

Imagens Malvadas¹

Autoria: Hito Steyerl²
Universidade de Artes de Berlim

Tradução: André Antônio Barbosa³
Universidade Federal de Pernambuco

Tradução: Hermano Callou⁴
Universidade Federal do Rio de Janeiro

Resumo: Neste texto, a artista e ensaísta Hito Steyerl aborda as imagens geradas por IA a partir dos múltiplos sentidos da palavra inglesa *mean*. Com seu humor característico, a autora afirma tratar-se de *imagens médias* (*mean*), pois são geradas através de uma média de dados estocados num banco para aprendizado de máquina. Porém, também são imagens malvadas (*mean*), pois a média que as gera se baseia num funcionamento estatístico que pode se prestar a fins higienistas e até mesmo eugenistas. Esteticamente, são imagens medíocres, médias, mas social e economicamente nefastas, pois têm como base uma rede perversa de microtrabalhadores precários do capitalismo contemporâneo e ensaiam a criação de um espectador que transforma a vida em jogo de cassino.

Palavras-chave: Inteligência Artificial; aprendizado de máquina; teoria da imagem; trabalho precarizado.

O escritor de ficção científica Ted Chiang descreveu recentemente os resultados gerados a partir de texto do ChatGPT como uma “JPEG embaçada de todo texto disponível na internet” — ou como uma “imagem pobre” semântica⁵. No entanto, os resultados gerados por redes de aprendizado de máquina possuem uma dimensão histórica adicional: a estatística. As imagens criadas por ferramentas de aprendizado de máquina são representações estatísticas, e não imagens de objetos que realmente existem. Elas deslocam o foco da indexicalidade fotográfica para a seleção estocástica. Não se referem mais à factualidade, muito menos à verdade, mas

¹ Nota dos tradutores: o ensaio de Hito Steyerl brinca ao longo do texto com os múltiplos sentidos da palavra *mean* em inglês, que nos esforçamos para conservar na tradução. O título *Mean images* exprime a compreensão da autora de que imagens geradas por aplicativos de IA são “imagens médias” (*mean images*), renderizações probabilísticas. O título também sugere que elas são imagens mesquinhas, medíocres e malvadas (*mean images*), uma formulação provocadora que preferimos manter no título.

² Hito Steyerl é professora de cinema e vídeo experimental e cofundadora do Centro de Pesquisa para Políticas de Procuração na Universidade de Artes de Berlim. Ela estudou cinematografia e documentário em Tóquio e Munique, e defendeu sua tese de doutorado na Academia de Belas Artes de Viena. Sua pesquisa concentra-se em mídia, tecnologia e na distribuição de imagens. E-mail: h.steyerl@udk-berlin.de.

³ Professor adjunto no Departamento de Artes da UFPE. Graduado e Mestre em Comunicação pela UFPE e doutor em Comunicação e Cultura pela UFRJ, linha de pesquisa Tecnologias da Comunicação e Estéticas. Foi professor substituto do bacharelado em Cinema e Audiovisual da UFPE entre 2017 e 2019. Foi também professor na Universidade Católica de Pernambuco (Unicap), lecionando no Curso Superior de Tecnologia em Fotografia com ênfase na relação entre Fotografia e Artes Visuais. É diretor e roteirista de cinema junto à Ponte Produtoras e ao coletivo Surto & Deslumbramento. Atua também como diretor de arte e montador de cinema e vídeo. ORCID: <https://orcid.org/0009-0006-8084-4837>. Lattes: <http://lattes.cnpq.br/5082481076087246>. E-mail: andre.abarbosa@ufpe.br.

⁴ É crítico, programador e pesquisador de cinema. Doutor em Comunicação e Cultura pela UFRJ, atualmente realiza estágio pós-doutoral na mesma instituição. ORCID: <https://orcid.org/0000-0002-3085-4838>. Lattes: <http://lattes.cnpq.br/2661025718263010>. E-mail: hermano.callou@gmail.com.

⁵ Ted Chiang, “Chatgpt is a Blurry jpeg of the Web”, *The New Yorker*, 9 de fevereiro de 2023.

Figura 2 – Imagem de Hito Steyerl da Stable Diffusion.



Fonte: Steyerl (2023, p. 83).

Então, como o Stable Diffusion foi do ponto A ao ponto B? Definitivamente, não é a justaposição de “antes e depois” mais lisonjeira; eu não recomendaria esse tratamento. Parece bastante malvado (*mean*), ou até mesmo depreciativo (*demeaning*) — mas esse é exatamente o ponto. A questão é: que maldade é essa (*what mean*)? Quem faz esse mal (*whose mean*)? E que mal, exatamente? O Stable Diffusion gera este retrato meu em uma situação de era do gelo, produzido por processos internos desconhecidos e relacionados de forma espúria aos dados de treinamento. Não é um algoritmo “caixa-preta” que tem que ser culpado, já que se conhece o código do Stable Diffusion. Em vez disso, poderíamos chamá-lo de um algoritmo caixa-branca, ou um filtro social. Essa é uma estimativa de como a sociedade, através de um filtro composto pelo lixo médio da internet, me enxerga. Basta remover o ruído da realidade das minhas fotos e extrair, em seu lugar, o sinal social; o resultado é uma “imagem média, malvada” (*mean*), uma renderização de médias correlacionadas — ou diferentes tons de medianidade.

A palavra inglesa “*mean*” tem vários significados; todos se aplicam aqui. “*Mean*” pode se referir a origens pobres ou humildes, à norma, à avareza ou à mesquinhez. Está relacionada ao sentido (*meaning*) como significação, com as ideias de comum, mas também com os meios (*means*) financeiros ou instrumentais. O próprio termo é um composto que mistura e sobrepõe camadas aparentemente incompatíveis de significação. Ele mistura valores morais, estatísticos, financeiros e estéticos, assim como posições de classe inferior e comum, em um único local indistintamente comprimido. Imagens médias estão longe de ser alucinações aleatórias; elas são produtos previsíveis do populismo de dados. Elas captam padrões sociais latentes que codificam significados conflitantes como coordenadas vetoriais e permitem visualizar atitudes sociais existentes, que alinham o comum com o status de classe inferior, a mediocridade e o comportamento mesquinho. São imagens residuais que persistem em telas e retinas muito tempo depois que sua fonte desapareceu. Elas realizam uma psicanálise sem *psique* nem análise para uma era de automação na qual a produção é aumentada pela fabricação em larga escala. Imagens

médias são sonhos sociais sem descanso, processando as funções irracionais da sociedade até suas conclusões lógicas. São expressões documentais das percepções da sociedade sobre si mesma, apreendidas através do sequestro e da captura caótica de dados em larga escala, e se apoiam em vastas infraestruturas de *hardware* poluentes e em trabalho subalterno precarizado, explorando conflitos políticos como recurso.

1 O PROBLEMA DE JANO

No outono de 2022, quando Dreamfusion, uma ferramenta de geração de imagens em 3D a partir de texto, foi colocada em teste, os usuários começaram a perceber uma falha interessante. Os modelos 3D gerados por aprendizado de máquina frequentemente tinham múltiplos rostos, que olhavam para direções diferentes (Figura 3). Essa falha foi apelidada de problema de Jano⁶. Qual era a sua causa? Uma resposta possível é que há uma ênfase excessiva em rostos no reconhecimento e análise de imagens por aprendizado de máquina; os dados de treinamento contêm relativamente mais rostos do que outras partes do corpo. As duas faces de Jano, deus romano dos começos e fins, olham para o passado e para o futuro; ele é, também, o deus da guerra e da paz, da transição de um estado social para outro.

Figura 3 – Modelo de um esquilo com três rostos em 3D gerado por aprendizado de máquina



Fonte: Steyerl (2023, p. 85).

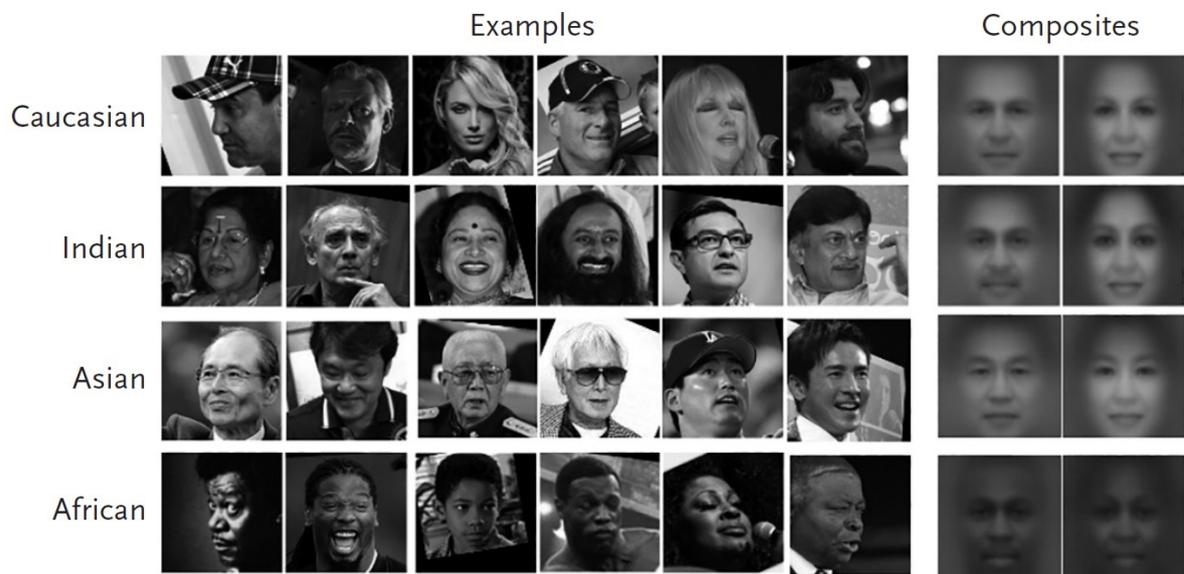
⁶ O problema de Jano foi identificado pela primeira vez por Ben Poole, um pesquisador do laboratório experimental Google Brain, e publicado em seu perfil no Twitter.

O problema de Jano experimentado no aprendizado de máquina toca em uma questão crucial: a relação entre o indivíduo e a multidão. Como retratar a multidão como um só indivíduo? Ou, inversamente, o indivíduo como multidão, como coletivo, grupo, classe ou Leviatã? Qual é a relação entre o indivíduo e o grupo, entre os interesses (e os patrimônios) privados e os comuns, especialmente em uma era em que representações estatísticas resultam de médias tiradas a partir de grupos?

2 PROBABILIDADES

Aqui (Figura 4) está outro compósito estatístico na qual meu rosto está emaranhado.

Figura 4 – Exemplos e retratos compostos de rostos racializados do *Wild Database*



Fonte: Steyerl (2023, p. 86).

Os borrões fantasmagóricos dotados de raça e gênero à direita poderiam ser denominados de foto vertical de grupo, na qual as pessoas não estão lado a lado, mas umas sobre as outras. Como surgiram essas imagens?

Em 2016, meu nome apareceu em um compêndio chamado ms-Celeb-1m, um banco de dados da Microsoft composto por 10 milhões de imagens de 100.000 pessoas encontradas na internet. A informação foi divulgada pelos artistas Adam Harvey e Jules Laplace como parte de seu projeto de pesquisa com bancos de dados, Megapixels⁷. Se o seu nome aparecesse nessa lista, a Microsoft incentivaria os pesquisadores a baixarem, da internet, fotos do seu rosto para construir um perfil biométrico. Eu havia me tornado parte de um dos conjuntos iniciais de dados usados para treinar algoritmos de reconhecimento facial. Mas quem os utiliza e para quê?

No fim das contas, o ms-Celeb-1m interessava a vários grupos e instituições. Foi usado,

⁷ Ver o site *Exposing.ai*. Uma pequena amostra dos nomes na lista incluía Ai Weiwei, Aram Bartholl, Astra Taylor, Bruce Schneier, Cory Doctorow, danah boyd, Edward Felten, Evgeny Morozov, Glenn Greenwald, Hito Steyerl, James Risen, Jeremy Scahill, Jill Magid, Jillian York, Jonathan Zittrain, Julie Brill, Kim Zetter, Laura Poitras, Luke DuBois, Michael Anti, Manal al-Sharif, Shoshanna Zuboff e Trevor Paglen. Como escrevem Harvey e Laplace, a definição de “celebridade” da Microsoft foi ampliada para jornalistas, ativistas e artistas, muitos dos quais eram “críticos vocais da própria tecnologia que a Microsoft está construindo usando seus nomes e informações biométricas”.

por exemplo, para otimizar formas de classificação racial pelos desenvolvedores de um outro conjunto de dados, chamado Racial Faces in the Wild. Eles lamentavam que as tecnologias de reconhecimento facial não funcionavam muito bem em pessoas não-brancas, então se dispuseram a “resolver” o problema. Carregaram imagens do conjunto de dados ms-Celeb-1m na interface de reconhecimento Face++ e usaram as pontuações raciais registradas para separar as pessoas em quatro grupos: caucasiano, asiático, indiano e africano. O objetivo declarado era reduzir a leitura enviesada dos programas de reconhecimento facial e diversificar os dados de treinamento⁸. Os resultados foram aparições fantasmagóricas de fenótipos racializados, ou uma ideia quase platônica da discriminação em si mesma.

Se essas renderizações fantasmagóricas evocam os famosos retratos compostos criados por Francis Galton na década de 1880, não é coincidência. Cientista social pioneiro, estatístico e eugenista, Galton desenvolveu um método de sobreposição fotográfica para criar retratos do que chamava de tipos, como “judeu”, “tuberculoso” e “criminoso”⁹. Eugenistas acreditavam no “aperfeiçoamento racial” e na “reprodução planejada”, alguns defendendo métodos como esterilização, segregação e até mesmo extermínio direto para livrar a sociedade de tipos que consideravam “inaptos”. Muitas vezes, os fantasmas acabaram convertidos em fotos policiais de grupos que, supostamente, deveriam desaparecer.

Muito já foi escrito sobre os antecedentes eugenistas dos pioneiros da estatística, incluindo, por exemplo, Ronald Fisher. Mas a estatística, como ciência, evoluiu desde então¹⁰. Como explica Justin Joque, os métodos estatísticos foram refinados ao longo do século XX para incluir mecanismos e parâmetros de mercado, como contratos, custos e benefícios, além de registrar os riscos econômicos de resultados falso-positivos ou falso-negativos. O resultado foi a integração da matemática de um cassino bem calibrado à ciência estatística¹¹. Utilizando dados, os métodos bayesianos puderam reverter o procedimento de Fisher de provar ou refutar a chamada hipótese nula. A nova abordagem funcionava ao contrário: partia-se dos dados para calcular a probabilidade de uma hipótese. Uma resposta específica podia ser retroativamente ajustada para corresponder à pergunta mais provável. Com o tempo, os métodos de cálculo de probabilidade foram otimizados para a lucratividade, adicionando mecanismos de mercado aos mecanismos selecionistas.

Renderizações estatísticas adicionam um efeito visual quasi-mágico a esse procedimento.

⁸ Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, Yaohai Huang, *Racial Faces in the Wild: Reducing Racial Bias by Information Maximization Adaptation Network*, artigo de pesquisa da Computer Vision Foundation, 2019.

⁹ Ver também o texto seminal de Allan Sekula sobre os retratos compostos de Galton, *The Body and the Archive*, *October*, vol. 39, inverno de 1986, p. 19.

¹⁰ Wendy Hui Kyun Chun, *Discriminating Data: Correlation, Neighbourhoods, and the New Politics of Recognition*, Boston, MA, 2021, p. 59; Lila Lee-Morrison, ‘Francis Galton and the Composite Portrait’, em *Portraits of Automated Facial Recognition*, Bielefeld, 2019, pp. 85–100. Ronald Fisher, uma figura fundadora da estatística, argumentou em *The Genetical Theory of Natural Selection* (1930) que as civilizações estavam em risco porque pessoas de “baixo valor genético” eram mais férteis do que pessoas com o chamado “alto valor genético”, e recomendou que limites fossem estabelecidos para a fertilidade das classes populares. Fisher criou o conhecido conjunto de dados *Iris*, publicando os resultados nos *Annals of Eugenics* em 1936, para provar que é possível discriminar entre diferentes espécies de flor de íris usando medições superficiais. As flores de Fisher eram representantes de uma ideia mais sinistra: se era possível discriminar diferentes espécies florais por meio de medições superficiais, também seria possível demonstrar a existência de diferentes raças pela medição de crânios. O conjunto de dados *Iris* ainda é ensinado a estudantes de Ciência da Computação como um exemplo fundacional.

¹¹ O método de Neyman-Pearson, por exemplo, incentiva a construção de duas hipóteses a serem selecionadas por um teste estatístico, dependendo das circunstâncias do experimento. Como um cassino lucrativo, “se os custos e riscos forem calculados adequadamente, a casa certamente perderá algumas rodadas, mas, com o tempo, acabará vencendo mais”. Justin Joque, *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism*, Londres e Nova York, 2022, pp. 124–125.

À medida que as categorias parecem emergir dos próprios dados, elas adquirem a autoridade de uma manifestação ou aparição imediata. Os dados não são mais apresentados através dos meios tradicionais, como gráficos, *clusters*, curvas, diagramas ou outras abstrações científicas. Em vez disso, eles são visualizados na forma da própria coisa da qual deveriam abstrair. Eles contornam a mediação para sugerir uma falsa imanência. Processos de abstração e alienação são substituídos por processos de retropropagação confusos — ou, de maneira mais simples: filtros sociais. Se Joque recorre ao conceito de abstração real de Sohn-Rethel para descrever os modos estatísticos de representação, as “imagens malvadas” (*mean*) poderiam ser descritas como “abstrações autenticistas”¹². Esse paradoxo reflete uma incoerência fundamental no cerne desse modo de representação. Embora essas renderizações se baseiem em médias correlacionadas, elas convergem para casos atípicos extremos e inatingíveis — ideais corporais anoréxicos, por exemplo. Um resultado irrealista e provavelmente insustentável é prescrito como norma: um modo de condicionamento social humano que antecede, há muito tempo, o aprendizado de máquina.

Nas “imagens médias” (*mean*), a estatística é integrada diretamente à aparência de um objeto, por meio da determinação de probabilidades. Se Galton realizou esse truque com rostos, as representações estatísticas expandem ainda mais seu método para o campo das ações, relações e objetos: o mundo em geral. A aparição aparentemente espontânea dessas distribuições disfarça as operações dentro das “camadas ocultas” das redes neurais, que distorcem as relações sociais existentes para convergir em direção a um caso “ótimo” altamente ideológico, utilizando todo tipo de ponderações e parâmetros baseados no mercado. Os mercados já eram vistos como computadores hipereficientes por von Mises e Hayek. De fato, dentro das mitologias econômicas liberais, os mercados desempenham o papel de inteligências artificiais gerais, ou IAGs — estruturas de nível superior, supostamente oniscientes, que não deveriam ser perturbadas ou reguladas. As redes neurais, portanto, são vistas como uma imitação da lógica de mercado, na qual a realidade está permanentemente em leilão.

É evidente essa integração da estatística no caso dos modelos 3D do Dreamfusion. A analogia estatística mais comum é a famosa moeda que, ao ser lançada, tem 50% de probabilidade de cair em cara ou coroa, caso seja justa e bem calibrada. Mas, no problema de Jano, a probabilidade de aparecer uma cara em vez de uma coroa é muito maior do que 50%. Na verdade, pode ser que não dê coroa de jeito nenhum. Os desenvolvedores apontam o problema básico de derivar renderizações 3D a partir de imagens 2D. Além disso, como mencionado acima, os dados podem estar enviesados, o algoritmo pode apresentar falhas ou algo pode estar faltando — ou o experimento em si e as ferramentas utilizadas podem nem ser justas nem bem calibradas. Seja qual for o caso, o Dreamfusion criou sua própria ramificação da teoria das probabilidades: em vez de cara ou coroa, a probabilidade é de cara e cara.

Como tudo isso se aplica aos retratos compostos de múltiplas cabeças do *Racial Faces in the Wild* com os quais me envolvi? Na lógica liberal da extração digital, a exploração e a desigualdade não são questionadas; elas são, no máximo, diversificadas. Nesse sentido, os autores do *RFW* tentaram reduzir o viés racial dos softwares de reconhecimento facial. Os resultados foram facilmente reaproveitados para identificar minorias com mais precisão por meio de algoritmos de visão computacional. Os departamentos de polícia estavam aguardando e torcendo para que o reconhecimento facial fosse otimizado para rostos não caucasianos. É exatamente isso que parece ter acontecido com a pesquisa surgida a partir do MS-Celeb-1M.

Uma empresa chamada SenseTime também percebeu isso¹³. A SenseTime é uma companhia de inteligência artificial que, até abril de 2019, fornecia software de vigilância

¹² Joque, *Revolutionary Mathematics*, p. 179.

¹³ Ver: *Megapixels*, conforme mencionado acima.

para autoridades chinesas, que o utilizavam para monitorar e rastrear uigures; a empresa fora apontada diversas vezes como potencialmente envolvida em casos de violação de direitos humanos¹⁴. Parece que a combinação do meu nome e rosto não foi apenas usada para otimizar a visão computacional para a classificação racial, mas essa otimização foi rapidamente colocada em prática para identificar e rastrear membros de uma minoria étnica na China. O simples fato da minha existência na internet foi suficiente para transformar meu rosto literalmente em uma ferramenta de discriminação, manejada por uma autocracia digital real e existente. A essa altura, a maioria dos rostos que apareceram na internet provavelmente já foi incluída em operações desse tipo.

3 OS MEIOS DE PRODUÇÃO MÉDIA

Há uma outra razão, mais pertinente, pela qual a suposta eliminação de vieses dentro dos conjuntos de dados cria mais problemas do que resolve. O processo limita as mudanças a partes do resultado, tornando-os mais aceitáveis para os consumidores ocidentais liberais, enquanto mantém intacta a estrutura da indústria e seus modos de produção. Mas o problema não está apenas na mídia-(social), e sim nos próprios meios de produção. Quem os possui? Quem são os produtores? Onde a produção acontece e como ela funciona?

Criar filtros para eliminar conteúdos prejudiciais e tendenciosos das redes é uma tarefa cada vez mais terceirizada para atores desfavorecidos, os chamados *microtrabalhadores* ou *trabalhadores fantasma*. Esses *microtrabalhadores* identificam e classificam materiais violentos, tendenciosos e ilegais dentro dos conjuntos de dados. Eles realizam essa tarefa na forma de *microtasks* mal remuneradas que transformam as redes digitais em linhas de montagem. Como relatado pela revista *Time* em janeiro de 2023, trabalhadores mal pagos no Quênia foram instruídos a alimentar uma rede “com exemplos classificados como de violência, discurso de ódio e abuso sexual”¹⁵. Esse detector agora é usado nos sistemas do ChatGPT da OpenAI. Nas metrópoles ocidentais, *microtrabalhadores* são frequentemente recrutados em comunidades excluídas do mercado de trabalho oficial devido a leis migratórias, como descreve esta entrevista anônima com um trabalhador digital em uma grande cidade alemã:

Trabalhador digital: Estávamos todos meio que na mesma situação, uma situação muito vulnerável. Éramos novos na cidade e no país, tentando nos integrar, e precisávamos desesperadamente de um emprego. Todo o pessoal no meu andar tinha pelo menos um mestrado, eu não era o único. Uma das minhas colegas era bióloga, especializada em pesquisa com borboletas, e teve que trabalhar nas mesmas tarefas que eu. Como era muito difícil encontrar um emprego de verdade, relativo à sua especialização, as pessoas simplesmente aceitavam esse tipo de trabalho de meio período. Eram pessoas altamente qualificadas, com diferentes origens linguísticas.

Entrevistador: Eram todos estrangeiros?

Trabalhador digital: Eram.

Entrevistador: Como era o trabalho?

Trabalhador digital: Terrível. E era assim para todos que conheci lá. Durante o treinamento, te informam que você vai ver pedofilia, violência gráfica, linguagem sexualmente explícita. E então, quando você realmente começa a trabalhar, você senta na

¹⁴ Anna Swanson e Paul Mozur, *US Blacklists 28 Chinese Entities over Abuses in Xinjiang*, *New York Times*, 17 de outubro de 2019.

¹⁵ Billy Perrigo, *Openai Used Kenyan Workers on Less Than \$2 Per Hour to Make Chatgpt Less Toxic*, *Time*, 18 de janeiro, 2023.

sua escrivanhinha e vê coisas inacreditáveis. Isso é mesmo real? Os efeitos de longo prazo desse trabalho são bem nocivos. Não havia ninguém no meu grupo que não tivesse tido problemas depois. Problemas como distúrbios do sono, perda de apetite, fobias, fobia social. Alguns até tiveram que fazer terapia. No primeiro mês, você passa por um treinamento muito, muito intenso. Tínhamos que aprender a reconhecer conteúdo que era extremo demais. Porque a IA ou os mecanismos de aprendizado de máquina não conseguiam decidir nos casos delicados. A máquina não tem sentimentos, não era precisa o suficiente.

Eu fiquei deprimido. Tive que fazer terapia. Fui medicado. Quando comecei lá no trabalho, minha tarefa principal era filtrar postagens com conteúdo sexualmente explícito e casos de “alta prioridade”, que geralmente tinham a ver com suicídio ou automutilação. Havia muitas fotos de ferimentos por corte. Eu tinha que analisar quais eram de automutilação e quais eram de suicídio. No segundo mês, pedi ao líder da minha equipe para me colocar em outro fluxo de conteúdo porque eu estava me sentindo mal.

Havia muitas regras no posto de trabalho: todo mundo sem telefone, sem relógio, nada com o que pudéssemos tirar fotos. Sem papel, sem canetas, nada com o que fazer anotações. Algumas vezes, vimos *drones* voando do lado de fora das janelas. Supostamente, espiões estavam tentando filmar o que estava acontecendo na empresa. Todos recebiam instruções para fechar as cortinas quando isso acontecia. Uma vez, um jornalista estava do lado de fora do prédio. Fomos orientados a não sair de lá e a não falar com ele. Para a empresa, os jornalistas eram como inimigos.

Entrevistador: Que tipo de IA se usava no seu trabalho?

Trabalhador digital: Eu não sei muito sobre a IA que era usada onde eu trabalhava.

Acho que eles tentavam esconder isso de alguma forma. O que sabíamos era que havia algum tipo de aprendizado de máquina acontecendo. Basicamente, eles queriam substituir os humanos pelo software de IA. Lembro que, em certo momento, tentaram usar esse software, mas ele era muito impreciso. Então, pararam de usá-lo.

Entrevistador: Que tipo de software era esse, qual era exatamente sua tarefa?

Trabalhador digital: Não tenho ideia. Esse tipo de conhecimento só era passado para o nível executivo da empresa. Eles mantinham essa informação em segredo. E, mesmo que ouvíssemos rumores aqui e ali, eles escondiam esse projeto de nós. Mas a IA é um fracasso, porque o algoritmo não é preciso o bastante. As pessoas falam sobre inteligência artificial, mas eu diria que a tecnologia por trás disso ainda é muito, muito básica. Por isso eles precisam de nós, humanos. As máquinas não têm sentimentos. Elas não conseguem sentir. O objetivo principal era fazer pessoas como eu trabalharem como robôs¹⁶.

Outra entrevista realizada como parte do mesmo projeto descreveu como, na Alemanha, trabalhadores digitais sírios tiveram que filtrar e revisar imagens de suas próprias cidades natais destruídas pelo recente terremoto na região — e, em alguns casos, dos escombros de suas antigas casas¹⁷. Essas imagens foram consideradas violentas demais para os consumidores das redes sociais, mas não para os habitantes da própria região, que haviam sido expulsos pela guerra e pela destruição e foram forçados a se tornar trabalhadores fantasmas no exílio. Convenientemente, a violência militar proporcionou às corporações digitais localizadas na Alemanha uma nova força de trabalho refugiada, extremamente explorável.

Ajustar a tecnologia para ser mais “inclusiva” pode, assim, levar a um melhor mecanismo

¹⁶ *Die Zeit*, Feuilleton Ausgabe no. 14, 30 de março de 2023 — um projeto coletivo de suplemento cultural sobre inteligência artificial de Marie Serah Ebcinoglu, Eike Kühn, Jan Lichte, Peter Neumann, Hanno Rauterberg, Malin Schulz, Hito Steyerl e Tobias Timm. Esta entrevista foi realizada por Tobias Timm.

¹⁷ Esta entrevista foi realizada por Marie Serah Ebcinoglu.

de identificação de minorias enquanto terceiriza trabalho traumático e mal remunerado. Isso pode otimizar a discriminação, higienizando superficialmente aplicações comerciais enquanto cria no processo hierarquias de classe abertamente exploratórias. Conflitos políticos e militares, assim como barreiras migratórias motivadas pelo racismo, são ferramentas importantes na criação dessa força de trabalho marginalizada. Talvez o enviesamento não seja uma falha, mas uma característica fundamental de um sistema de produção de médias (*mean*). O viés é produtivo não apenas no nível da representação, ao degradar visualmente as pessoas. Sua suposta eliminação é igualmente produtiva, ao ajudar a consolidar hierarquias de classe mantidas por guerras, conflitos energéticos e sistemas de fronteira racistas, podendo ser explorada dentro de um sistema de produção malvado (*mean*).

A eliminação do enviesamento não é a única tarefa dos microtrabalhadores. Eles também classificam fotografias de vias urbanas para carros autônomos e categorizam imagens de objetos e pessoas a fim de ajudar redes de aprendizado de máquina a distinguir uns dos outros. Como muitos autores já observaram, os trabalhadores fantasma humanos são o motor da automação: os carros autônomos não poderiam funcionar sem eles. A automação avança a partir das microdecisões médias de destacamentos inteiros de humanos sub-remunerados, e não a partir de um computador superinteligente. Em alguns casos, isso leva a que pessoas se façam passar por inteligências artificiais, mesmo onde não existe nenhum uso de aprendizado de máquina. Um pesquisador relata o seguinte:

Entrevistamos K., um empreendedor parisiense e fundador de uma *start-up* que acusava seus concorrentes de dizer que trabalhavam com IA, enquanto, na verdade, terceirizavam todo o trabalho para humanos recrutados no exterior por meio de plataformas. Ele chegou a afirmar que “Madagascar lidera a inteligência artificial na França”. Ainda mais indignado estava S., um estudante que fez estágio em uma *start-up* de IA que oferecia recomendações personalizadas de viagens de luxo para pessoas de alta renda. A estratégia de comunicação da empresa enfatizava a automação, com um sistema de recomendação supostamente baseado nas preferências dos usuários extraídas de redes sociais. Mas, nos bastidores, todos os processos eram terceirizados para microprestadores de serviços em Madagascar. A empresa não realizava nenhum aprendizado de máquina, e o estagiário não conseguiu adquirir as habilidades de alta tecnologia com as quais sonhava¹⁸.

As camadas ocultas das redes neurais também escondem a realidade do trabalho humano, assim como o absurdo das atividades realizadas. A magia aparentemente não mediada das imagens que surgem automaticamente de um amontoado de dados é baseada em uma exploração e expropriação massivas no nível da produção. Talvez as emanções aparentemente fantasmagóricas de rostos em renderizações estatísticas sejam, na verdade, retratos dos microtrabalhadores ocultos, assombrando e permeando as imagens médias (*mean*).

O trabalho oculto também é crucial para os conjuntos de dados usados para treinar geradores de *prompts*. Os 5,8 bilhões de imagens e legendas retiradas da Internet e compiladas no LAION-5B, o conjunto de dados de código aberto utilizado para treinar o modelo *Stable Diffusion*, são produto do trabalho humano não remunerado, “desde programadores e *designers* de páginas da Internet até usuários que carregam e compartilham imagens nessas páginas”¹⁹. Não é preciso dizer que nenhuma dessas pessoas recebeu remuneração, ou participação comercial nos conjuntos de dados, ou nos produtos e modelos construídos a partir deles. Os direitos de

¹⁸ Paolo Tubaro, Antonio A. Casilli, and Marion Coville, “The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence”, *Big Data & Society*, vol. 7, no. 1, junho de 2020, p. 7.

¹⁹ Chloe Xiang, “ai Isn’t Artificial or Intelligent”, *Vice Motherboard*, 6 de dezembro de 2022.

propriedade privada, no capitalismo digital e fora dele, só são pertinentes quando se referem aos proprietários ricos. Todos os demais podem ser roubados de forma rotineira.

4 DA MÉDIA AO COMUM

Agora fica mais claro por que os modelos 3D com cara de Jano mostram muito mais caras que coroas: a “moeda” está viciada. Seja saindo cara ou “cara”, na automação ou — na expressão de Astra Taylor — *fauxtovation*, a casa sempre vence. Mas a questão das condições de trabalho pode permitir uma observação mais geral sobre a relação entre estatísticas e realidade, ou sobre a questão da correlação e da causalidade. Muitos autores, incluindo eu mesma, interpretaram a mudança de uma ciência baseada em causalidade para suposições baseadas em correlação de dados como um exemplo de pensamento mágico ou um retrocesso para a alquimia. Mas e se esse retrocesso capturar também um aspecto importante da realidade? Uma realidade que, em vez de ser governada pela lógica ou pela causalidade, está, na verdade, se estruturando muito mais como um cassino?

Um *post* de *blog* sobre depressão e videogames fornece um exemplo esclarecedor. O autor descreve sua atividade de jogar videogames em momentos de depressão, desfrutando em especial as tarefas pequenas e repetitivas que levam a algum tipo de resultado construtivo — uma colheita sendo plantada, uma casa sendo construída:

O que torna o trabalho nos videogames tão empolgante é a chance de desfrutar integralmente da recompensa pelos seus esforços. Você recebe de volta o que você dedicou de maneira muito direta. Os jogos que oferecem maior satisfação são, na verdade, simulações fantasiosas nas quais você pode viver segundo um preceito marxista básico: o trabalho tem direito a tudo o que cria²⁰.

O autor descreve uma relação causal entre *input* e *output*, trabalho e recompensa. A conclusão surpreendente é que tal causalidade dentro do capitalismo realmente existente é rara, especialmente no trabalho precarizado. Por mais esforço que se faça, não se vai obter de forma linear a recompensa adequada — um salário digno ou uma forma justa de compensação. Formas precárias e altamente especulativas de trabalho não geram retornos lineares; causa e efeito estão desconectados. Isso também introduz um aspecto de classe na distribuição de causalidade *versus* correlação na vida real. Empregos pagos por hora mantêm um grau maior de causa e efeito do que aqueles que ocorrem no reino desregulamentado do acaso, em ambas as extremidades da escala salarial. Assim, é mais “racional” para muitas pessoas tratar a existência cotidiana como um cassino e esperar por ganhos especulativos. Se tudo o que você pode esperar dentro de um paradigma causal é renda zero, então comprar um bilhete de loteria se torna uma opção eminentemente racional. O trabalho começa a se assemelhar aos jogos de azar. Phil Jones descreve o *microtrabalho* de forma semelhante:

O trabalhador, assim, opera cada vez mais em uma economia quase mágica de aposta e loteria. O *microtrabalho* representa o ápice sombrio dessa trajetória, onde a possibilidade de que a próxima tarefa seja paga seduz os trabalhadores a retornarem repetidamente por mais. Sistemas complexos de recompensa e preços negociáveis *gamificam* tarefas e criam efetivamente uma nova embalagem para a superfluidade e a precariedade como formas de trabalho-como-lazer novas e emocionantes²¹.

²⁰ Tabitha Arnold, “Depression and Videogames”, *Labour Intensive Art Substack*, 6 de janeiro de 2023.

²¹ Phil Jones, *Work Without the Worker: Labour in the Age of Platform Capitalism*, Londres e Nova York 2021, p. 50ff.

Quando o “salário se transforma em uma aposta”, a probabilidade deixa de ser apenas uma estimativa de um resultado real para se tornar parte do próprio resultado²². As representações estatísticas levam isso em conta. Quando a causalidade social é substituída parcialmente por correlação, as relações de trabalho retrocedem para os tempos da exploração do trabalho pela indústria vitoriana, as imagens se transformam em apostas e a fábrica se torna uma casa de jogo. A fotografia indexical era, pelo menos em parte, baseada em uma relação de causa e efeito. Mas, nas representações estatísticas, a causalidade fica à deriva em meio a um emaranhado de processos quase não lineares, que não são contingentes, mas sim manipulados de maneira opaca.

O trabalho precarizado na indústria de aprendizado de máquina, e os processos repetitivos de condicionamento e treinamento que ele exige, levanta a questão: quem ou o que está sendo treinado? Claramente, não são apenas as máquinas, ou mais exatamente as redes neurais. As pessoas também estão sendo treinadas — como microtrabalhadores, mas também como usuários em geral. Voltando à excelente questão de Herndon e Dryhurst: “eu fui treinada?” A resposta é: sim. Não apenas minhas imagens, mas eu, eu mesma. Geradores de imagens baseados em *prompt*, como o DALL-E, de fato, dependem do treinamento de modelos de aprendizado de máquina. Mas, de maneira muito mais importante, eles treinam os usuários sobre como usá-los e, ao fazer isso, os integram em novos fluxos de produção, em conjuntos de software e hardware alinhados a aplicações de aprendizado de máquina patenteadas.

Eles normalizam um ambiente de produção sitiado no qual os usuários têm que alugar constantemente algum sistema de nuvem, não apenas para serem capazes de realizar suas tarefas, mas até mesmo para acessar as ferramentas e resultados de seu próprio trabalho. Um exemplo disso é o Azure, segundo a Microsoft, a “única nuvem pública global que oferece supercomputadores de IA com capacidades massivas de crescimento vertical e horizontal”. O Azure cobra pelo aluguel de capacidade de processamento e de aplicações computacionais de aprendizado de máquina, enquanto a Microsoft estabeleceu uma estrutura de propriedade vertical que inclui *hardware* e *software* preparados para aprendizado de máquina, navegadores, modelos, interfaces de programação de aplicativos (APIs), etc.. A Adobe, o mais odiado, complexo e extrativista quasi-monopólio mundial para trabalhadores de imagem, avança rapidamente na mesma direção. Dwayne Monroe chama esse tipo de quasi-monopólio de “estrutura super-rentista”, na qual as corporações digitais privatizam os dados dos usuários e vendem os produtos de volta para eles mesmos: “a indústria de tecnologia sequestra uma variedade de bens comuns e depois nos cobra pelo que deveria ser de livre acesso”²³.

Muitos trabalhadores digitais e administrativos estão ameaçados pela automação baseada em aprendizado de máquina (ML - Machine Learning); entre eles, sem nenhuma ordem específica, programadores, profissionais de relações públicas, *designers* de sites e contadores. Mas é mais provável que muitos deles sejam forçados a “atualizar” seus serviços pagando para usar plataformas construídas com base no trabalho roubado deles mesmos, para continuar “competitivos”, do que sejam totalmente substituídos pela automação mediante aprendizado de máquina. Hoje, formar esses profissionais significa acostumá-los com o desaparecimento de suas funções no médio prazo e condicioná-los à dependência de serviços monopolizados para poder continuar trabalhando e acessando os resultados de seu próprio trabalho.

Isso chama a atenção para outro aspecto das “imagens malvadas” (*mean*). Assim como os NFTs, as representações estatísticas são ferramentas de integração a ambientes tecnológicos específicos. No caso dos NFTs, trata-se de um ambiente criptografado gerido por

²² Ibid, p. 50.

²³ Dwayne Monroe, “Chatgpt: Super Rentier”, *Computational Impacts*, 20 de janeiro de 2023.

meio de ferramentas como carteiras digitais, transações ou registros. No caso do aprendizado de máquina, a infraestrutura é composta por arquiteturas de nuvem massivas, devoradoras de energia e verticalizadas, baseadas nos cliques mal pagos realizados por pessoas em regiões em conflito, ou por refugiados e imigrantes nos centros metropolitanos. Os usuários estão sendo integrados a um sistema gigantesco de extração e exploração, que cria uma pegada de carbono imensa.

Levar o problema de Jano a sério, portanto, significaria nos descondicionar de um sistema de extorsão e extração múltipla. Um primeiro passo seria ativar o outro aspecto da cabeça de Jano, aquele que olha para as transições, para os finais como novos começos, em vez de olhar para um passado feito de dados roubados. Por que não mudar a perspectiva para outro futuro — para um era de pequenas tecnologias resilientes, usando configurações viáveis mínimas, alimentadas por energia renovável, que não exijam roubo, exploração e regimes monopolistas sobre os meios digitais de produção? Isso significaria descondicionar a nossa visão de um futuro dominado por algum tipo de esquema pirâmide digital-oligárquico, sustentado pelo esforço de microtrabalhadores ocultos, no qual as relações de causa e efeito são substituídas por correlações manipuladas. Se uma das faces de Jano olha para a média, a outra se volta para o comum.

Referências

STEYERL, H. Main Images. *New Left Review*, [S. l.], n. 140-141, p. 82-97, 2023. Disponível em: <https://newleftreview.org/issues/ii140/articles/hito-steyerl-mean-images>. Acesso em: 16 jun. 2025.

Submissão: 02/03/2025

Aprovação: 30/06/2025

