só podem ser apreciadas em sua forma original e em locais específicos. Exposições, igrejas e coleções privadas são os espaços onde essas imagens podem ser vistas, o que restringe seu acesso a um público limitado. O deslocamento físico da obra é necessário para que ela seja exibida, e a experiência estética está diretamente vinculada ao espaço em que a obra se encontra. Essa limitação física contribui para a aura de exclusividade da obra de arte pré-fotográfica, tornando o ato de sua contemplação algo raro e, por isso, mais valorizado (Nöth e Santaella, 1998).

Papel do Receptor

O espectador, ao se deparar com uma obra pré-fotográfica, não se limita a observá-la passivamente, mas estabelece uma relação ativa com a imagem. A experiência de contemplação vai além do campo visual, envolvendo uma dimensão sensorial completa. A presença física da obra e seu contexto de exibição influenciam profundamente a percepção do receptor, que vivencia uma experiência única e irrepetível ao estar diante de um objeto singular. Ao contrário da reprodutibilidade moderna, em que a imagem pode ser vista em qualquer lugar e a qualquer momento, a obra pré-fotográfica exige a presença do espectador no mesmo espaço que ela, criando um encontro irrepetível entre a obra e o público. Essa experiência, como aponta Walter Benjamin, é enriquecida pela aura da obra, pela singularidade de sua existência e pela impossibilidade de replicar a vivência estética em outros contextos (Nöth e Santaella, 1998).

Assim, o paradigma pré-fotográfico revela um universo em que todos os aspectos da criação e da experiência artística estão profundamente interligados. A

materialidade dos meios de produção, a singularidade das imagens, a centralidade do artista e a relação íntima com o receptor definem o caráter único desse paradigma, proporcionando uma forma de vivência estética que está essencialmente ancorada na autenticidade e na presença física da obra.

1.2 Segundo Paradigma: Fotográfico

O segundo paradigma da imagem, denominado por Santaella (1998) como fotográfico, surge a partir da invenção da fotografia no século XIX, marcando uma reviravolta profunda no modo como as imagens são produzidas e, conseqüentemente, compreendidas. Mas, apesar de ter causado grande fascínio em seus primeiros observadores, a fotografia não pode ser entendida como fruto de uma invenção repentina ou isolada. Ela é, na verdade, a continuidade de um longo processo evolutivo, diretamente ligado ao desenvolvimento da câmara obscura. Esse dispositivo, já amplamente conhecido e utilizado desde o século XIV, projetava imagens do mundo externo sobre uma superfície interna por meio de uma pequena abertura em um ambiente escuro. A luz que passava por essa abertura reproduzia a cena do outro lado de maneira invertida e em detalhes, de modo que era comum aos artistas renascentistas recorrerem a essa ferramenta para tracejarem estas projeções e melhorarem a precisão de seus desenhos. Dessa maneira, apesar de já antecipar a ideia de captura do real, faltava à câmara obscura, ainda, um suporte sensível à luz para a captura automática da imagem, de um lado, e o negativo para a automatização da reprodução dessa imagem original, de outro (Nöth e Santaella, 1998).

Tais limitações foram superadas por William Henry Fox Talbot, com seu invento descrito pela primeira vez em "Processo Pelo Qual Os Objetos Naturais Podem Ser Delineados Sem A Ajuda Do Lápis Do Artista" (1839), que daria início à fotografia. Tomando como base a câmara obscura, Talbot adicionou ao mecanismo de projeção sais de prata aplicados a um suporte de papel, que, por serem compostos químicos sensíveis à luz, permitiam fixar as imagens de maneira duradoura. Esse método também transformava a imagem projetada em um negativo, do qual não apenas uma, mas múltiplas cópias positivas podiam ser feitas de maneira automática, sem a dependência da habilidade manual de um artista para reproduzi-las.

E, se como mecanismo óptico, a fotografia tem como sua antecessora a câmara obscura, do ponto de vista da capacidade técnica de reprodução de sua imagem, suas precursoras são as diversas técnicas artísticas de replicabilidade. Como descreve Walter Benjamin (2012), a fundição, a cunhagem, a xilogravura e a litografia são todos métodos que renunciavam a capacidade de reprodução técnica do negativo fotográfico. Criadas em períodos diferentes, esses processos possuíam em comum o princípio de gerar múltiplas cópias de uma mesma matriz original, permitindo que esculturas, moedas, ilustrações e formas fossem replicadas e disseminadas para além do alcance de um único exemplar.

Contudo, cada técnica apresentava limites, uma vez que todas dependiam, em maior ou menor grau, do trabalho manual do artista ou artesão, seja para a criação da peça matriz (no caso da xilogravura, do entalhe no pedaço de madeira ou, na litografia, em pedra) ou para seu correto manuseio para replicação, suscetível a variações e perda de precisão. Foi apenas "com a fotografia [que] pela primeira vez a mão é dispensada das tarefas artísticas essenciais nos processos de reprodução de imagem, que agora cabem exclusivamente ao olho que vê por meio da objetiva". (Benjamin, 2012, p. 11).

Além disso, a fotografia introduziu um elemento inédito em comparação aos métodos anteriores: a ancoragem direta no mundo físico. Enquanto gravuras, litografias e outras técnicas reproduziam tanto interpretações dos artistas do mundo ao seu redor quanto imagens imaginárias, a fotografia atua como uma representação diretamente ligada à realidade, necessitando de objetos reais preexistentes e capturando o visível com a precisão de um reflexo. Essa característica conferiu à imagem fotográfica um valor documental, fazendo com que as imagens fossem percebidas como registros visuais automáticos e exatos da realidade externa, marcando o início de um paradigma em que as imagens não apenas se reproduzem tecnicamente, mas, também, estão relacionadas à fidelidade do mundo visível (Nöth e Santaella, 1998).

Essa característica é, inclusive, potencializada com a chegada do cinema e do vídeo, também componentes do paradigma fotográfico. Ao introduzirem o tempo e o movimento na representação visual com uma sequência contínua de imagens, essas mídias fortalecem o vínculo perceptivo com o espectador, oferecendo uma experiência

sensorial expandida que aprofunda a imersão no registro técnico da realidade. No caso do vídeo, essa transformação visual se torna ainda mais poderosa ao incorporar a transmissão em tempo real, permitindo acesso imediato e possibilitando uma documentação contínua e instantânea do mundo ao redor (Nöth e Santaella, 1998).

Por tudo isso, o paradigma fotográfico pode ser definido como uma ruptura em relação ao paradigma artesanal, no qual a criação de imagens era intrinsecamente ligada ao trabalho manual e à fisicalidade dos materiais. Neste novo paradigma, o processo de produção imagética se torna dependente não apenas de um humano, mas, principalmente, de uma máquina e de um processo químico (e, posteriormente, eletrônico), inaugurando uma nova era de captação do mundo visível. A principal característica dessa nova fase é a “gênese automática”, como bem coloca André Bazin (apud Dubois, 1994), ao referir-se ao modo como a fotografia rompe com a intervenção direta do artista sobre a obra.

Walter Benjamin (2012) destaca como traço fundamental deste novo momento na produção de imagens sua capacidade de ser reproduzida numa escala nunca antes vista, cunhando o termo “reprodutibilidade técnica”. Ao permitir que cópias infinitas e uniformes sejam feitas de uma mesma imagem, a fotografia alinha o mundo pictórico com os processos industriais, rompendo com o valor único e aurático que caracterizava as obras no paradigma artesanal. Para Benjamin, essa reprodutibilidade técnica da fotografia não apenas torna a imagem mais acessível, mas também altera sua própria função cultural, promovendo uma nova experiência estética baseada na proximidade e na circulação em massa. O valor de culto da obra cede lugar a um valor de exposição, focado no potencial de disseminação das imagens. Dessa forma, o paradigma fotográfico transforma a relação do espectador com a imagem, que deixa de ser um objeto de contemplação única para se tornar parte de uma cultura visual compartilhada, moldando a percepção coletiva e ampliando o alcance das representações do mundo visível.

Delimitadas as propriedades gerais deste paradigma e as particularidades de seus meios de produção, abre-se agora a possibilidade para uma análise mais profunda de seus desdobramentos. Adotando critérios comparativos como os propostos por Santaella, amplia-se o entendimento sobre como o paradigma fotográfico

se diferencia nos aspectos do papel de seu agente produtor, armazenamento, transmissão e recepção da imagem.

Papel do Agente Produtor

No paradigma fotográfico, o papel do agente produtor sofre uma transformação radical em relação ao paradigma artesanal, uma vez que a produção da imagem passa a ser mediada por uma máquina. Lucia Santaella (1998) aponta que o fotógrafo se distancia da figura do artesão, cujas habilidades manuais e sensibilidade estética imprimiam uma marca única e pessoal na obra, para se tornar um “operador técnico” de uma prótese óptica, que captura o mundo com objetividade e precisão. Essa reconfiguração do papel do agente está associada ao uso da câmera como extensão do olhar, onde o controle manual cede lugar a uma interação técnica, na qual o fotógrafo decide o enquadramento e o momento da captura, mas depende inteiramente da máquina para a criação visual. Em certo sentido, como Walter Benjamin coloca, a mão do artista é substituída pelo olho mecânico da câmera, redefinindo o processo artístico como uma interação entre percepção humana e precisão tecnológica.

A câmera, neste contexto, torna-se uma ferramenta que encapsula o saber técnico da perspectiva e a sensibilidade da luz em seus mecanismos, dando ao fotógrafo um papel de seletor e observador do real, mais do que de criador propriamente dito. Essa seletividade é fundamental: o fotógrafo não “faz” a imagem de forma direta, mas escolhe o ponto de vista e o instante preciso que serão eternizados, enquanto o aparato técnico assegura que a cena captada seja registrada sem a subjetividade explícita do artista. Esse processo de captura de um fragmento da realidade constitui uma espécie de “roubo” do real, em que o agente produtor exerce um papel de mediador e controlador da imagem, mas sem a intervenção direta sobre o conteúdo visual, capturando o visível por meio de um aparato que age quase de forma autônoma.

A função do fotógrafo é marcada, portanto, por uma dualidade entre controle e passividade. Santaella (1998) caracteriza essa interação como uma dinâmica de confronto entre o olhar humano e o dispositivo técnico, onde o fotógrafo precisa, ao

mesmo tempo, dominar a máquina e aceitar os limites de sua própria atuação sobre a imagem. Esse papel dual transforma o fotógrafo em um “caçador de imagens”, um agente que, embora com controle sobre o enquadramento e o momento do clique, não pode intervir diretamente na construção visual que a máquina realiza. O resultado é uma imagem que registra o real como índice, uma testemunha da realidade que depende da precisão mecânica e química da câmera, distanciando o fotógrafo da criação direta da imagem e aproximando-o de um processo de captura e documentação.

Flusser (1985) contribui com esse pensamento ao descrever essa relação entre fotógrafo e câmera como um processo de exploração dos limites e potencialidades do dispositivo, onde o fotógrafo atua como um "jogador" em um “campo programado”. Ele observa que a câmera oferece uma estrutura rígida de possibilidades, mas o fotógrafo tenta explorar cada uma dessas possibilidades ao máximo, “brincando” com a luz, a composição e o momento de captura. Dessa forma, o ato de fotografar é menos uma criação ex nihilo e mais uma descoberta de composições e ângulos já previstos nas potencialidades do aparelho, mas ativados pela intuição e percepção do fotógrafo.

Assim, o “gesto de fotografar” é visto por Flusser (1985) como um gesto interdependente e programado, em que o fotógrafo atua dentro das “regras” impostas pelo aparelho, mas ao mesmo tempo tenta subverter essas regras para capturar algo único. O fotógrafo se torna, portanto, um intérprete das possibilidades mecânicas e ópticas da câmera, utilizando o aparelho para "revelar" uma visão singular do mundo. Esse processo resulta em uma obra que é tanto produto do dispositivo quanto das intenções e decisões do fotógrafo, mostrando como o ato fotográfico é um exercício dialético entre controle técnico e expressão subjetiva.

Meios de Armazenamento

No paradigma fotográfico, os meios de armazenamento da imagem introduzem uma durabilidade que supera os suportes únicos do paradigma artesanal, ampliando a capacidade de preservação das imagens e, conseqüentemente, de sua replicabilidade.

O negativo fotográfico emerge como um arquivo “latente”, um suporte que não apenas conserva a imagem, mas também permite sua reprodução indefinida. Santaella (1998) observa que o armazenamento fotográfico desloca a imagem da exclusividade do objeto material e abre espaço para novos conceitos de conservação e acessibilidade.

Além disso, a estrutura do negativo redefine a ideia de memória visual ao dissociar o momento de captura do momento de exibição, o que significa que a imagem pode ser preservada sem necessariamente estar exposta. Esse processo de armazenamento gera uma relação indireta com o tempo, em que o valor da imagem passa a residir não na sua presença constante, mas na potencialidade de ser acessada e reproduzida em momentos futuros. A conservação da imagem torna-se, assim, não mais uma questão de cuidar do objeto material, mas de manter o negativo intacto, o que altera a dinâmica de temporalidade e preservação em relação ao paradigma artesanal (Santaella, 1998).

Meios de Transmissão

O ato de expor uma imagem fotográfica transcende a ideia de contemplação única e se alinha a uma lógica de acessibilidade e reprodutibilidade. Exposições fotográficas, publicações em jornais, revistas, álbuns de família e, posteriormente, galerias e museus, tornam-se espaços que democratizam o acesso às imagens, permitindo que um mesmo registro visual possa ser visto por inúmeras pessoas em diferentes lugares e momentos. Essa multiplicação da presença da imagem marca uma ruptura com a aura de exclusividade que permeia as obras no paradigma artesanal, e reforça o valor de exposição das imagens fotográficas, promovendo uma experiência visual coletiva e compartilhada (Benjamin, 2012).

Além disso, os meios de exposição no paradigma fotográfico são adaptáveis e mutáveis, permitindo que as imagens possam ser transportadas, ampliadas, reproduzidas em diferentes tamanhos e suportes, e até mesmo integradas em novos contextos culturais e sociais. O mesmo registro fotográfico pode ser encontrado em um jornal, um cartaz de rua, um livro ou uma galeria, adaptando-se ao propósito e ao público-alvo. Essa flexibilidade de exposição amplia o impacto cultural das imagens e

insere a fotografia no cotidiano, conferindo-lhe um papel central na construção de memórias coletivas, identidades culturais e discursos sociais (McLuhan, 1964).

A possibilidade de disseminação em massa também leva à criação de novas formas de circulação da imagem, como as fotomontagens, os cartazes publicitários, os álbuns familiares e, em tempos mais recentes, as plataformas digitais. Cada uma dessas modalidades de exposição carrega suas próprias implicações semióticas e contextuais, moldando a forma como a imagem é percebida e consumida.

Papel do Receptor

O paradigma fotográfico redefine significativamente o papel do espectador em relação ao paradigma pré-fotográfico. Enquanto a experiência estética nas obras artesanais dependia da presença física diante da obra, a fotografia possibilita que as imagens sejam consumidas à distância, em múltiplos formatos e contextos. A circulação massiva de imagens fotográficas torna a arte acessível a um público muito mais amplo, democratizando sua fruição. Como argumenta Walter Benjamin (2012), essa democratização permite que as imagens sejam reproduzidas e disseminadas em escala global, mas não sem custo: a "aura" única da obra de arte, associada à sua presença física e singularidade, se perde nesse processo.

Além disso, o cinema, que emerge como um elemento central no paradigma fotográfico, transforma ainda mais a recepção artística. Diferente da interação isolada e contemplativa que caracteriza o engajamento com obras tradicionais, o cinema apresenta uma forma de recepção coletiva, moldada pela sucessão dinâmica de imagens. Segundo Benjamin (2012), a experiência cinematográfica se destaca por provocar uma "percepção de choque", envolvendo o espectador em uma absorção quase inconsciente. A massa, ao assistir a um filme, interage de maneira ativa, porém dispersa, em contraste com a concentração requerida pela pintura ou escultura.

A ideia de "percepção distraída" introduzida pelo cinema foi alvo de críticas, principalmente por parte dos detratores da cultura de massa, que a consideravam uma ameaça à experiência artística mais profunda. Intelectuais como Duhamel lamentavam

que o cinema fosse um "passatempo para a ralé", promovendo uma forma de diversão que não exigia concentração, representando uma degradação do envolvimento estético. No entanto, Benjamin defende que essa nova forma de recepção não é simplesmente um sinal de decadência, mas uma adaptação às exigências do mundo moderno. A distração, para ele, faz parte do ritmo da vida contemporânea, oferecendo oportunidades de engajamento mais dinâmicas com a arte.

Em conclusão, o segundo paradigma da imagem, conforme definido por Lucia Santaella (1998), transforma radicalmente a relação entre o criador, a imagem e o mundo ao introduzir a máquina como mediadora central no processo de produção imagética. A fotografia, com seu caráter técnico e reproduzível, estabelece uma conexão direta e documental com a realidade visível, prometendo um registro "objetivo" e preciso do mundo. Essa característica legitima a fotografia como um novo modelo de representação, permitindo a circulação massiva de imagens e amplificando seu alcance cultural e comunicacional em uma escala global. A reproduzibilidade técnica desloca o foco de valor das obras de arte, conferindo centralidade à acessibilidade e multiplicação das representações em diferentes contextos sociais. O cinema e o vídeo, que integram esse paradigma, aprofundam a ruptura ao introduzirem o tempo e o movimento às representações visuais, criando experiências sensoriais imersivas e coletivas que ultrapassam os limites das imagens estáticas. Com o avanço das tecnologias digitais, o próximo paradigma pós-fotográfico levará essa evolução ainda mais adiante, possibilitando a criação de realidades totalmente virtuais e desvinculadas do mundo físico—a ser abordado em maior profundidade no capítulo seguinte.

1.3 Terceiro Paradigma: Pós-Fotográfico

De acordo com Santaella (1998), o terceiro paradigma, intitulado como pós-fotográfico, tem seu início a partir do surgimento da computação gráfica e das reviravoltas que esta provocou na natureza do paradigma fotográfico. Segundo a autora, esse novo paradigma introduz uma estrutura de produção triádica, que se diferencia das dinâmicas predominantemente indiciais e diádicas do paradigma anterior. Esse novo processo de produção é inaugurado pela infografia, que traz uma

mutação radical nos modos de criar imagens ao substituir o primado secular da imagem ótica, estabelecido desde o Renascimento com a câmara obscura e a perspectiva monocular, por uma nova economia simbólica baseada na abstração digital. Essa mudança histórica rompeu com os sistemas ópticos e físico-químicos que dominavam a morfogênese imagética, deslocando o eixo de criação para um ambiente computacional, onde as imagens emergem como construções numéricas.

O suporte das imagens sintéticas não é mais matérico, como na produção artesanal, nem maquínico e físico-químico, como nas imagens fotográficas. Em vez disso, ele resulta da interação entre o computador e a tela de vídeo, ambos mediados por operações abstratas, como modelos matemáticos, algoritmos e cálculos simbólicos. Conforme Santaella (1998), o computador, embora seja uma máquina, é de uma categoria especial, pois não opera diretamente sobre a realidade física, mas sobre um substrato simbólico: a informação. Essa distinção é fundamental, pois a imagem infográfica não se forma a partir de um referente externo, mas de modelos numéricos que podem ser manipulados de forma infinita dentro da memória computacional.

Os modelos são o ponto de partida fundamental desse novo modo de produção de imagens, funcionando como estruturas abstratas que descrevem as propriedades de um fenômeno real ou imaginário. Segundo Santaella (1998), o modelo não é apenas uma representação formal, mas uma abstração que pode ser manipulada, transformada e recombina de maneira infinita, formando a base para a criação da imagem digital. Arlindo Machado complementa essa visão ao definir o modelo como "um sistema matemático que procura colocar em operação propriedades de um sistema representado. O modelo é, portanto, uma abstração formal — e, como tal, passível de ser manipulado, transformado e recomposto em combinações infinitas —, que visa funcionar como a réplica computacional da estrutura, do comportamento ou das propriedades de um fenômeno real ou imaginário" (Machado, 1993, p. 117). Essa flexibilidade permite que a imagem digital transcenda a mera representação e se torne um espaço de experimentação visual e simbólica, onde realidades potenciais são modeladas e testadas.

Não só isso, a simulação é outro elemento central desse paradigma. Enquanto o modelo representa uma estrutura formal, a simulação é o ato de experimentar simbolicamente essa estrutura no ambiente digital. Santaella (1998) observa que a simulação possibilita experiências que não ocorrem no espaço e tempo reais, mas em um domínio virtual onde os comportamentos, funções e dinâmicas podem ser testados e ajustados. Essa capacidade de simular e recriar fenômenos complexos sem a necessidade de interagir com objetos físicos redefine o papel das imagens. Elas deixam de ser representações do mundo visível para se tornarem ferramentas de experimentação, capazes de modelar realidades virtuais autônomas.

O papel do computador nesse processo é, portanto, crucial. Ele não apenas gera as imagens, mas também conecta o domínio simbólico ao domínio visual. O pixel — a unidade básica das imagens digitais — é controlado por valores numéricos que podem ser constantemente ajustados, permitindo a reconfiguração da imagem a qualquer momento. Essa interação entre números e pixels une dois mundos distintos: o abstrato, dos algoritmos e programas, e o físico, da tela e da luz, criando uma conexão semiótica entre símbolo, índice e ícone. Como resultado, a imagem digital não é estática, mas um fluxo contínuo de possibilidades, onde o programador atua como mediador entre a abstração matemática e a visualização sensorial (Santaella, 1998).

Em suma, o paradigma pós-fotográfico desloca o foco da aparência física e do rastro dos objetos do mundo para os comportamentos e funções que podem ser modelados e simulados digitalmente. Na visão de Santaella, as duas palavras-chave desse paradigma são "modelo" e "simulação", pois é por meio delas que a imagem digital transcende sua função representacional para se tornar um meio de experimentação e criação de novas realidades. A imagem infográfica, como produto desse sistema, é não apenas visualmente icônica, mas também profundamente abstrata e simbólica, um reflexo da integração entre a inteligência humana e os poderes da inteligência computacional.

Assim compreendidas as premissas fundamentais desse novo regime de produção imagética, torna-se possível então investigar, de maneira aprofundada, os

impactos e as implicações dessa transformação nos elementos estruturantes da dinâmica imagética.

Papel do Agente Produtor

O agente produtor, neste paradigma, assume o papel de um manipulador lógico e experimental, cuja principal habilidade está em modelar dados e operar abstrações simbólicas no ambiente digital. Esse papel contrasta com o do fotógrafo, que capturava fragmentos do real, e com o do pintor, que projetava sua subjetividade sobre um suporte material. Segundo Santaella, o programador infográfico "é um manipulador, sujeito antecipador e ubíquo" que interage com o computador para produzir mudanças constantes nas imagens. Esse novo agente rompe com a centralidade do sujeito criador, característica da pintura, e com a captura performática do real, típica da fotografia, destacando-se como um operador da lógica e da simulação.

Meios de Armazenamento

O paradigma pós-fotográfico substitui os suportes físicos do passado, como negativos e fitas magnéticas, pela memória computacional. Santaella descreve esse armazenamento como uma transição para o universo lógico-matemático, no qual as imagens existem como "entidades abstratas da memória", prontas para serem atualizadas em qualquer momento e lugar. Essa abordagem redefine a relação entre a imagem e o tempo, criando um ambiente no qual as imagens não apenas resistem à deterioração, mas permanecem disponíveis para transformação e reutilização infinita. Diferentemente do paradigma fotográfico, que trouxe a reprodutibilidade para o armazenamento, o paradigma infográfico inaugura o conceito de "disponibilidade", onde o suporte digital oferece uma durabilidade quase eterna, livre das limitações do espaço físico e da degradação material (Santaella, 2013).

Meios Transmissão

A transmissão das imagens no paradigma pós-fotográfico está inserida na era da comunicação planetária e da interatividade. Santaella ressalta que as imagens digitais são "indefinidamente conserváveis" e acessíveis em terminais de computador, o que

permite sua circulação global em tempo real. Essa nova lógica desloca a imagem da esfera da comunicação de massas, típica da fotografia, para a esfera da "comutação", onde o acesso é individualizado e descentralizado. A interatividade emerge como uma característica central, possibilitando que os usuários manipulem e redefinam as imagens, configurando um modelo de transmissão que privilegia o contato e a personalização em detrimento da mera projeção massiva.

Papel do Receptor

No paradigma pós-fotográfico, o receptor estabelece uma relação imersiva e interativa com as imagens digitais, que vão além da simples observação. Segundo Santaella, a necessidade de controlar a imagem à medida que ela é criada levou ao desenvolvimento de modos de programação que priorizam respostas rápidas e dinâmicas às ações do usuário. Esse caráter interativo elimina a distância tradicional entre o espectador e a obra, promovendo uma experiência de mergulho e navegação nas complexidades internas da imagem. A interface entre receptor e imagem torna-se quase orgânica, integrando corpo e mente de forma fluida. Ao manipular as imagens por meio de teclas ou mouses, o receptor experimenta uma relação simbiótica que borra as fronteiras entre o observador e o objeto observado (Santaella, 2013).

1.4 Quarto Paradigma: Híbrido

Publicada em 1998, originalmente, a teoria dos paradigmas da imagem restringia-se a três momentos: o pré-fotográfico, o fotográfico e o pós-fotográfico. No entanto, a complexidade e a proliferação de novos modos de produção, armazenamento, exposição e recepção das imagens, amplamente impulsionados pela convergência digital a partir dos anos 2010 tornaram evidente a necessidade de revisitar essa classificação. Como a própria Santaella (2013) argumenta, o "pós-fotográfico", um termo cunhado em resposta ao impacto da computação gráfica e à ruptura com os sistemas ópticos tradicionais, revelou-se insuficiente para abarcar as intrincadas misturas e interpretações midiáticas que configuram a produção de imagens nos anos que se seguiram.

Assim, em 2013, quinze anos após a formulação inicial de sua teoria, Santaella introduziu o quarto paradigma em sua obra. Como ela explica, a proposta inicial, é motivada, em grande parte, pela hibridização radical das linguagens e dispositivos imagéticos, impulsionados pela convergência digital, pela popularização da internet e pela integração das redes sociais às práticas visuais cotidianas. Essa expansão tecnológica e cultural trouxe à tona dinâmicas inéditas, nas quais a interação entre diferentes dispositivos, linguagens e meios tornou-se central, evidenciando a necessidade de um novo paradigma para abarcar essas hibridizações.

Além disso, as discussões de teóricos como Philippe Dubois e Arlindo Machado influenciaram diretamente a formulação deste novo paradigma. Os debates em torno da dissolução das fronteiras entre fotografia, cinema e vídeo, assim como a proliferação de termos como "cinema expandido" e "fotografia híbrida", evidenciaram que os modos de produção imagética não se limitavam mais a categorias estanques. Para Santaella (2013), tal multiplicidade não era apenas um sintoma de diversidade, mas uma redefinição das bases ontológicas da imagem, preparando o terreno para uma era de hibridizações mais abrangentes e constitutivas do quarto paradigma da imagem, moldado por processos que misturam tecnologias, meios e linguagens de maneira inextricável.

Nesse paradigma emergente, essas misturas encontraram seu auge no ambiente computacional moderno. Se, no pós-fotográfico, os primeiros computadores serviram como uma mídia para a criação de imagens sintéticas e até então desvinculadas das outras formas de produção de imagem dos paradigmas anteriores, agora, as novas tecnologias computacionais passam a abarcar diversos modos de produção visual de maneira integrada, dissolvendo as barreiras entre mídias tradicionais e digitais. Essa integração não se limita à simples convivência entre diferentes linguagens imagéticas, mas promove uma interação constante, na qual propriedades e metodologias são compartilhadas e transformadas. Como afirma Santaella (2013): “Quando o computador deixou de ser uma caixa fechada para produzir imagens, textos e guardar arquivos, mais ainda, quando as interfaces gráficas

abriram as comportas para o envio, troca e compartilhamento de dados multimídia, as misturas entre mídias e linguagens tornaram-se regra”.

O computador, assim, transformou-se em um laboratório experimental no qual diferentes mídias podem se encontrar e suas técnicas e estéticas se combinar na geração de novas espécies sígnicas. “Quando uma mídia é simulada no computador, propriedades e métodos de trabalho lhe são adicionados até o ponto de transformar a identidade dessa mídia. Isso ocorre porque os *softwares*, como as espécies em uma ecologia comum – nesse caso o ambiente computacional compartilhado –, uma vez liberados, começam a interagir, mutar e gerar híbridos” (Santaella, 2007, p.266).

Manovich (2013) corrobora e complementa essa visão ao afirmar que, mais do que o computador como hardware em si, é a “softwarização da cultura” que redefine profundamente os processos criativos ao transformar técnicas, métodos e meios tradicionais em ferramentas virtuais. Para ele, a computação virtualizou praticamente todas as técnicas de criação e modificação de mídia, extraíndo-as de seus meios físicos particulares e transformando-as em algoritmos. Nesse contexto, o software não apenas simula as mídias existentes, mas também possibilita hibridizações entre técnicas antes separadas, o que Manovich denomina como “remixabilidade profunda”.

Essa noção é aparente na evolução da estética visual e na prática cultural contemporânea, onde softwares como After Effects, Illustrator, Photoshop e Canva exemplificam o impacto do ambiente computacional na criação artística. Manovich (2013) observa que essas ferramentas promovem novas formas de expressão ao permitir combinações entre animação, fotografia, gráficos 3D e tipografia, consolidando uma estética híbrida e adaptável. Assim, a softwarização implica um deslocamento do foco criativo nas limitações impostas pelos meios físicos para um ambiente virtual onde os elementos culturais podem ser continuamente transformados, remixados e recriados.

Por tudo isso, nota-se uma forma radicalmente nova de produção de imagens, consolidando o quarto paradigma de Santaella (2013). Nele, a criação imagética é definida por um modelo híbrido, onde dispositivos analógicos e digitais não apenas

coexistem, mas interagem de maneira profunda em um ambiente virtual e contribuem para a criação de imagens que transcendem as propriedades de uma única mídia. A imagem não é apenas um produto final estático, mas torna-se um processo ativo e dinâmico, gerado em um espaço compartilhado entre algoritmos, ferramentas digitais e intervenções humanas.

Além disso, os meios de produção no paradigma híbrido são marcados por uma forte interatividade. Ferramentas como softwares de edição em tempo real, dispositivos de captura volumétrica e tecnologias de realidade aumentada e virtual possibilitam que o processo criativo seja ajustado e modificado continuamente, muitas vezes em resposta a estímulos externos ou à intervenção do próprio espectador. Essa dinâmica transforma o ato de produzir imagens em uma experiência colaborativa e fluida, na qual o criador interage constantemente com as possibilidades oferecidas pelos dispositivos e pelos ambientes digitais (Manovich, 2013).

Entendidas as propriedades produtivas deste quarto paradigma, torna-se agora pertinente investigar em mais detalhes suas consequências em outros espectros do universo imagético, como o papel do agente produtor, o modo de armazenamento, distribuição e recepção dessas imagens. No texto original sobre os três paradigmas da imagem, Santaella (1998) reserva um espaço específico para elaborar tal discussão. Entretanto, para a postulação do paradigma híbrido (Santaella, 2013), a autora opta por outra estratégia de escrita, de modo que, apesar de presentes, algumas dessas consequências específicas não são diretamente abordadas. Por essa razão, as reflexões de Lev Manovich em “Software Takes Command” (2013) serão aqui utilizadas de forma a complementar os dizeres de Santaella a esse respeito, entendendo que, apesar de não utilizar a nomenclatura de paradigma, o autor se refere ao mesmo momento da produção de imagens.

Papel do Agente Produtor

No quarto paradigma da imagem, o papel do agente produtor apresenta uma mudança fundamental em relação ao paradigma pós-fotográfico, mantendo a ideia de coautoria entre humano e computador, mas deslocando o tipo de conhecimento técnico

necessário para a criação. Enquanto no paradigma pós-fotográfico o criador de imagens era, necessariamente, um artista-programador, amplamente conhecedor dos modelos matemáticos e dos códigos que sustentavam o funcionamento das máquinas, no paradigma híbrido essa barreira técnica é reconfigurada. A evolução do computador pessoal e o desenvolvimento de softwares com interfaces mais acessíveis não eliminam a necessidade de habilidades especializadas, mas transferem o foco do domínio da programação para o domínio das ferramentas e linguagens específicas desses softwares.

Como observa Manovich, “o papel da vanguarda midiática não é mais desempenhado por artistas individuais em seus estúdios, mas por uma variedade de agentes, de empresas gigantes como Microsoft, Adobe e Apple, até programadores independentes, hackers e designers” (Manovich, 2013). Essa reconfiguração não elimina a necessidade de conhecimento técnico, mas redefine sua natureza: o domínio exigido desloca-se da programação para a compreensão das funcionalidades e potencialidades dos softwares. A sofisticação das ferramentas digitais, com interfaces gráficas intuitivas, permite que operações complexas sejam realizadas sem que o usuário precise lidar diretamente com os algoritmos subjacentes. No entanto, o agente produtor deve ser proficiente no uso dessas interfaces e na combinação de elementos como animação, fotografia e tipografia para criar composições híbridas. Assim, o papel do criador no paradigma híbrido continua exigindo expertise, agora centrada na exploração prática dos softwares e na mediação entre as possibilidades tecnológicas e as demandas criativas da cultura contemporânea.

Meios de Armazenamento

No quarto paradigma da imagem, as dinâmicas de armazenamento observadas no terceiro paradigma são amplificadas com o advento do armazenamento na web, que transforma de maneira definitiva as relações entre as mídias e seus suportes. Se no paradigma pós-fotográfico já se observava o deslocamento dos meios físicos para o espaço digital, agora, esse processo atinge um novo patamar, com o armazenamento

digital não apenas centralizando os conteúdos, mas conectando diferentes formas de mídia em um ecossistema global e acessível.

Como sugere o nome do paradigma, o hibridismo emerge como elemento central, uma vez que todas as mídias — textuais, visuais, sonoras ou cinematográficas — passam a ser softwarizadas e integradas em formatos digitais, dissolvendo as fronteiras que antes delimitavam suas especificidades. Esse fenômeno é exemplificado por Lev Manovich: “em 2012, a Europeana fornecia informações e links para 20 milhões de objetos culturais digitalizados, incluindo pinturas, desenhos, mapas, livros, jornais, diários, música e gravações de voz, filmes, cinejornais e transmissões de rádio e TV, contribuídos por 1500 instituições europeias” (Manovich, 2013; p.227). Tal fato ilustra como materiais antes dispersos em formatos analógicos são reunidos, acessados e compartilhados em um único ambiente digital.

Assim, no quarto paradigma, o armazenamento digital não apenas dá continuidade às inovações do paradigma anterior, mas expande suas possibilidades, potencializando o hibridismo e criando um espaço onde todas as formas de mídia coexistem e se reconfiguram em novos modos de uso e interação.

Meios de Transmissão

As redes sociais configuram-se como o principal meio de transmissão e exposição no paradigma híbrido, desempenhando um papel determinante na circulação de conteúdos visuais e na mediação das interações culturais. Lev Manovich (2013) observa que “os serviços de redes sociais eliminaram, em muitos casos, as fronteiras entre produção, distribuição e consumo de mídia”, destacando como essas plataformas transformaram as dinâmicas tradicionais de criação e disseminação de conteúdos. Neste contexto, a criação imagética deixa de estar restrita a estúdios especializados, sendo integrada diretamente ao espaço digital, onde produção, transmissão e interação se fundem em um único ecossistema.

Além disso, redes sociais como YouTube, Instagram e TikTok exemplificam o hibridismo central ao quarto paradigma, pois integram, de forma nativa, ferramentas de

edição, remixagem e compartilhamento. Essas plataformas sintetizam funções antes separadas, permitindo que imagens, vídeos e outros conteúdos sejam simultaneamente criados, consumidos e redistribuídos em ciclos contínuos. Tal integração não apenas viabiliza a circulação massiva e adaptável de conteúdos, mas também os reconfigura constantemente, moldando-os às dinâmicas sociais, estéticas e técnicas de cada ambiente digital. Assim, o ato de postar em redes sociais transcende a mera apresentação, transformando-se em um processo de ressignificação que acompanha e potencializa as possibilidades tecnológicas do paradigma híbrido (Manovich, 2013).

Papel do Receptor

O papel do receptor é transformado pelas interfaces digitais e pelo ambiente ainda mais interativo das redes sociais. Lev Manovich (2013) argumenta que “as interfaces de aplicativos de acesso à mídia, como navegadores da web e motores de busca [...] encorajam as pessoas a 'navegar', movendo-se rapidamente tanto horizontalmente entre mídias [...] quanto verticalmente, através dos artefatos de mídia”. Essa observação revela que o receptor não é mais um espectador passivo, mas um agente ativo, que navega, interage e influencia o fluxo de conteúdos visuais.

Essa dinâmica de navegação contínua entre plataformas e formatos é um dos traços mais marcantes do paradigma híbrido, pois reflete o hibridismo presente não apenas na criação, mas também na experiência de consumo das imagens. O ato de “navegar” implica em escolhas ativas e personalizadas, permitindo que os usuários estabeleçam suas próprias trajetórias de exploração imagética, remixando e reorganizando conteúdos de acordo com seus interesses e contextos (Manovich, 2013).

Capítulo 02: IA Generativa Como Um Novo Modo De Produzir Imagens

O ano de 2021 testemunhou um marco significativo na história da criação visual assistida por inteligências artificiais com o lançamento do DALL-E, da empresa OpenAI. Nomeado a partir da junção dos nomes do pintor surrealista Salvador Dalí e do icônico personagem robótico de animação da Pixar “*Wall-E*”, o sistema inaugurou uma nova era na síntese de imagens digitais, utilizando modelos avançados de aprendizado profundo (*deep learning*) para gerar composições visuais complexas e diversas a partir de comandos textuais em linguagem natural. Dessa forma, através de uma breve descrição em texto (como "um estudante de mestrado digitando sua tese em um computador em cima de uma moto pegando fogo", conforme Figura 01), um usuário qualquer, mesmo sem formação tecnológica ou artística, tornou-se capaz de, em poucos segundos, ter suas ideias materializadas em imagens digitais de alta qualidade e estilos diversos (Manovich, 2023).

Figura 01 - Imagem gerada com DALL-E-3 a partir da descrição "um estudante de mestrado digitando sua tese em um computador em cima de uma moto pegando fogo"



Fonte: autoria própria (2025)

A partir desse marco, o que se sucedeu já no ano seguinte foi uma rápida expansão do campo intitulado “arte generativa”, como o surgimento de diversas outras aplicações de I.As generativas (IAGs) texto-para-imagem (*text-to-image*). Entre elas, sistemas como MidJourney, Stable Diffusion e DALL-E 2 se consolidaram como as mais utilizadas, cada uma com suas características únicas que ampliaram as possibilidades criativas e tecnológicas do cenário: MidJourney destacou-se por suas variações estéticas e precisão, capaz de gerar imagens que simulam desde o fotorrealismo a qualquer tipo de mídia física ou modelos 3Ds; Stable Diffusion, por sua abordagem de código aberto que fomentou personalizações e inovações em diferentes campos; e DALL-E 2, a segunda versão do sistema pioneiro, por suas melhorias na precisão e interatividade, permitindo interpretações mais complexas e ajustes detalhados nas imagens geradas (Manovich, 2023).

Com igual celeridade, softwares já amplamente difundidos do mercado de edição de imagens, como Canva e Photoshop, promoveram adaptações estratégicas para incorporar as inovações das inteligências artificiais generativas em seus ecossistemas tecnológicos, o que popularizou ainda mais o seu uso. Em Março de 2023, A Adobe, empresa criadora do Photoshop, anunciou o Firefly, seu modelo de I.A generativa que foi incorporado a todos aos seus softwares de edição de imagem, permitindo não apenas gerar imagens do zero mas também adicionar, modificar ou remover partes específicas de composições existentes a partir de simples comandos textuais (Lorenzo, 2023). Em Outubro do mesmo ano, o Canva – plataforma de design amplamente reconhecida por sua acessibilidade e uso tanto por designers profissionais quanto por educadores, empreendedores e criadores de conteúdo em geral – apresentou o Magic Studio, incorporando a IA generativa para realizar as mesmas funcionalidades supracitadas (Pereira, 2023).

Mais recentemente, a produção de imagens por meio de comandos de texto com IAs generativas atingiu um novo patamar de disseminação: sua integração ao aplicativo de mensagens instantâneas mais utilizado no mundo, o WhatsApp. Anunciada em 2024, tal inovação possibilitou que os usuários incorporassem a criação de imagens diretamente a suas trocas de mensagens cotidianas, apenas adicionando “@metaAI”

ao início de uma frase e escrevendo a descrição desejada. Além disso, a nova funcionalidade também promoveu um aspecto colaborativo à essa tecnologia, de modo que as pessoas num mesmo *chat* podem responder a uma imagem gerada anteriormente por outro usuário com novos comandos, permitindo a criação sequencial de imagens que se complementam ou reinterpretam o conteúdo inicial (Mozelli, 2024).

Essa trajetória de expansão vertiginosa — ainda em plena ascensão durante a elaboração desta tese — indica que a inteligência artificial generativa não constitui uma tecnologia efêmera ou nichada, mas sim o início de um momento inédito na produção imagética contemporânea, marcada por uma nova dinâmica criativa entre humanos e máquinas e sua ampla adoção não apenas por designers e artistas, mas também por um público amplo que abrange até mesmo usuários comuns em interações cotidianas.

Por tudo isso, o objetivo deste capítulo consiste em aprofundar a compreensão no modo como as imagens das inteligências artificiais generativas texto-imagem são produzidas, investigando de maneira sistemática os princípios operacionais que regem essas tecnologias, os processos técnicos que culminaram em sua invenção e as propriedades criativas que as distinguem, a fim de criar uma base comparativa para inseri-las na lógica paradigmática de Santaella (Nöth e Santaella, 1998).

Para tal, antes, será indispensável empreender um olhar retrospectivo e analítico sobre os processos que levaram ao surgimento desse tipo de IA, elucidando como práticas computacionais e estéticas desenvolvidas ao longo de décadas estabeleceram os fundamentos necessários para sua emergência enquanto fenômeno contemporâneo e de que forma elas se diferenciam de seus antecedentes.

2.1 IA na Produção de Imagens

Contrariando a impressão de novidade disruptiva, o surgimento da IA para produção de imagens deve ser compreendido como uma consequência de processos históricos e tecnológicos gradativos. Suas raízes encontram-se em décadas de pesquisa dedicada à inteligência artificial e ao aprendizado computacional. Arthur I. Miller, em seu livro *The Artist in the Machine* (2019), oferece uma narrativa meticulosa

sobre a evolução histórica dos diversos experimentos que culminaram na ascensão da inteligência artificial como ferramenta criativa.

Entre os marcos delineados por Miller (2009), o Deep Dream, desenvolvido por Alexander Mordvintsev no Google em 2015, destaca-se como uma das inovações primordiais. Este sistema, projetado inicialmente para visualizar os processos internos das redes neurais convolutivas (CNNs), revolucionou a compreensão de como essas arquiteturas processam e reinterpretem dados visuais. Para compreender sua relevância, é essencial situar as CNNs em seu contexto original e descrever como suas funções tradicionais foram ampliadas e ressignificadas por essa Mordvintsev.

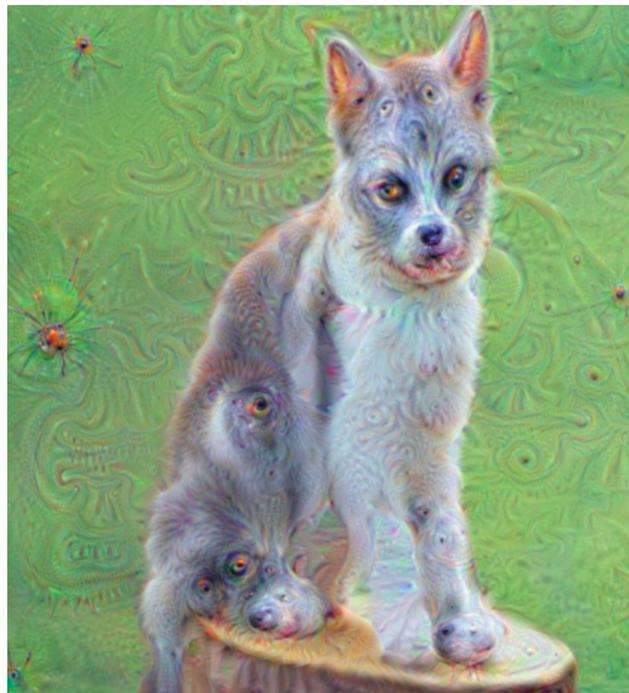
As CNNs foram inicialmente projetadas para processar imagens e identificar padrões visuais a partir de grandes conjuntos de dados. Seu funcionamento é baseado em um processo de treinamento supervisionado, no qual a rede é exposta a milhares ou milhões de imagens rotuladas pertencentes a uma mesma categoria, como "gatos" ou "cachorros". Durante o treinamento, a rede utiliza camadas convolutivas para decompor as imagens, identificando características elementares, como bordas e texturas, nas camadas iniciais, e combinando essas informações em representações mais complexas, como formas ou objetos inteiros, nas camadas mais profundas. Esse aprendizado iterativo ajusta os pesos das conexões neurais a cada tentativa, refinando a capacidade da rede de reconhecer com precisão os padrões associados a um determinado tema. É dessa maneira que, ao receber uma imagem nova, um computador é capaz de identificar com alta confiabilidade se o objeto representado é um gato, um cachorro ou outro elemento previamente aprendido (Miller, 2019).

O Deep Dream, no entanto, subverteu essa lógica analítica ao explorar os padrões aprendidos pelas CNNs de maneira inovadora. Mordvintsev estava interessado em compreender o que as redes "viam" ao processar as imagens, uma questão que se tornou o ponto de partida de sua experimentação. Como relatado em entrevista a Miller (2019), após alimentar a rede com uma imagem específica, o engenheiro programou o sistema para interromper o processamento em uma de suas etapas intermediárias de identificação. Nesse estágio, a rede já havia detectado

características específicas, mas ainda não as havia integrado em representações completas. A partir daí, o sistema foi configurado para amplificar automaticamente as características detectadas até aquele ponto, realçando elementos supostamente reconhecidos, como olhos, focinhos ou texturas específicas.

Em seu primeiro experimento, Mordvintsev forneceu a imagem de um gato a uma CNN treinada para distinguir gatos de cachorros e, ao seguir os passos citados anteriormente, obteve como resultado uma composição que, apesar de partir de características visuais reconhecíveis, logo se transformava em algo peculiar e inesperado: olhos adicionais, focinhos duplicados e texturas exageradas emergindo em uma paisagem surrealista (Figura 02). Embora simples aos olhos das tecnologias contemporâneas, esse feito embrionário revelou o potencial inexplorado das IAs para a geração de imagens que transcendiam os objetivos funcionais para os quais haviam sido projetadas (Miller, 2019).

Figura 02 - Imagem criada por Alexander Mordvintsev usando Deep Dream



Fonte: Miller (2019)

Mike Tyka, cientista também vinculado ao Google, deu o passo subsequente na exploração dessa tecnologia, empregando o Deep Dream para transformar padrões abstratos em criações visualmente complexas. Tyka alimentou a rede neural com imagens compostas por "ruído" – um arranjo aleatório de pontos sem significado aparente –, o que fez com que o sistema, acessando seu vasto repertório de padrões armazenados no banco de dados ImageNet — como formas de cães, gatos, carros ou passaros —, tentasse identificar semelhanças latentes, mesmo onde elas não existiam objetivamente. As imagens resultantes, como exemplificada na Figura 03, se assemelham a obras surrealistas, nas quais observa-se uma proliferação de elementos que evocam criaturas familiares, como répteis, peixes e insetos, imersos em texturas caleidoscópicas e multicoloridas (Miller, 2019).

Figura 03 - Imagem gerada por Mike Tyka utilizando o Deep Dream



Fonte: miketyka.com (2016)

Outro dos vários experimentos descritos por Miller (2019) que antecede a criação das IAGs atuais é o Style Transfer, uma técnica desenvolvida por Leon Gatys, Alexander Ecker e Matthias Bethge em 2015. Diferentemente do Deep Dream, que amplifica padrões internos, o Style Transfer utiliza redes neurais convolutivas (CNNs)

para sintetizar elementos estéticos de uma obra de arte ou imagem e aplicá-los sobre outra, gerando uma fusão visual que combina estrutura e estilo de maneira única.

Em termos técnicos, essa abordagem baseia-se na análise simultânea de duas imagens pela CNN: uma imagem base, que define a composição estrutural e o conteúdo, e uma imagem de referência, da qual são extraídos os traços estilísticos. Por meio de suas camadas convolutivas, a rede decompõe a imagem base em características essenciais, como bordas, formas e texturas, enquanto, da imagem de referência, extrai informações associadas ao estilo, como paletas de cores, padrões de pinceladas e texturas específicas. Essas informações são, então, integradas pela rede em um processo de otimização que gera uma nova imagem, preservando a composição da imagem base, mas revestindo-a com os atributos estilísticos da referência (Miller, 2019). Assim, o Style Transfer possibilita a reinterpretação de uma fotografia ordinária em uma composição que evoca as características estéticas de artistas como Van Gogh, Picasso ou Monet. Enquanto a estrutura original da imagem permanece intacta, ela é transformada à luz das paletas cromáticas, texturas distintivas e pinceladas que definem o estilo dos pintores de referência, como exemplificado na Figura 04.

Figura 04 - Fotografia da Neckarfront em Tübingen (canto superior esquerdo). Pintura “A Noite Estrelada”, de Van Gogh (canto inferior esquerdo). Combinação das imagens (direita)



Fonte: Miller (2019)

Dando seguimento ao panorama de inovações que precederam as inteligências artificiais generativas atuais, as GANs (Redes Adversárias Generativas) ocupam um papel central como um dos avanços mais significativos no campo da produção autônoma de imagens. Contemporânea às tecnologias anteriores, a primeira versão desse sistema, introduzida por Ian Goodfellow e seus colaboradores, não apenas ampliou as possibilidades técnicas da IA, mas também redefiniu as bases conceituais da aprendizagem de máquina (Miller, 2019).

As GANs operam a partir de uma arquitetura dual composta por duas redes neurais que competem entre si: o gerador e o discriminador. Enquanto o gerador tem como objetivo criar imagens que sejam indistinguíveis de exemplos reais, o discriminador é treinado para avaliar se as imagens geradas são autênticas ou sintéticas. Esse processo adversarial, baseado em aprendizado profundo, resulta em um ciclo de refinamento mútuo, onde o gerador se torna progressivamente mais habilidoso em produzir imagens convincentes, enquanto o discriminador aprimora sua capacidade de identificação (Miller, 2019).

Como explica Miller (2019), inicialmente, as imagens geradas pelo gerador são abstratas e confusas, sendo prontamente rejeitadas pelo discriminador, que funciona como um "detetive" avaliando a autenticidade das imagens com base em seu treinamento anterior. O gerador, então, retorna às camadas intermediárias do modelo, ajustando os pesos e parâmetros de acordo com o feedback do discriminador. Esse processo iterativo permite que o gerador aprenda a criar imagens cada vez mais realistas, abandonando gradativamente sua dependência do espaço latente inicial. Com o tempo, o gerador desenvolve a capacidade de produzir imagens que enganam o discriminador, que por sua vez aprimora continuamente suas habilidades de detecção. Eventualmente, o sistema atinge um equilíbrio estável, no qual o gerador consegue produzir imagens tão convincentes que se tornam indistinguíveis dos dados reais utilizados no treinamento do discriminador.

Antes do surgimento das GANs, o ajuste de modelos geradores era realizado de forma manual, exigindo que os pesquisadores definissem explicitamente os parâmetros que determinavam como os dados deveriam ser criados. Essa abordagem não apenas limitava a flexibilidade dos modelos, mas também tornava o processo trabalhoso e, em muitos casos, impossível, dada a quantidade de variáveis e a dificuldade de otimizar modelos em espaços de alta dimensionalidade. Assim, as GANs revolucionaram esse cenário ao automatizar e dinamizar o processo de ajuste por meio do aprendizado adversarial (Miller, 2019).

Figura 05 - Um dos rostos gerados por Myke Taka, em 2017

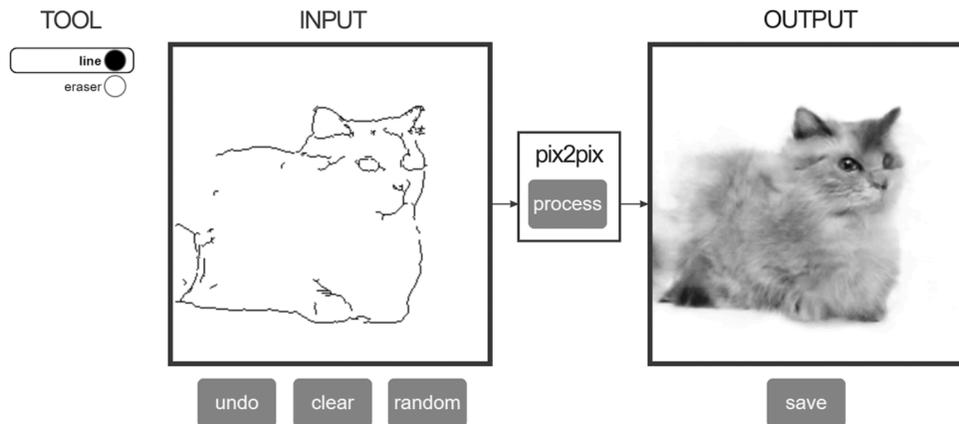


Fonte: Miller (2019)

Além disso, uma das contribuições mais notáveis das GANs para as IAs atuais é sua capacidade de criar imagens que não são meramente reproduções de dados existentes, mas sim composições originais que combinam características aprendidas a partir do conjunto de treinamento. Um exemplo seminal dessa potencialidade é o experimento *Portraits of Imaginary People* (2017), no qual o artista Mike Tyka (o mesmo do Deep Dream), explorou as possibilidades criativas das GANs ao alimentar o discriminador com milhares de retratos de pessoas. Os resultados, conforme a Figura 05 apresenta, foram imagens de rostos completamente inéditos, que não reproduziam diretamente nenhuma das imagens do banco de dados, mas sintetizavam seus padrões para gerar "fotografias" de pessoas que jamais existiram, com composições visuais estilizadas e próximas da realidade. Embora rudimentares em comparação com as tecnologias de texto-imagem contemporâneas, experimentos como esse desempenharam um papel crucial ao demonstrar a capacidade das IAs de gerar imagens originais, pavimentando as bases para os avanços posteriores no campo da criação visual automatizada (Miller, 2019).

Com o avanço do desenvolvimento da técnica de aprendizado profundo (*deep learning*) e o aumento das bases de dados disponibilizadas online, o impacto das GANs foi amplificado por inovações subsequentes. As CGANs (Redes Adversariais Generativas Condicionais), aprimoraram a técnica adversarial ao condicionar o gerador por meio de uma imagem de entrada, possibilitando experimentos em que, receber desenhos tracejados ou inacabados dos usuários, esses "se transformavam" em imagens fotorrealistas. Exemplos célebres, como os criados por Chris Hesse ainda em 2017, incluem um modelo treinado para converter esboços de gatos em retratos realistas de felinos (Figura 06), e outro que utilizava simples desenhos de bolsas para gerar representações fotorrealistas, com texturas e detalhes precisos (Miller, 2019).

Figura 06 - Chris Hesse, edges2cats, 2017.



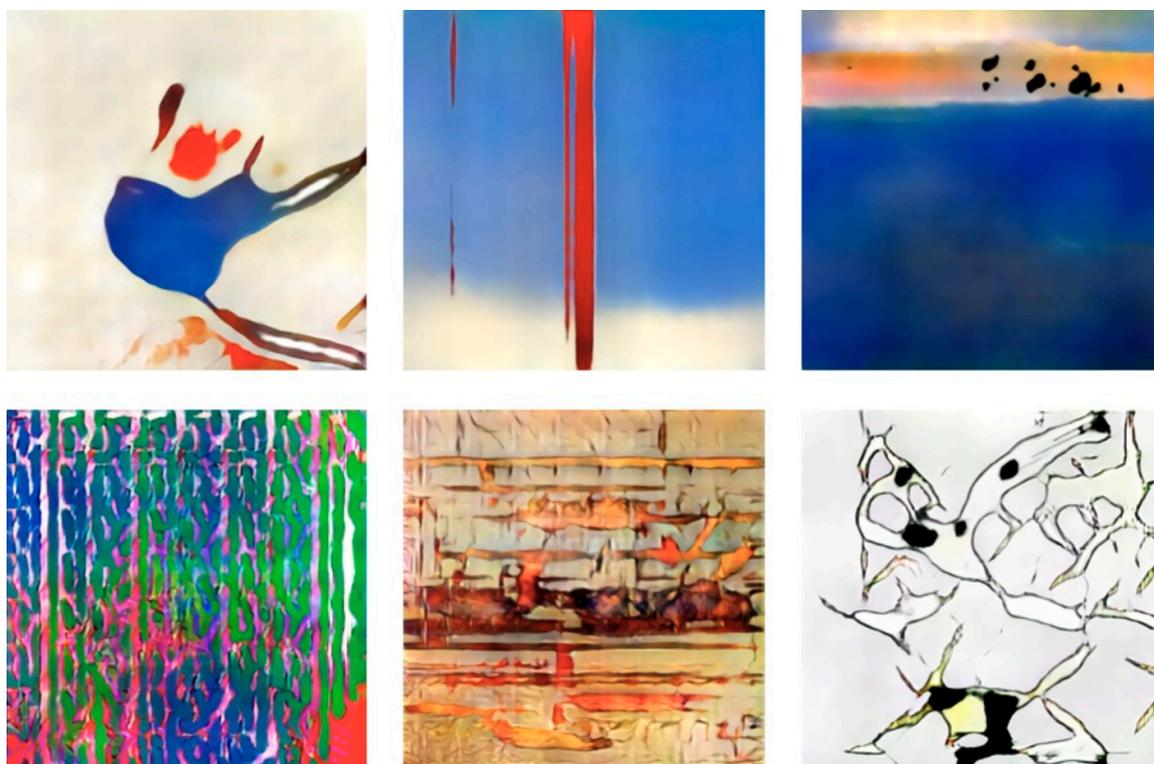
Fonte: Miller (2019)

Ainda nessa linha evolutiva, as CANs (*Creative Adversarial Networks*, ou Redes Adversárias Criativas), foram introduzidas como mais uma variação inovadora das GANs, projetadas especificamente para impulsionar a chamada “criatividade computacional”. Diferentemente das GANs tradicionais, que buscam replicar padrões dos dados de treinamento com alta fidelidade, as CANs exploram uma abordagem mais ousada: incentivar o gerador a sair das convenções estilísticas dos dados de treinamento, criando imagens que desafiam as categorizações estabelecidas (Miller, 2019).

Conforme explica Elgammal, desenvolvedor da tecnologia: “Nosso objetivo era estudar o processo criativo artístico e como a arte evolui de um ponto de vista perceptivo e cognitivo. [...] O processo simula como os artistas assimilam obras de arte anteriores até que, em determinado momento, rompem com estilos estabelecidos e criam novos estilos” (Mazzone, Elgammal, 2019; p.02). Assim, o desenvolvimento das CANs envolveu o treinamento de redes neurais em uma base de dados com mais de 80 mil imagens representando cinco séculos da história da arte ocidental, com o qual o sistema aprendeu não apenas as características definidoras de cada período, mas também identificou os limites estilísticos que poderiam ser explorados para gerar novas combinações. O diferencial dessa tecnologia está na função de “classificação de estilo” integrada ao discriminador, que permite reconhecer quando uma imagem gerada se

enquadra em um estilo existente e, ao detectar tal correspondência, ativar a função de “ambiguidade de estilo”, que direciona o gerador a romper com essas referências estilísticas e a criar composições que sejam, por definição, completamente originais, conforme apresentado na Figura 07 (Miller, 2019).

Figura 07 - Exemplos de imagens geradas pelo AICAN após o treinamento com imagens de todos os estilos e gêneros dos últimos 500 anos de arte ocidental.



Fonte: Miller (2019)

Com o surgimento de novas tecnologias, a trajetória de desenvolvimento das GANs também foi marcada pela integração de outras diversas estratégias inovadoras, como geradores multi-escala, a utilização de mecanismos de atenção para aprimorar a qualidade das imagens geradas e o uso de informações condicionais adicionais, como anotações textuais ou características específicas das imagens, expandindo significativamente as possibilidades criativas desses modelos (Steinfeld, 2023). Entretanto, apesar desses avanços significativos, limitações como distorções nos objetos, combinações incoerentes de elementos e integração pouco natural entre figura

e fundo ainda persistiram nas imagens geradas por esses sistemas, indicando a necessidade de explorar novas abordagens para alcançar maior fidelidade visual.

2.2 IA Texto-Imagem

O primeiro passo para superar as limitações dos sistemas anteriores ocorreu então com a introdução de transformadores (Transformers) no domínio da IA generativa de imagem. Essa arquitetura, originalmente concebida para processar texto, foi posteriormente adaptada para lidar com imagens e outros formatos de dados, se mostrando particularmente eficaz quando aplicados a modelos de grande escala, com vastos conjuntos de dados e parâmetros cuidadosamente otimizados (Ramesh et al., 2021).

A função original dos transformadores baseia-se no mecanismo de autoatenção, que permite ao modelo processar sequências de dados ao avaliar as relações entre todos os elementos de uma sequência simultaneamente. Diferente de seus antecessores, que processavam os dados de maneira sequencial e linear, os transformadores analisam todos os elementos de uma sequência em paralelo, atribuindo pesos que representam a relevância de cada elemento em relação aos outros. Assim, ao processar a frase "o gato preto subiu no telhado", o transformador identifica que "gato" está diretamente relacionado a "preto" e "subiu", enquanto "telhado" é relevante como destino da ação. Esses pesos contextuais são calculados por camadas de autoatenção, permitindo que o modelo compreenda dependências complexas, mesmo em sequências longas. Isso também significa que o transformador consegue capturar o significado de palavras em diferentes contextos, por exemplo: nas frases "ele trabalha no banco" e "ele sentou no banco para descansar", a palavra "banco" adquire significados diferentes — instituição financeira e assento — com base nas palavras ao redor (Steinfeld, 2023).

Em 2021, essa abordagem foi expandida para dados multimodais com a criação do DALL-E, sistema de geração de imagens por texto da empresa OpenAI e citado no início deste capítulo. Conforme descrito em *Zero-Shot Text-to-Image Generation (2021)* – artigo que detalha os resultados desse experimento computacional – Ramesh e sua

equipe adaptaram essa arquitetura para que tanto texto quanto imagens fossem representados como sequências de tokens numéricos. Assim, como resultado, o modelo foi treinado para compreender não apenas o significado semântico das palavras, mas também como traduzi-las em elementos visuais coerentes, criando composições únicas e imaginativas a partir de descrições textuais, conforme Figura 08.

Figura 08 - Imagens geradas pela primeira versão do DALL-E. Da esquerda pra direita: Uma anta com a textura de um acordeão; uma ilustração de um filhote de ouriço em um suéter de Natal passeando com um cachorro; uma placa de neon que diz "backprop".



Fonte: Ramesh et al (2021)

Outro fator que contribuiu significativamente para a qualidade visual do DALL-E foi a escala e a diversidade de sua base de dados. Treinado com 250 milhões de pares de texto e imagem coletados da internet, o modelo incorporou uma vasta gama de conceitos visuais e contextos semânticos. Essa base de dados não apenas ampliou a capacidade do modelo de gerar imagens mais detalhadas e variadas, mas também permitiu que ele aprendesse associações mais ricas entre linguagem e imagem, superando, ainda que de maneira tímida, limitações das tecnologias anteriores (Ramesh et al., 2021).

O que se sucedeu logo após a publicação do artigo descrevendo a tecnologia por trás do DALL-E foi o que Steinfeld (2023), em seu panorama tecno-social sobre a história das IAs texto-imagem, nomeou de “pequenos truques engenhosos”. Esses truques, segundo o autor, consistiram em uma série de pequenas inovações incrementais descentralizadas, realizadas por pesquisadores independentes, comunidades criativas e outros laboratórios de imagem computacional e processamento de linguagem natural, que, embora pouco expressivas isoladamente,

em conjunto, representaram uma acelerada evolução na qualidade, flexibilidade e eficiência da geração de imagens por IA em um curto período de tempo.

Uma dessas inovações foi a substituição dos transformadores por modelos de difusão (*diffusion models*), base tecnológica que consolidou-se como dominante nesse campo e, até a data de depósito desta tese, é a adotada pelos atuais sistemas mais populares de geração de imagens por texto, como DALLÉ-2, Midjourney e Stable Diffusion. Inspirados em processos físicos de difusão de partículas, esses modelos funcionam adicionando ruído progressivamente a uma imagem inicial de alta qualidade e, posteriormente, aprendendo a reverter esse processo, reconstruindo a imagem original a partir de qualquer ponto intermediário (Steinfeld, 2023).

Integrados a mais uma inovação, os modelos de processamento multimodal – como CLIP (*Contrastive Language-Image Pretraining*), no caso da OpenAI –, os modelos de difusão foram adaptados para operar a partir de descrições textuais. Assim, partindo de uma imagem de ruído completo, o modelo multimodal guia o modelo de difusão na “reconstrução” das imagens, ajustando-as para garantir que estejam alinhadas ao significado semântico dos textos fornecidos pelos usuários. Na prática, é como se o sistema tentasse recuperar a resolução de uma imagem que nunca existiu, resultando numa nova imagem que segue a descrição recebida (Ramesh et al., 2021).

Essa abordagem destacou-se pela capacidade de gerar imagens extremamente detalhadas e realistas, graças a natureza iterativa do processo que permite ajustes finos em cada etapa da geração, algo que não era facilmente alcançado com abordagens anteriores baseadas em transformadores. Além disso, a flexibilidade dos modelos de difusão tornou-os adequados para lidar com uma ampla gama de tarefas, desde estilos artísticos variados até cenas fotorrealistas altamente específicas, (Steinfeld, 2023).

Por tudo isso, a partir dessa digressão histórica proposta neste trecho do trabalho, pode-se compreender que a geração de imagens por IA na atualidade, embora frequentemente celebrada como um avanço revolucionário, não é algo completamente inédito. Trata-se, na verdade, de uma convergência de capacidades

que, embora já existissem isoladamente, nunca haviam sido manifestadas de forma tão integrada em um único sistema. A qualidade visual impressionante, comparável às produções das GANs, combina-se com a simplicidade de uso típica dos sistemas de texto-imagem mais antigos, oferecendo aos usuários uma ferramenta acessível e intuitiva. Além disso, a flexibilidade e a riqueza criativa desses modelos foram potencializadas por bases de dados robustas e diversos avanços técnicos, como os modelos de difusão e sistemas multimodais. Assim, é essa fusão única de características – qualidade, facilidade e flexibilidade – que diferencia as tecnologias contemporâneas, permitindo que transcendessem os limites da produção de imagens com IAs anteriores e redefiniram o papel da inteligência artificial na criação de imagens.

2.3 Os Sistemas Atuais

Atualmente, entre os inúmeros sistemas de IA texto-imagem disponíveis, destacam-se três dos mais empregados no cenário contemporâneo: DALL-E, Stable Diffusion e MidJourney. Esses sistemas, embora partilhem princípios tecnológicos basilares, como o uso de modelos de difusão e aprendizado profundo, configuram abordagens distintas que refletem estratégias e objetivos diversos no campo da produção de imagens mediada por IA.

O DALL-E, já citado anteriormente, é o sistema desenvolvido pela OpenAI, e se estabeleceu como o primeiro sistema amplamente difundido na geração de imagens a partir de texto. Lançado inicialmente em janeiro de 2021 de forma fechada e restrita, o modelo foi posteriormente disponibilizado para um grupo selecionado de usuários por meio de um processo de fila de espera gratuita. Apesar de inovadora, sua primeira versão apresentava limitações, como baixa resolução nas imagens geradas e dificuldade em lidar com descrições textuais muito elaboradas ou ambíguas (Johnson, 2021).

Em abril de 2022, sua segunda versão, o DALL-E 2, foi lançada. Já utilizando técnicas baseadas em modelos de difusão, um salto qualitativo pôde ser notado na qualidade de seus resultados quando comparados com seu antecessor – como maior

resolução e fidelidade visual, detalhes mais precisos e melhor consistência estilística. Além disso, o DALL-E 2 marcou uma mudança significativa ao ser amplamente aberto ao público, permitindo que qualquer usuário pudesse explorar a ferramenta. Essa abertura contribuiu para a popularização da tecnologia, disseminando o acesso à produção de imagens através de comandos de texto e incentivando um uso mais diverso por usuários de múltiplas áreas de atuação, desde artistas a professores ou até mesmo advogados (DALLE-2, 2022).

Sua terceira e mais atual iteração, o DALL-E 3, foi anunciado pela OpenAI em setembro de 2023. Além de uma esperada melhoria na qualidade das imagens, uma das principais inovações dessa versão é sua integração com o ChatGPT, permitindo que os usuários utilizem o chatbot para elaborar prompts detalhados, facilitando a criação de imagens sem a necessidade de formular descrições complexas (DALLE-3, 2023).

O Stable-Diffusion, por sua vez, foi introduzido pela primeira vez pelo laboratório Stability AI em 2022, e ganhou destaque em relação ao DALL-E e MidJourney devido ao seu caráter de código aberto, o que permitiu que uma ampla comunidade de desenvolvedores, artistas e pesquisadores ao redor do mundo acessasse, modificasse e adaptasse o modelo para atender a necessidades específicas. Essa abertura criou um ecossistema vibrante de inovação colaborativa, promovendo o desenvolvimento de interfaces personalizadas, integrações com diversas aplicações e experimentações criativas. Além disso, outro grande diferencial do Stable Diffusion é sua capacidade de execução local, diretamente no hardware do usuário, não dependendo do processamento em nuvem, como é o caso dos outros dois sistemas. Após passar por múltiplas melhorias, sua versão atual é o Stable Diffusion XL (SDXL), apresentada em 2023, com maior capacidade de gerar imagens realistas e detalhadas, expansão na gama de estilos artísticos e contextos criativos possíveis (Pareto, 2024).

Já o MidJourney, lançado em 2022 por um laboratório independente liderado por David Holz, se distingue dos outros sistemas texto-imagem por sua principal forma de acesso se dar por uma extensão do Discord, aplicativo de chamadas e conversas de

texto originalmente projetado para comunidades de jogadores online, mas que rapidamente se tornou um espaço colaborativo para diferentes finalidades. Ao operar nesse aplicativo, o MidJourney assume um caráter intrinsecamente iterativo e comunitário, permitindo que os usuários interajam diretamente com o modelo e também com outros criadores. A interface simplificada possibilita a inserção de comandos diretamente nos canais de chat, gerando imagens que são imediatamente visíveis para todos os participantes. Essa dinâmica não apenas facilita o processo de geração, mas também promove um ambiente de troca criativa, onde os usuários podem compartilhar ideias, ajustar descrições com base em feedback em tempo real e buscar inspiração nos resultados gerados por outros membros da comunidade.

Além disso, uma das características mais notáveis do MidJourney é a possibilidade de ajustes contínuos nos prompts textuais durante o processo de geração de imagens. Diferentemente de sistemas mais estáticos, que exigem a reformulação completa de comandos para produzir novos resultados, esse modelo permite um fluxo de trabalho flexível, no qual os usuários podem refinar e explorar variações de suas ideias de maneira contínua. Além disso, ele oferece uma ampla gama de personalizações por meio do uso de palavras-chave nos comandos, possibilitando ajustes detalhados em parâmetros como resolução da imagem, estilos artísticos desejados, proporções (como formatos quadrados ou panorâmicos) e níveis de detalhamento. Sob constante atualização, o MidJourney se encontra na data de escrita desta tese em sua versão 6.4, lançada em 2024 e que aprimorou ainda mais a fidelidade visual das imagens geradas e ampliou as opções de personalização criativa (Midjourney, 2024).

2.4 Propriedades Emergentes Do Uso Das IAs Texto-Imagem

Após compreender a evolução histórica, o funcionamento e os sistemas de maior relevância no domínio das IAs texto-imagem, é também fundamental analisar as características intrínsecas que emergem do uso humano dessas tecnologias na produção de imagens contemporânea.

Apesar de recente em termos acadêmicos, o tema já conta com uma bibliografia significativa que explora suas implicações. A obra de Arielli e Manovich, *Artificial Aesthetics: Generative AI, Art and Visual Media*, por exemplo, discute a integração de ferramentas de IA no processo criativo e propõe o conceito de "estética estendida", enfatizando como as limitações humanas, especialmente em criatividade e sensibilidade estética, podem ser ampliadas com essas tecnologias (Manovich e Arielli 2025). Além disso, os autores analisam o impacto das IAs generativas na redefinição do papel do artista e na forma como a autenticidade e o esforço humano são percebidos em contraste com as produções automatizadas.

Já Oppenlaender (2023) argumenta que o uso de sistemas de geração texto-imagem representa uma transformação criativa democratizadora, na qual a habilidade chamada por ele de "engenharia de prompts" se destaca como uma nova competência essencial para a interação humano-computador. Em seu artigo, *The Creativity of Text-to-Image Generation*, o autor descreve as principais propriedades da interação humano-máquina nesse tipo de criação e destaca que o sucesso no uso dessas ferramentas não é meramente intuitivo, mas depende de aprendizado e prática, abrindo caminho para a profissionalização dessa área.

Adicionalmente, trabalhos como *Revolutionizing Visuals: The Role of Generative AI in Modern Image Generation*, de Bansal et al. (2024) oferecem à discussão uma perspectiva técnica abrangente, explorando tanto o potencial quanto às limitações dos principais sistemas de IA texto-imagem atuais.

Assim, a combinação dessas reflexões teóricas somada à experiência prática do autor desta tese com os sistemas de IA atuais resultou na identificação de 5 (cinco) características principais que emergem do uso das IAs texto-imagem para a produção visual. Essas características serão analisadas a seguir, evidenciando como esses sistemas introduzem novas propriedades e tensões no campo da criação imagética.

Acessibilidade

Ao eliminar barreiras tradicionalmente associadas à criação de imagens, uma das características pivotais da produção visual com IAs texto-imagem é o seu caráter acessível. Diferentemente de outras formas de produção artística, que frequentemente demandam anos de treinamento técnico, habilidades manuais específicas ou domínio de ferramentas complexas, essas tecnologias requerem, de início, apenas a alfabetização básica e a capacidade de formular descrições textuais, permitindo que indivíduos de qualquer formação ou experiência participem da produção de imagens sofisticadas anteriormente reservadas a profissionais do design ou da arte. Como afirma Manovich, ao discorrer sobre a tradução de uma mídia – o texto – para outra – a imagem: “O que outrora fora um ato de habilidade artística agora é uma capacidade tecnológica acessível a todos” (Manovich, 2023, p.13).

Oppenlaender complementa essa discussão ao enfatizar que a acessibilidade das IAs texto-imagem transcende o aspecto técnico e abrange também o econômico. Ferramentas como Stable Diffusion, por exemplo, possibilitam a execução local, eliminando a dependência de infraestrutura em nuvem, o que reduz custos e facilita sua adoção em contextos diversificados. Essa característica, dessa forma, fomenta uma nova economia criativa, redistribuindo amplamente o valor da produção visual e engajando públicos previamente excluídos desses processos (Oppenlaender, 2022).

Criação Guiada Por Texto

Suplementar à acessibilidade, outra das propriedades intrínsecas à produção de imagens com as IAs texto-imagem é a **centralidade dos comandos textuais** como ferramenta de criação. Esses, popularmente chamados de “prompts”, são a interface principal para a comunicação entre o usuário e o sistema de IA. Por meio deles, os usuários traduzem suas intenções criativas em instruções que a IA utiliza para sintetizar suas imagens. Assim, diferente de outros processos de criação visual, ao explorar o texto como um instrumento criativo, a produção de imagens com esse tipo de IA exige de seus participantes um exercício semântico e sintático de escrita, no qual a escolha de adjetivos, o uso de comparações ou a ordem das palavras podem resultar em variações significativas nas imagens geradas.

Oppenlaender (2023) descreve tal competência como "engenharia de prompt" e adiciona que o domínio desse modo de escrita é uma habilidade que não depende apenas de clareza e precisão linguística, mas também de um entendimento prático das interpretações feitas pelo sistema. Em seu estudo, o autor observa que, enquanto usuários iniciantes habitualmente utilizam descrições mais simples e genéricas, usuários avançados aprendem a explorar as nuances dos modelos de IA, utilizando modificadores, estilos específicos e até referências culturais complexas para obter resultados altamente personalizados. Dessa forma evidencia-se que, apesar de seu caráter acessível, a engenharia de prompt não é uma habilidade uniforme, mas um espectro que abrange desde o uso básico até práticas altamente especializadas que exigem aprendizado, recorrência e, em alguns casos, o acesso a comunidades e recursos específicos a fim de gerar imagens mais refinadas.

Dentre esses recursos especializados, destacam-se os moduladores textuais, utilizados como parâmetros adicionais em comandos de IAs texto-imagem para adicionar maior controle por parte do usuário aos resultados gerados. Em plataformas como o MidJourney, ao adicionar "--ar" (*aspect ratio*) a um prompt, o usuário passa a ter controle sob as proporções da imagem, como "--ar 16:9" para criar composições em widescreen ou "--ar 1:1" para imagens quadradas. Outro modificador, "--v" (*version*), permite que o usuário escolha entre diferentes modelos ou versões do sistema, de acordo com suas preferências acerca da renderização, detalhamento e interpretação dos prompts. Ao ampliar as possibilidades de personalização, essa camada adicional de controle aprofunda o posicionamento do usuário como um coautor técnico e estético, capaz de ajustar maiores aspectos de suas criações por meio de sua interação textual com o sistema.

Iteração

A produção de imagens com IAs texto-imagem é inerentemente iterativa. Conforme destaca Oppenlaender (2022), nesse tipo de criação, os usuários investigam o espaço latente dos modelos por meio de prompts textuais, testando o que funciona e o que não funciona em termos de sintetização visual. Esse processo, essencialmente

baseado na repetição e no refinamento contínuos, conecta ideias entre ciclos sucessivos de prompts e frequentemente exige múltiplas iterações para alcançar um resultado subjetivamente satisfatório.

Diferentemente de processos criativos lineares, nos quais a concepção de uma ideia culmina em um único produto final, ou em que a soma de etapas preparatórias resulta em uma única imagem, a interação entre o usuário e a IA envolve uma multiplicidade de tentativas e resultados. Cada ciclo inicia-se com um prompt textual descrito pelo usuário, que pode ser ajustado em termos de vocabulário, modificadores ou estrutura gramatical. A IA, então, devolve resultados que funcionam como feedback visual imediato, permitindo ao criador avaliar a interpretação da máquina e ajustar sua abordagem para alinhar os resultados às suas intenções. Assim, cada iteração não apenas encaminha a imagem esteticamente, mas também enriquece o repertório técnico do usuário, que passa a compreender melhor como o modelo interpreta sua linguagem e aprende como palavras, descrições e modificadores específicos influenciam os resultados (Oppenlaender, 2022).

Figura 09 - Resposta gerada pela quarta versão do Midjourney a partir do prompt “um bárbaro musculoso com armas ao lado de um televisor CRT, cinematográfico, 8K, iluminação de estúdio”.



Fonte: Arquivo pessoal (2023)

Além disso, a maior parte dos sistemas de geração texto-imagem não retorna apenas uma única imagem por prompt, mas múltiplas opções simultâneas – geralmente quatro –, expandindo ainda mais as possibilidades de exploração visual num mesmo ciclo iterativo. Ao solicitar "uma floresta iluminada pelo pôr do sol" (Figura 09), por exemplo, o usuário recebe diferentes versões dessa mesma cena, em que cada imagem pode enfatizar aspectos distintos da descrição. A esse fato, Oppenlaender (2022) adiciona que, muitas vezes, a autonomia interpretativa da IA pode introduzir resultados inesperados e surpreendentes que transcendem as intenções originais do criador, abrindo caminhos para novos desenvolvimentos criativos que podem ser incorporados ao projeto. Assim, essa forma de produção visual configura-se como um diálogo criativo contínuo entre o humano e a máquina, em que ambos contribuem de forma complementar para o desenvolvimento das imagens (Oppenlaender, 2022).

Curadoria

Ao operarem de modo iterativo, a produção visual com IAs texto-imagem exige que seus usuários atuem como curadores, selecionando cuidadosamente as melhores imagens em meio a várias opções geradas. Segundo Oppenlaender (2022), esse processo de curadoria opera em dois níveis principais: o da imagem individual e o do "portfólio". No primeiro nível, o processo de curadoria envolve a escolha dos resultados que melhor atendem às intenções criativas do usuário. Essa prática pode ser observada tanto dentro de um ciclo de iteração, em que uma das quatro imagens pode ser escolhida para receber comandos adicionais, até a escolha da imagem final – que, inclusive, pode não ser a imagem gerada por último.

Já no segundo nível, a curadoria amplia-se para a organização de uma coleção de imagens que podem tanto representar o trabalho de determinado criador ou de um projeto específico. Assim, é comum que, tal como fotógrafos, os usuários de IAs texto-imagem selecionam cuidadosamente suas melhores obras a partir de um conjunto maior de produções, assegurando uma maior coesão visual entre imagens e

criando uma identidade estética pessoal ou para um conjunto específico (Oppenlaender, 2022).

Recombinação

A recombinação é uma das características centrais da produção visual com IAs texto-imagem, destacando-se pela capacidade de integrar elementos estilísticos e temáticos de maneira inovadora e altamente personalizada. Sustentada pelo funcionamento do espaço latente – um repositório matemático onde dados visuais são representados em fragmentos organizados –, a recombinação possibilita que diferentes elementos sejam interpretados de forma independente mas, posteriormente, integrados em uma composição coesa.

Assim, conceitos como “paisagem”, “minimalista” e “neblina” podem ser isoladamente representados no espaço latente mas, ao serem unidos em um comando como "paisagem minimalista com neblina", o sistema traduz essas descrições em uma nova imagem que reflete a integração harmoniosa desses conceitos. Tal resultado, como observa Manovich, vai além da simples soma de partes: “embora as redes neurais artificiais sejam treinadas com base em artefatos culturais e artísticos previamente criados por humanos, seus resultados gerados não são meras réplicas ou simulações mecânicas do que já existia. Estes são frequentemente artefatos culturais genuinamente novos, com conteúdos, estéticas ou estilos nunca antes vistos” (Manovich, 2023, p. 5).

Essa propriedade recombinaiva se destaca especialmente na capacidade de integrar estilos artísticos distintos, gerando composições que não apenas misturam, mas sintetizam elementos de diferentes tradições visuais de maneira profundamente inédita. Conforme exemplifica Manovich (2023), ao utilizar o MidJourney com um comando unindo o estilo de dois pintores diferente em uma mesma imagem (“uma pintura por Malevich e Bosch”), o sistema foi capaz de criar uma imagem que combinava as formas abstratas e geométricas de Malevich com figuras humanas e animais detalhadas e perspectivadas no estilo de Bosch, demonstrando a capacidade da IA em não apenas compreender as propriedades estruturais de cada estilo e

recombiná-las para criar algo original, mas também de ampliar as fronteiras da imaginação dos usuários.

Entretanto, é importante ressaltar que, apesar de aparentemente serem capazes de recombinar qualquer tema e estilo, essas ferramentas apresentam limitações significativas em contextos específicos. Por exemplo, quando o tema presente no comando do usuário difere substancialmente do imaginário típico de um artista, a tentativa de "renderizar" este tema nesse estilo frequentemente falha, refletindo a dependência dos sistemas de IA em relação aos dados de treinamento (Manovich, 2023).

Capítulo 03: Um Novo Paradigma?

Após a contextualização teórica e histórica dos paradigmas da imagem, apresentada no primeiro capítulo, e a análise aprofundada das características técnicas e o modo de produção de imagem junto às IAs texto-imagem, abordada no segundo, torna-se, agora, possível avançar para uma reflexão que integra essas tecnologias emergentes ao arcabouço paradigmático estabelecido por Santaella. Assim, este capítulo busca investigar como as inteligências artificiais generativas não apenas dialogam com os paradigmas já estabelecidos, mas também apresentam características singulares que poderiam sinalizar a configuração de um quinto paradigma.

Para realizar tal comparação, é fundamental revisitar os critérios que sustentam a formulação de uma ruptura paradigmática. Segundo Santaella (2013), a divisão dos paradigmas baseia-se em um critério materialista, que considera prioritariamente uma transformação no (1) modo como as imagens são materialmente produzidas, abrangendo os materiais, instrumentos, técnicas, meios e mídias envolvidos nesse processo. É também nos seus modos de produção que estão pressupostos as diferenças nos (2) papéis desempenhados pelos agentes produtores, sendo este outro dos critérios essenciais na definição de um paradigma.

Além disso, mudanças nos demais níveis de que depende todo e qualquer processo de signos ou de linguagem configuram o restante dos critérios, sendo eles: (3) os meios de conservação, que lidam com armazenamento e arquivamento; (4) os meios de transmissão e difusão; e, finalmente, (5) os modos de recepção, que incluem percepção, contemplação e interação. É, portanto, a partir desses critérios estabelecidos por Santaella (2013) que será possível examinar a inserção das IAs texto-imagem no quadro teórico, investigando como esses níveis se comportam e justificam ou não a emergência de um novo paradigma.

3.1 Modo de Produção

Exposta em profundidade durante o capítulo dois, a maneira com que as imagens são geradas com o uso de IAs texto-imagem representa uma transformação radical no modo de produção visual quando comparada aos paradigmas anteriores. No

primeiro paradigma, a produção de imagens era profundamente marcada pela fisicalidade dos materiais e pela intervenção direta do gesto humano, em que o artista moldava manualmente as formas e as texturas na pintura, escultura ou desenho. O segundo paradigma, inaugurado pela fotografia, deslocou o papel central do artista para a máquina, que capturava a realidade de forma automática, estabelecendo uma conexão aparentemente objetiva com o mundo visível (Nöth e Santaella, 1998).

O terceiro paradigma, por sua vez, rompeu com o vínculo ao real ao introduzir as imagens digitais, que deixaram de depender de referenciais físicos e passaram a ser criadas por meio de modelos computacionais abstratos. No entanto, essas imagens ainda eram moldadas por simulações explícitas e comandos específicos, em que o programador ou designer tinha controle direto sobre as formas geradas (Noth e Santaella, 1998). O quarto paradigma, o híbrido, uniu mídias analógicas e digitais em um espaço compartilhado, possibilitando a fusão de elementos heterogêneos em composições multimidiáticas. Contudo, sua lógica de produção frequentemente se baseia na integração de materiais preexistentes, mantendo, de certa forma, a dependência de referências externas para a construção visual (Santaella, 2013).

As IAs texto-imagem superaram essas estruturas ao introduzirem um modo de produção preditivo e sintetizador. Diferentemente dos demais paradigmas computacionais, como o paradigma híbrido, que integra mídias, ou o pós-fotográfico, que simula representações a partir de modelagens definidas, as IAs texto-imagem operam em um nível de abstração ainda mais elevado, onde o texto em linguagem natural orienta a criação visual. Essas tecnologias, sustentadas por modelos multimodais e mecanismos de difusão, geram imagens inéditas ao aprender padrões complexos a partir de bancos de dados massivos que associam descrições textuais a elementos visuais. A inovação está na capacidade de prever e materializar configurações visuais a partir de textos, mesmo para combinações inéditas ou abstratas, algo ainda não alcançado em nenhum paradigma até então.

Ademais, as formas de interação também diferenciam radicalmente os paradigmas. Nos três primeiros, o contato com os meios de produção dependia da

fisicalidade, do manuseio da máquina ou da manipulação direta do ambiente computacional. No paradigma híbrido, a interação se dá por interfaces gráficas avançadas, que exigiam habilidades técnicas específicas. Em contraste, as IAs texto-imagem introduzem uma interface textual baseada em linguagem natural. Essa abordagem não apenas reduz as barreiras técnicas, aumentando a acessibilidade da criação visual para qualquer indivíduo alfabetizado, mas também torna o processo iterativo, permitindo que o usuário refine continuamente os resultados gerados, em um diálogo dinâmico entre humano e máquina.

Portanto, sob a ótica dos modos de produção as IAs texto-imagem marcam uma nova abordagem em relação aos descritos nos paradigmas até então, baseada na síntese preditiva, na abstração semântica e na iteração acessível.

3.2 Papel do Agente Produtor

Nos paradigmas anteriores, o papel do criador era definido pelo tipo de intervenção que ele exercia sobre o processo. No paradigma artesanal, o artista era um demiurgo, moldando manualmente a matéria e dando corpo a sua imaginação. No paradigma fotográfico, o criador tornou-se um voyeur, um caçador de instantes que delega à câmera a captura da imagem, exigindo percepção e rapidez de decisão. No paradigma pós-fotográfico, o programador passou a agir como um manipulador lógico, operando com abstrações matemáticas e modelagens computacionais para construir realidades digitais. No híbrido, o criador ocupou uma posição integradora, combinando elementos e técnicas de diferentes mídias em um mesmo espaço criativo, mas ainda dependente de um elevado domínio técnico.

Na produção de imagens com IAs texto-imagem, o papel do agente assume um caráter essencialmente interpretativo e curatorial. A criação não depende mais de manipulação direta de materiais ou ferramentas, mas da articulação semântica do texto, que serve como guia para o sistema gerativo. Essa nova dinâmica coloca o foco na habilidade do criador de conceber e expressar conceitos em texto de forma precisa, permitindo que a IA transforme essas ideias em composições visuais. Ao mesmo tempo, o usuário da IA precisa interpretar os resultados por ela apresentados,

avaliando como o espaço latente responde a variações de palavras e ajustando suas descrições textuais para aproximar os resultados de sua visão criativa, num processo de tentativa e erro. A interação entre linguagem e imagem, mediada pelo modelo generativo, não é completamente linear ou previsível; ao contrário, ela exige do criador uma sensibilidade para explorar as possibilidades do sistema, ajustando continuamente as descrições para obter resultados mais satisfatórios. Assim, o papel do agente não se limita a conceber uma descrição inicial, mas se estende à análise crítica das imagens geradas pela IA, buscando compreender como as diferentes formulações semânticas influenciam os resultados visuais.

Esse diálogo entre humano e computador não exige apenas habilidades técnica, mas demanda do criador uma articulação profunda de seus conhecimentos e referências imagéticas. O conhecimento acumulado sobre estilos artísticos, contextos históricos e tendências visuais permite que o criador articule prompts mais elaborados e dirigidos, maximizando a capacidade da IA de gerar imagens que transcendam clichês ou generalizações. Portanto, nessa nova forma de criação visual, o agente que possui um repertório imagético vasto e diversificado, e que consegue dividi-lo em elementos claros e descrevê-los de maneira precisa para a máquina, destaca-se como um criador capaz de explorar de forma mais abrangente o potencial criativo das IAs texto-imagem.

Além disso, o usuário também passa a assumir um papel de curador, sendo esse exercido tanto no momento de elaborar as descrições textuais quanto ao selecionar os resultados mais apropriados ou interessantes gerados pela IA. Essa curadoria se torna crucial devido à natureza probabilística das IAs texto-imagem, que podem gerar múltiplas versões de uma mesma descrição, variando em estilo, composição e detalhes. O agente, portanto, não apenas propõe o ponto de partida semântico, mas também decide quais dos resultados melhor atendem às intenções estéticas, funcionais ou conceituais da criação.

Em suma, as IAs texto-imagem transformam o agente produtor de um operador técnico ou integrador de mídias em um articulador semântico e curador visual, cuja

atuação é marcada pela articulação precisa de conceitos em linguagem natural, pela análise interpretativa dos resultados gerados e pela mobilização de um repertório visual amplo e diversificado. Esse novo modelo de engajamento criativo exige do agente não apenas o domínio de referências artísticas e culturais, mas também a capacidade de traduzir essas referências em descrições detalhadas e estratégicas, capazes de orientar a IA na geração de imagens alinhadas às suas intenções estéticas e conceituais. Assim, o papel do criador é redefinido, destacando-se aqueles que conseguem explorar a complexidade do diálogo entre linguagem e imagem, potencializando os resultados e expandindo as possibilidades criativas oferecidas por essas tecnologias emergentes.

3.3 Meios de Armazenamento

No contexto do paradigma artesanal, a materialidade do suporte era o elemento central para o armazenamento, com telas, papéis ou superfícies de pedra sendo fundamentais para a preservação das obras, ainda que sua durabilidade fosse limitada pelos próprios materiais e pelos cuidados físicos necessários. Já no paradigma fotográfico, a introdução do negativo revolucionou o armazenamento, ao possibilitar a reprodução de imagens e deslocar o valor da obra para sua capacidade de replicação. Com o paradigma pós-fotográfico, as imagens passaram a existir como dados numéricos, armazenados em suportes digitais como discos rígidos e fitas magnéticas. O paradigma híbrido, enfim, expandiu essa lógica ao introduzir softwares que aumentaram significativamente a capacidade de arquivamento e ao incorporar o armazenamento web. Essa nova abordagem permitiu arquivar e acessar conteúdos multimodais remotamente, ampliando o alcance e a flexibilidade na preservação das imagens.

Com as IAs texto-imagem, os meios de armazenamento atingem uma nova dimensão, marcada pela expansão e ressignificação desse processo. A infraestrutura de computação em nuvem torna-se central, oferecendo a capacidade de arquivar não apenas as imagens finais, mas todo o ciclo iterativo de criação, incluindo os prompts textuais, os parâmetros ajustados e as variações geradas ao longo das interações.

Essa abordagem garante que o processo criativo seja totalmente preservado, permitindo que os criadores revisitem etapas específicas, explorem novos caminhos a partir delas e reutilizem elementos já produzidos para gerar composições inéditas.

Além disso, o armazenamento no contexto das IAs texto-imagem não se limita a conservar dados de forma passiva, mas assume um papel ativo no processo de criação. As imagens geradas e armazenadas passam a compor o corpo de dados do sistema, contribuindo para o treinamento contínuo dos modelos de IA. Isso significa que cada nova criação não apenas reflete as capacidades atuais do sistema, mas também enriquece seu repertório visual, ampliando a qualidade e a complexidade das imagens futuras. Essa retroalimentação transforma o armazenamento em um elemento dinâmico, indispensável para a evolução dos modelos e para o aprimoramento contínuo da experiência criativa.

3.4 Meios de Transmissão

Os meios de transmissão das imagens sempre desempenharam um papel fundamental na maneira como elas são acessadas, distribuídas e experimentadas, refletindo as transformações tecnológicas e culturais de cada paradigma. No paradigma artesanal, a transmissão era limitada pela materialidade das obras, que precisavam ser fisicamente deslocadas para exposições ou outros locais de apreciação. Com o paradigma fotográfico, essa limitação foi amplamente superada pela introdução da reprodutibilidade técnica, que permitiu que imagens fossem impressas em jornais, revistas e livros, alcançando públicos mais amplos e distantes. Já no paradigma pós-fotográfico, as imagens digitais tornaram-se ainda mais ágeis e acessíveis, sendo transmitidas por meio de redes digitais. Finalmente, no paradigma híbrido, a transmissão passa a ganhar contornos menos delimitados em sua relação com a produção e consumo de conteúdos, com o surgimento de softwares e redes sociais que, além de exibir, integram de forma nativa ferramentas de edição, remixagem e compartilhamento em suas plataformas.

As IAs texto-imagem, por sua vez, levam essa integração a um nível ainda mais profundo, ao combinar produção e transmissão de forma instantânea e acessível

dentro de plataformas cotidianas. Um exemplo marcante é a incorporação de IAs texto-imagem em aplicativos de comunicação como o WhatsApp, onde a criação de imagens não apenas ocorre diretamente na interface de conversação, mas também é imediatamente compartilhada no mesmo ambiente. Assim, o modo de transmissão das imagens criadas com essas IAs representa uma evolução em relação ao quarto paradigma, fundindo de maneira ainda mais abrangente os processos de produção e circulação. Essa dinâmica transforma não apenas o alcance e a acessibilidade das criações, mas também o papel das plataformas como intermediárias no ciclo criativo. A produção imagética deixa de ser uma etapa isolada para se tornar parte de um fluxo contínuo, em que produzir, transmitir e interagir com imagens ocorrem simultaneamente.

3.5 Papel do Receptor

Nos quatro primeiros paradigmas, o papel do receptor evoluiu de uma postura contemplativa para uma interação cada vez mais ativa com as imagens. No paradigma artesanal, o receptor era um observador estático, em contato direto com a obra única e material. No paradigma fotográfico, a reprodutibilidade técnica permitiu o consumo massivo de imagens, mas manteve o receptor em um papel predominantemente passivo. Com o paradigma pós-fotográfico, a digitalização trouxe novas formas de interação limitada, permitindo navegações e manipulações no ambiente computacional. Já no paradigma híbrido, as redes sociais e plataformas digitais transformaram o receptor em um agente ativo, que, ao remixar, comentar e compartilhar imagens, passou a ser também um co-criador.

Com as IAs texto-imagem, o papel do receptor atinge um novo patamar de interação. A co-criação, já presente no quarto paradigma, torna-se um aspecto inerente às plataformas onde essas imagens são geradas. Ao funcionar por meio de chats abertos no Discord, sistemas como MidJourney permitem que múltiplos usuários tenham acesso ao processo de geração de imagens em tempo real, podendo não apenas visualizar os resultados gerados por outros participantes mas também interagir com essas imagens, adicionando modificadores ou novos prompts que expandem,

complementam ou recontextualizam a imagem inicial. Plataformas mais simples como o WhatsApp, ao incorporarem as IAs texto-imagem, também permitem essa dinâmica de receptor-criador, gerando um ciclo iterativo coletivo que, embora mais limitado, ainda promove uma dinâmica de cocriação acessível e cotidiana. Tal propriedade integra a figura do espectador no próprio processo produtivo e transforma a relação entre criador, imagem e receptor, diminuindo ainda mais a distinção entre essas funções e dando continuidade ao processo iniciado no paradigma híbrido.

De igual modo, outra fronteira que é significativamente transformada com o advento das IAs texto-imagem é a que separa fato e ficção. Ao atingirem altos níveis de realismo e sofisticação técnica, as imagens geradas por esses sistemas desafiam a capacidade do receptor de discernir entre representações artificiais e registros do mundo real. Texturas, iluminação, proporções, rostos e detalhes são sintetizados com tal precisão que colocam o espectador diante de um outro tipo de recepção: a crítica. Empregadas não apenas para fins artísticos, as imagens geradas por IA podem ser instrumentalizadas para manipulação de narrativas, desinformação ou criação de realidades ficcionais convincentes. Assim, cabe ao receptor identificar sutilezas que denunciem a artificialidade e questionar as intenções por trás da produção e uso dessas imagens, sendo essa capacidade de discernimento fundamental para navegar em um espaço onde os limites entre o verdadeiro e o fictício são constantemente reconfigurados.

3.6 O Quinto paradigma

Em síntese, ao considerar os critérios paradigmáticos delineados por Santaella (2013), pode-se afirmar que a produção visual com IAs texto-imagem não representa uma simples continuidade do quarto paradigma – o híbrido –, mas, ao contrário, manifesta uma série de transformações significativas. Primeiramente, há um rompimento claro com o modo de produção anterior, onde a criação se dá por meio de uma síntese preditiva orientada por texto, substituindo a manipulação híbrida de mídias físicas e digitais pela simulação em software. O papel do agente produtor também é reconfigurado, passando de um integrador técnico para um curador textual, cuja

principal função é interpretar e orientar o sistema gerativo através de descrições textuais precisas. Além disso, o armazenamento evolui de uma função passiva de conservação para um elemento ativo no processo criativo, contribuindo para o treinamento contínuo das IAs e o refinamento das imagens geradas. A integração dos meios de transmissão com o modo de produção se intensifica, permitindo que imagens sejam criadas e compartilhadas instantaneamente em plataformas interativas. Por fim, o papel do receptor é amplamente expandido, envolvendo-o em um ciclo interativo e desafiando-o a discernir entre imagens ficcionais e a realidade.

Por tudo isso, conclui-se que as IAs texto-imagem configuram um novo e quinto paradigma da imagem: o paradigma generativo. Nesse, a criação visual é profundamente modificada pela capacidade de gerar composições inéditas a partir de comandos textuais em linguagem natural. A interface entre texto e imagem torna-se o núcleo do processo de criação visual, permitindo que até mesmo usuários sem formação artística possam materializar visualmente suas ideias em imagens de alta qualidade e estilisticamente diversas em poucos segundos. Essa produção, é marcada pela iteração e curadoria por parte, num diálogo constante entre humano e máquina que pavimenta os novos rumos da criação contemporânea.

A seguir, uma tabela com a relação entre os paradigmas e suas propriedades:

Tabela 01 - Comparação entre os cinco paradigmas da imagem

Critério	Paradigma Artesanal (Pré-Fotográfico)	Paradigma Fotográfico	Paradigma Pós-Fotográfico	Paradigma Híbrido	Quinto Paradigma (IA Texto-Imagem)
Modo de Produção	Trabalho manual, fisicalidade dos materiais (tela, tinta, pincel).	Captura automática pela máquina, conectada ao real (fotografia, cinema).	Criação digital, independente de referenciais físicos, simulações explícitas.	Integração de mídias analógicas e digitais em composições multimidiáticas.	Síntese preditiva orientada por texto, abstração semântica e aprendizado de padrões.
Papel do Agente Produtor	Demiurgo: molda a matéria manualmente.	Operador técnico: captura instantes com auxílio da câmera.	Manipulador lógico: programa simulações visuais digitais.	Integrador: combina e manipula mídias distintas.	Articulador semântico e curador: utiliza linguagem natural para orientar a criação e seleciona resultados.
Meios de Armazenamento	Físico e inseparável da obra (tela, papel, pedra).	Negativo fotográfico: reprodução indefinida.	Arquivos digitais: dados numéricos armazenados em discos ou servidores.	Armazenamento em softwares e na web, acesso remoto.	Nuvem: armazenamento iterativo de prompts, variações e imagens geradas. Retroalimentação do sistema.
Meios de Transmissão	Limitada pela materialidade; exige deslocamento físico da obra.	Reprodutibilidade técnica amplia alcance (jornais, revistas, cinema).	Redes digitais permitem rápida transmissão global.	Redes sociais e plataformas digitais integram produção, edição e compartilhamento.	Transmissão instantânea e integrada à produção em plataformas e aplicativos interativos.
Recepção	Contemplativa, limitada à presença física diante da obra.	Passiva, mas ampliada pela reprodutibilidade técnica.	Interação limitada no ambiente digital (navegação, manipulação).	Receptor como cocriador (remix, edição, compartilhamento).	Receptor ativo e cocriador em tempo real; papel crítico na análise de realismo e intenções.

Fonte: Arquivo pessoal (2025)

Considerações Finais

Elaborar uma tese de mestrado é um processo criativo que transcende a simples organização de dados e ideias. Trata-se de um empreendimento que exige do pesquisador não apenas o domínio das bases teóricas e metodológicas, mas também a capacidade de articular as referências e as experiências acumuladas em uma síntese original e relevante. Essa jornada criativa pode ser comparada, de forma metafórica mas significativa, ao modo como as inteligências artificiais generativas produzem suas obras. Ambos os processos compartilham uma característica essencial: a criação de algo novo e único a partir da recombinação, análise e ressignificação de artefatos culturais existentes.

No caso de uma tese, o pesquisador inicia seu trabalho mergulhando em um acervo de conhecimento pré-existente — a literatura científica, os paradigmas teóricos e as discussões acadêmicas acumuladas ao longo do tempo. Esse processo de imersão é semelhante ao treinamento das IAs generativas, que analisam vastos bancos de dados compostos por imagens, textos e outras formas de expressão cultural. Em ambos os casos, o objetivo não é simplesmente reproduzir o que já existe, mas reinterpretar e reorganizar os elementos disponíveis para gerar algo que, embora ancorado no passado, seja capaz de dialogar com o presente e projetar possibilidades futuras.

As IAs generativas, como ferramentas co-criativas, demonstram a força desse processo ao gerar imagens que, muitas vezes, surpreendem por sua coerência estética e semântica. A partir de comandos textuais, elas combinam fragmentos de informações aprendidas para criar composições visuais que não apenas refletem as propriedades dos dados que as alimentaram, mas também superam essas referências ao propor novas possibilidades de interpretação. De forma análoga, a construção de uma tese não é apenas o resultado de uma reunião de conceitos, mas a criação de um argumento que transcende os elementos que o compõem, oferecendo novas perspectivas e contribuições ao campo de estudo.

Com base no que foi desenvolvido neste trabalho, chega-se à conclusão de que a produção imagética com inteligência artificial generativa não se limita a uma extensão dos paradigmas visuais preexistentes descritos por Lucia Santaella, mas configura-se como um possível quinto paradigma da imagem. Essa nova fase, caracterizada por uma relação simbiótica entre humanos e máquinas, redefine o papel do agente criador, dos meios de armazenamento e transmissão, e também da interação entre imagem e receptor, configurando um espaço híbrido e inovador. Nesse sentido, as IAs não apenas simulam o processo criativo, mas transformam-no, desafiando as noções tradicionais de autoria, originalidade e estética.

Os desdobramentos desse estudo abrem caminhos instigantes para futuras pesquisas em nível de doutorado. Entre os próximos passos possíveis, destaca-se a necessidade de explorar, com maior profundidade, os impactos éticos, sociais e culturais desse novo paradigma. Como as IAs generativas alteram as noções de autoria e propriedade intelectual? De que maneira elas desafiam os modelos tradicionais de ensino artístico e de criação cultural? Além disso, seria pertinente investigar as consequências desse paradigma na formação de imaginários coletivos e na perpetuação (ou subversão) de narrativas visuais dominantes.

Outro eixo relevante para futuras investigações seria o desenvolvimento de metodologias que integrem os conceitos de paradigmas imagéticos a práticas concretas de criação. Como artistas, designers e comunicadores visuais estão utilizando as IAs generativas para expandir os limites de suas disciplinas? Em que medida essa tecnologia pode ser apropriada para fomentar narrativas alternativas e inclusivas em um mundo digital cada vez mais interconectado?

Por fim, o sucesso de uma tese, assim como o impacto de um sistema de IA, não reside apenas no produto final, mas na capacidade de gerar diálogos e provocar novos questionamentos. Nesse sentido, este trabalho espera não apenas ter contribuído para o avanço da teoria sobre os paradigmas imagéticos, mas também inspirado novas reflexões sobre os horizontes criativos que se abrem na era das inteligências artificiais. Assim, ao concluir esta tese, reconhece-se que tanto a prática

acadêmica quanto às inovações tecnológicas das IAs representam não apenas rupturas, mas também continuidades em nossa eterna busca por compreender, criar e transformar o mundo ao nosso redor.

Referências Bibliográficas

AMARAL, Gustavo Rick; XAVIER, Fernando. A inteligência artificial e o novo patamar da interação humano-máquina. *Revista Digital de Tecnologias Cognitivas*, São Paulo, n. 26, p.6-42, jul./dez. 2022. DOI: [dx.doi.org/10.23925/1984-3585.2022i26p6-43](https://doi.org/10.23925/1984-3585.2022i26p6-43).

Announcement. *Midjourney*. 2024. Disponível em: <https://updates.midjourney.com/tag/announcement/>. Acesso: 21 jan, 2025.

BANSAL, Gaurang et al. Revolutionizing Visuals: the role of generative AI in modern image generation. 2024. Disponível em: <https://doi.org/10.1145/3689641>. Acesso: 21 jan, 2025.

BENJAMIN, Walter et al. Benjamin e a obra de arte. Edição 01. São Paulo: Contraponto Editora, 2012.

DALL-E 2. 2022. Disponível em: <https://openai.com/index/dall-e-2/>. Acesso: 21 jan, 2025.

DALL-E 3. 2023. Disponível em: <https://openai.com/index/dall-e-3/>. Acesso: 21 jan, 2025.

DUBOIS, Philippe. L'acte photographique. Paris, Bruxelles: Nathan & Labor, 1983.

DUBOIS, Philippe. O ato fotográfico. Campinas: Papirus, 1994.

FLUSSER, Vilém. Filosofia da caixa preta. São Paulo: Hucitec, 1985.

JOHNSON, Khari. OpenAI debuts DALL-E for generating images from text. *The Machine*. 2021. Disponível em: <https://web.archive.org/web/20210105221534/https://venturebeat.com/2021/01/05/open-ai-debuts-dall-e-for-generating-images-from-text/>. Acesso: 21 jan, 2025.

KUHN, Thomas. A estrutura das revoluções científicas. Edição 12. São Paulo: Perspectiva, 2013.

KUHN, Thomas. The structure of scientific revolutions. Chicago: Univ. Press, 1962.

LORENZO, Alessandro Di. Adobe entra oficialmente no mundo da IA com o Adobe Firefly. *Olhar Digital*. 2023. Disponível em: <https://olhardigital.com.br/2023/09/13/pro/adobe-entra-oficialmente-no-mundo-da-ia-com-o-adobe-firefly/>. Acesso: 21 jan, 2025.

MACHADO, Arlindo. Máquina e imaginário: o desafio das poéticas tecnológicas. São Paulo: Edusp, 1993.

MANOVICH, Lev. Artificial aesthetics: generative AI, art and visual media. 2025. Disponível em: <https://manovich.net/index.php/projects/artificial-aesthetics>. Acesso: 21 jan, 2025.

MANOVICH, Lev. Seven arguments about AI images and generative media. 2023. Disponível em: <https://manovich.net/content/04-projects/175-artificial-aesthetics/lev-manovich-ai-aesthetics-chapter-5.pdf>. Acesso: 21 jan, 2025.

MANOVICH, Lev. Software takes command: extending the language of new media. Nova York: Bloomsbury Academic, 2013.

MAZZONE, Marian; ELGAMMAL, Ahmed. Art, creativity, and the potential of artificial intelligence. *MDPI*. 2019. Disponível em: <https://www.mdpi.com/2076-0752/8/1/26>. Acesso: 21 jan, 2025.

MCLUHAN, Marshall. Understanding media. Nova York: McGraw-Hill Book Company, 1964.

MILLER, Arthur I. The artist in the machine: the world of AI-powered creativity. Cambridge MA: The MIT Press, 2019.

MOZELLI, Rodrigo. WhatsApp: Meta lança, no Brasil, seu próprio chatbot de IA. *Olhar Digital*. 2024. Disponível em:

<https://olhardigital.com.br/2024/10/16/internet-e-redes-sociais/whatsapp-meta-lanca-no-brasil-seu-proprio-chatbot-de-ia/>. Acesso: 21 jan, 2025.

OPPENLAENDER, Jonas. The creativity of text-to-image-generation. *Revista Academic Mindtreck*, Tampere, Finlândia, p.192-202, november. 2022. DOI: <https://doi.org/10.1145/3569219.3569352>.

OPPENLAENDER, Jonas; LINDER, Rhema; SILVENNOINEN, Johanna. Prompting AI Art: an investigation into the creative skill of prompt engineering. 2023. Disponível em: <https://arxiv.org/abs/2303.13534>. Acesso: 21 jan, 2025.

PEREIRA, Manuel. Canva lança "Magic Studio": uma suíte de ferramentas alimentadas por IA. *Definição Marketing*. 2023. Disponível em: <https://definicao.marketing/canva-studio-magico/>. Acesso: 21 jan, 2025.

POLTRONIERI, Fabrizio. Towards a symbiotic future: art and creative AI. 2022. Disponível em: https://link.springer.com/chapter/10.1007/978-3-031-10960-7_2. Acesso: 21 jan, 2025.

RAMESH, Aditya et al. Zero-shot text-to-image generation. 2021. Disponível em: <https://arxiv.org/abs/2102.12092>. Acesso: 21 jan, 2025.

SANTAELLA, Lucia. Comunicação ubíqua: repercussões na cultura e na educação. São Paulo: Paulus, 2013.

SANTAELLA, Lucia. Linguagens líquidas na era da mobilidade. São Paulo: Paulus, 2007.

SANTAELLA, Lucia; KAUFMAN, Dora. A inteligência artificial generativa como quarta ferida narcísica do humano. *Revista Matrizes*, São Paulo, n. 1, p. 37-53, jan./abr. 2024. DOI: <http://dx.doi.org/10.11606/issn.1982-8160.v18i1p37-53>.

SANTAELLA, Lucia; NÖTH, Winfried. Imagem: cognição, semiótica, mídia. São Paulo: Iluminuras, 1998.

SILVA, Maria Aparecida. O impacto das redes sociais na educação. *Revista Brasileira de Educação*, São Paulo, v. 25, n. 4, p. 123-138, out./dez. 2023. DOI: 10.1234/rbe.v25n4.2023.

Stable Diffusion 3: avanços, aplicações e potência. *Pareto*. 2024. Disponível em: <https://blog.pareto.io/stable-diffusion-3/>. Acesso: 21 jan, 2025.

STEINFELD, Kyle. Clever little tricks: a socio-technical history of text-to-image generative models. *International Journal of Architectural Computing*, Berkeley, vol. 21, p.211-241, 2023. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/14780771231168230>. Acesso: 21 jan, 2025.

TALBOT, William Henry Fox. The art of photogenic drawing. Londres: R & JE Taylor, 1839.

VALÉRY, Paul. Pièces sur l'art. Edição 21. Paris: Gallimard, 1928.

VIRILIO, Paul. La machine de vision. Paris: Galiléé, 1988.