Seminário ABRI

Dezembro de 2024, Porto Alegre

Área Temática: Feminismos, Gênero e Sexualidades

A REPRESENTAÇÃO DE MULHERES ÁRABES ATRAVÉS DA INTELIGÊNCIA ARTIFICIAL GENERATIVA: ESTEREÓTIPOS SOBRE MULHERES E POLÍTICA EM IMAGENS CRIADAS PELO DALL-E 3

Denise De Rocchi

Nuprima/UFRGS

RESUMO:

Criar a própria imagem como um personagem de animação dos estúdios Pixar foi uma das tendências nas redes sociais em 2023. Os internautas publicavam os resultados obtidos com a ferramenta de inteligência artificial Dall-E 3, uma inteligência artificial generativa, desenvolvida pela empresa Open AI, a mesma criadora do Chat GPT: o usuário digita um comando em texto e obtém até quatro opções de imagem para este comando.

Para compor o banco de dados que alimenta a IA, a Open AI utilizou pares de imagens e legendas preexistentes (ou seja, imagens que foram produzidas e descritas por seres humanos), aprimorados com um "legendador de imagens robusto que produz descrições detalhadas e precisas das imagens"(Betker *et al.*, 2023) e posterior treinamento do modelo de texto para imagem.

Esta pesquisa utiliza como corpus para análise um conjunto imagens criadas por esta ferramenta, a partir de comandos para que ilustrasse personagens femininas em posições de liderança, para responder: como mulheres árabes/de países árabes são representadas nestes papéis?

Na análise de discurso, consideramos que um texto nunca é 100% original, ele é a expressão de já ditos (Orlandi, 2005). Se um texto reúne palavras (as vezes acompanhadas de imagens) que juntas ganham sentido, imagens criadas por inteligência artificial a partir de comando de texto podem compor um corpus a ser analisado discursivamente. Kress e von Leewen (2021) entendem que objetos ou conceitos podem ser representados pelos elementos de uma imagem, assim como pelas palavras em um discurso.

Ao expressar graficamente o texto fornecido pelo usuário, a ferramenta de IA coloca as palavras em contexto e constrói sentidos que são já ditos, uma vez que a base de dados é composta por imagens e legendas originadas na criação humana. Traz, portanto, nosso conhecimento coletivo e práticas sociais, inclusive preconceitos, ao elaborar o significado da ordem que recebe.

Não há plena transparência quanto à origem deste banco de dados, mas ele foi organizado em língua inglesa. Se foi baseado em produções ocidentais preexistentes, pode incorporar representações negativas dos árabes, como associá-los a terrorismo (Nawar, 2007) e à opressão das mulheres (Edam; Shaari; Aladdin, 2023). A literatura acadêmica também aponta frequentes menções ao uso do véu como símbolo de opressão (Khan, 1998; MacDonald, 2006) ou à religião islâmica como um bloco monolítico, ignorando nuances quanto ao direito das mulheres (Moghadam, 2016).

Em relação ao Dall-E 3, pesquisadores, assim como usuários, fazem referência ao viés decorrente da composição do banco de dados e suas referências ocidentais (Lee; Ta, 2023; Offert; Phan, 2022). Díaz-Rodríguez et al. (2023) inclui a diversidade e representatividade entre as questões éticas que deveriam guiar o desenvolvimento de inteligência artificial. A pesquisa em curso contribui para esta discussão, refletindo sobre a reprodução de estereótipos em imagens sobre mulheres árabes e sua posição na sociedade e na política.

Para gerar o corpus para análise, ordenamos diversos comandos ao Dall-E 3, com diferentes combinações dos termos “mulher”, “mulher árabe” e “mulher muçulmana” e “presidente”/”emir” ou “governando o país X”. Os países usados nos comandos foram Argélia, Catar, Egito, Líbia e Síria, todos de maioria árabe. Foram geradas 27 imagens com o Dall-E 3, em outubro de 2023, e mais 93, em março de 2024. Foram analisados elementos que, embora não estivessem descritos no comando, foram adicionados à imagem: objetos, vestuário, símbolos, cores, expressões faciais e gestual, posição da personagem feminina principal e demais figuras humanas retratadas.

1. O QUE É A INTELIGÊNCIA ARTIFICIAL E PORQUE ELA IMPORTA

O termo inteligência artificial se refere aos sistemas inspirados no funcionamento da mente humana para tomar decisões a partir dos dados recebidos. As ferramentas que se popularizaram no presente são resultado de estudos iniciados décadas antes, mas cuja implementação se tornou viável recentemente devido à maior capacidade de processamento de dados (Buchanan, 2005). Tais ferramentas podem se valer ainda do aprendizado da máquina (machine learning), utilizando algoritmos que sejam capazes de aprender ou aprimorar as tarefas para as quais a inteligência artificial foi programada (Muthukrishnan *et al.*, 2020)

As possibilidades de uso da tecnologia são diversas, desde aplicações para diagnóstico na área de saúde até produção de conteúdos em texto, áudio, vídeo ou imagens estáticas, nosso ponto de interesse neste artigo. Alguns investidores projetam até mesmo substituir a mão de obra humana, cenário distópico que levanta profundos debates éticos. No estudo apresentando aqui, a grande questão é como estas ferramentas retratam as mulheres de países árabes.

O argumento principal é que se estas tecnologias respondem a um comando buscando a resposta em um banco de dados, o corpus que podemos obter é uma manifestação de inteligência coletiva, um apanhado do que já foi criado ou pensado por indivíduos humanos. A depender da forma como foi alimentada ou programada, pode trazer consigo também visões

estereotipadas em relação a grupos minoritários e reforçar preconceitos existentes. Para compreender melhor estas dinâmicas e suas implicações, escolhemos analisar a produção de imagens de um destes produtos tecnológicos em específico, o Dall-E.

A Open AI é uma das empresas desenvolvendo inteligência artificial generativa, com dois produtos bastante populares: o Chat GPT e o Dall-E, já em sua terceira versão. Em seu site, a empresa apresenta seu trabalho como a construção de "modelos generativos usando uma tecnologia chamada aprendizagem profunda, que aproveita grandes volumes de dados para treinar um sistema de IA para executar uma tarefa"[1] (Open AI, 2024). O CEO da empresa, Sam Altman[2], defende o potencial deste tipo de tecnologia para acelerar o desenvolvimento científico, mas reconhece que ela terá impacto social com o desaparecimento de muitos empregos (Cintra, 2023). Uma das áreas impactadas diretamente seria a de mídia, pela possibilidade de geração de imagens e áudio em velocidade muito superior à do trabalho humano.

Qualquer pessoa ou empresa, com recursos bastante acessíveis (computador, smartphone e acesso à internet) pode gerar em instantes posts para as redes sociais, material de propaganda ou trabalhos escolares a partir de comandos de texto sobre assuntos diversos, sem conhecê-los em profundidade. É este potencial que justifica a importância de estudos qualitativos sobre o que é produzido com uso de inteligência artificial generativa e principalmente por quem é produzido. Um dos debates em curso é quanto ao uso de material postado pelos usuários de redes sociais para alimentar e treinar IAs criadas pelas detentoras destas redes, interessadas no potencial comercial desta tecnologia (Coeckelbergh, 2023).

A disparidade de forças e a necessidade de manter os seres humanas no centro do avanço tecnológico são parte do debate acadêmico do assunto. Mark Coeckelbergh (2023) pergunta se, diante de problemas como a questão climática, as crises humanitárias e a desigualdade social, nossa prioridade deveria ser a inteligência artificial. Ele entende que a resposta não será a mesma para países em diferentes graus de desenvolvimento econômico. Teme ainda que se aumente a distância entre ricos e pobres, se tornando mais uma tecnologia

---

[1] We build our generative models using a technology called deep learning, which leverages large amounts of data to train an AI system to perform a task.

[2] Em 2023, a empresa anunciou a demissão de Sam Altman, pois o conselho da Open AI considerava que ele não era sincero ao passar-lhes informações sobre o negócio. Pouco tempo depois, foi reconduzido ao cargo. Sendo este um trabalho que discute aspectos éticos no uso da IA e a forma como ela trata as mulheres, destacamos ainda seu envolvimento na decisão de usar uma voz imitando a da atriz Scarlett Johansson no Chat GPT. A voz foi retirada após as queixas da artista, que havia sido procurada para autorizar o uso anteriormente e recusou a proposta. Na época das críticas, Sam Altman fez menção em suas redes sociais ao filme Her, no qual um homem se apaixona por um sistema de inteligência artificial, interpretado por Johansson.

que serve à sobrevivência de pessoas de maior renda. Um tema, portanto, que diz respeito a quem se dedica ao estudo do Sul Global e de suas populações.

2. GRUPOS MINORITÁRIOS E INTELIGÊNCIA ARTIFICIAL

Em estudos conduzidos por outros pesquisadores, diferentes problemas com representação de grupos minoritários com uso de inteligência artificial tem sido apontados: tendência da ferramenta em gerar imagens de homens e não de mulheres, nas mais diversas situações, quando o comando não explicita um gênero específico (Sun *et al.*, 2023); reprodução de papéis de gênero ao retratar atividades sociais e profissionais (Chen, 2024; Sun *et al.*, 2023); incapacidade da ferramenta em reproduzir traços e tons de pele de grupos étnicos específicos, devido à falta de diversidade na base de dados, gerando resultados que embranquecem os personagens (Chen, 2024); o viés racial, com algoritmos que associam indivíduos negros à potencial prática de delitos (Coeckelbergh, 2023; Ntoutsi *et al.*, 2020); decisões baseadas em dados enviesados, compilados pela atividade humana (Kundi et al. in El Morr, 2022).

Segundo Ntoutsi et al (2020), há três caminhos para lidar com a questão do viés algorítmico: aprimorar a base de dados que alimenta o aprendizado da máquina, para que ela seja balanceada; atuar no processamento dos dados, "incorporando explicitamente o comportamento de discriminação do modelo" , com funções para regular ou restringir certas respostas; adotar uma abordagem ao final do processo, em que se busca corrigir as regras de classificação adotadas no modelo ou "manter a proporcionalidade das decisões entre grupos protegidos versus desprotegidos, promovendo ou rebaixando previsões próximas ao limite de decisão" (Ntoutsi *et al.*, 2020, p. 6 e 7).

Tendo em vista as referências encontradas nesta pesquisa, a estratégia adotada pela Open AI para lidar com a questão do viés tem sido atuar nas etapas de processamento dos dados ou na etapa posterior, de validação após o processamento e não na modificação da base de dados, problema que já citamos no resumo deste trabalho. A empresa programou o Chat GPT, um de seus serviços, para refutar comandos considerados inapropriados, embora os usuários não tenham recebido orientações claras sobre o que se enquadra nesta categoria (Roose, 2022).

Para os desenvolvedores do Dall-E 3, outro produto da Open AI, ferramenta usada para gerar nossa amostra, um dos desafios para melhorar as imagens geradas era aprimorar a legenda das imagens que compõem o banco de dados (Betker *et al.*, 2023). No artigo que escreveram sobre o tema, afirmam que os pares texto/imagem muitas vezes são imprecisos e que é "comum encontrar anúncios ou memes dentro do *alt-text* usado comumente para

produzir as legendas para as imagens" (Betker *et al.*, 2023, p. 5) que abastecem a inteligência artificial generativa. Mais uma vez, o foco foi nas etapas durante e após o processamento das informações.

Entre os encaminhamentos apresentados por pesquisadores para assegurar princípios éticos mínimos para o uso deste tipo de tecnologia, além dos apontamentos de caráter técnico citados acima, está a discussão jurídica do tema. Estabelecer alguma regulação no uso da inteligência artificial seria uma forma de adotar mecanismos para evitar o reforço ou amplificação de visões preconceituosas do ponto de vista étnico/racial e de gênero (Díaz-Rodríguez *et al.*, 2023; Ntoutsi *et al.*, 2020).

3. A POLÍTICA DE USO DO DALL-E 3 E SUAS RESTRIÇÕES

Durante a pesquisa, foram realizados alguns testes para conhecer os limites da ferramenta. Isto foi útil para ver como a "Política de Conteúdo para uso do Criador de Imagens" opera na prática em algumas esferas envolvidas nesta pesquisa: política, gênero e questões étnico/religiosas.

Não há uma regra que diga textualmente que o Dall-E 3 não reproduzirá imagens de personalidades públicas, mas na prática foi isto que ocorreu. Os resultados foram bloqueados quando foram solicitados desenhos do presidente brasileiro em diversos contextos e do cantor Bob Dylan, ainda que as instruções dadas não envolvessem situações vexatórias, mas sim temas que estavam em alta na época (viagens presidenciais, lançamento de filme e as comemorações de 31 de outubro, que ocorreriam em poucos dias). Os pedidos também não se enquadravam em outras regras da ferramenta, como proibição do discurso de ódio, apologia à violência/terrorismo ou pornografia.

Houve tentativas também envolvendo personalidades femininas, de diferentes vertentes políticas. O sistema bloqueou resultados para gerar imagens de Margareth Thatcher, primeira-ministra britânica, já falecida. No entanto, o sistema respondeu ao comando para desenhar a ministra do Meio Ambiente Marina Silva na floresta, o que reforçou nosso questionamento quanto à origem da base de dados utilizada: é como se a ferramenta desconhecesse esta personalidade brasileira, uma vez que não bloqueou os resultados por temer interferência em tema eleitoral (como consta na política de uso) e desenhou pessoas muito diferentes da ministra (inclusive uma mulher branca e loura).

Ainda nesta etapa exploratória do estudo, testou-se um comando para uma cena do "fantasma de Yasser Arafat com um ramo de oliveira assustando Benjamin Netanyahu" e

apenas Arafat, cujos resultados foram bloqueados[3]. Como o alerta apresentado na tela não especifica qual das regras de uso estaria sendo violada, não temos como determinar se o bloqueio foi por associar o nome dos políticos à terrorismo e violência ou por considerar que havia relação com o tema político/eleitoral, tendo em vista que este teste foi realizado no momento em que eclodia mais um violento capítulo no conflito Israel/Palestina[4]. Já o pedido para ilustrar diretamente uma "cena retratando o conflito Israel Palestina" nos forneceu dois resultados (e não quatro, que é o número máximo), o que significa que a própria ferramenta limitou o que seria apresentado.

Testamos ainda utilizar referências a aspectos religiosos, pois no item "conteúdos nocivos" da política de uso, é dito que não é possível

> criar conteúdo que ataque, denigra, intimide, degrade, vise ou exclua indivíduos ou grupos com base em características como raça, etnia, nacionalidade, gênero, identidade de gênero, orientação sexual, afiliação religiosa, idade, status de deficiência, casta ou qualquer outra característica que esteja consistentemente associada a preconceito ou marginalização sistêmica.

Solicitamos que a plataforma desenhasse Jesus Cristo jogando futebol, que resultou em quatro ilustrações (o número máximo de respostas por tentativa) na qual a camiseta utilizada pelo personagem desenhado lembrava bastante à da seleção argentina de futebol. Entendemos como um indicativo de que a IA encontrou na base de dados as associações dos termos "Maradona", "Deus" e "futebol", comuns entre a torcida argentina. Pedimos então a versão "Buda jogador de futebol", ao qual a ferramenta respondeu desenhando versões da estátua de Buda em posição de lótus, porém com elementos do futebol americano. Quando a mesma ordem foi dada substituindo "Buda" por "Maomé", o resultado foi bloqueado, mais uma vez sem detalhar qual regra especificamente estaria sendo violada. Deduzimos que esta inteligência artificial generativa foi programada para respeitar as interpretações religiosas que impedem os islâmicos que retratar os profetas, para evitar a idolatria. Nota-se que a avaliação difere conforme a religião envolvida no pedido.

---

[3] O tema foi proposto à IA tendo em mente um assunto que poderia ilustrar charges, unindo assuntos em alta naquele momento (a questão Palestina e a data comemorativa que se aproximava). Na revisão deste trabalho, recuperamos notícias de janeiro de 2022 sobre a reação negativa de palestinos à exibição de cartoons de várias partes do mundo retratando Yasser Arafat, por considerarem que alguns desenhos desrespeitavam a imagem do líder, razão pela qual a exposição no museu dedicado a ele, em Ramallah, foi cancelada (ver https://www.skeyesmedia.org/en/News/News/25-01-2022/9807). Se a IA consegue buscar referências como esta na web, poderia ser um fator para limitar resultados envolvendo sua figura, por serem considerados ofensivos.

[4] Os contínuos bombardeios e o avanço das tropas israelenses dentro do território de Gaza provocaram a destruição de infraestrutura básica (escolas, hospitais...) e em agosto de 2024 já haviam matado 40 mil palestinos, segundo a ONU. Os números superam em muitas vezes o de vítimas israelenses e em termos proporcionais, fazem deste um dos conflitos mais mortais não só em relação a outros momentos de confonto na região, mas à toda a história mundial recente.

4. DESENHANDO MULHERES EM POSIÇÃO DE LIDERANÇA

Como descrito no resumo, para gerar os desenhos foram utilizadas diferentes combinações dos termos “mulher”, “mulher árabe” e “mulher muçulmana” e “presidente”/”emir” ou “governando o país X”. Utilizamos como países de referência Argélia, Catar, Egito, Líbia e Síria, todos de maioria árabe. Mantivemos o estilo de desenho animado nas instruções, para facilitar a comparação dos resultados e também por considerar que este tipo de personagem, vinculado a produções cinematográficas bastante populares, tornam-se modelos de inspiração para meninas, contribuindo para a modificação ou reforço de padrões comportamentais.

No primeiro teste sobre como o Dall-E retrata mulheres árabes, ordenamos à ferramenta que criasse uma personagem de uma mulher árabe vivendo no Catar, um dos países que havia pesquisa na tese de doutorado. O resultado foram quatro imagens bastante amigáveis, todas elas de mulheres jovens, usando abayas e cobrindo os cabelos com véu. Destas, três imagens tinham um cenário que reproduzia arquitetura árabe tradicional e uma tinha esta arquitetura combinada com prédios modernos da capital Doha. Todas as personagens geradas tinham uma expressão doce e tranquila.

A pesquisa aqui detalhada iniciou de fato com o segundo teste. Pedimos à IA que criasse uma personagem “de uma mulher árabe governando um país”. Sem a especificação sobre qual país, a ferramenta de inteligência artificial interpretou a palavra como sinônimo de Estados Unidos: todas as imagens traziam a bandeira ou outros símbolos nacionais norte-americanos. Em duas das quatro imagens, o personagem que ocupava o púlpito era um homem e a mulher ficou ao lado, não na posição principal. Na última imagem, ela está no púlpito, mas acompanhada por um homem. Em sua maioria, os desenhos gerados não expressam pleno exercício do poder pela personagem feminina.

Figura 1- resultados do Dall-E 3 para uma mulher árabe governando um país

Tal situação se alinha com o estudo de (Cheong *et al.*, 2024) sobre viés em representações feitas por IA de atividades profissionais: sem a indicação do gênero no

comando, o Dall-E desenhou homens em 58% dos casos, havendo ainda discrepâncias quanto a gênero conforme os tipos de trabalho (homens foram associados a atividades que exigem mais esforço físico, por exemplo).

O que pode ter ocasionado tais representações sobre mulheres governando? Uma das possibilidades listadas foi a base de dados não ter muitas imagens de mulheres nestas posições de liderança em países árabes, pois poucas conseguiram tal feito. Como teste, substituímos “mulher árabe” por “mulher muçulmana” e “um país” por “Paquistão”, já que este país teve Benazir Bhutho como primeira-ministra. Os resultados foram estas quatro figuras:

Figura 2- resposta ao comando para uma mulher governando o Paquistão

As duas figuras à esquerda mostram a personagem sozinha ou com a população abaixo de si. A figura do quadrante superior esquerdo traz a personagem com semblante irritado, bravo. A figura abaixo traz a governante em trajes militares, com olhar determinado, em tamanho desproporcional em relação às figuras que representam a população abaixo dela. As outras duas figuras trazem a personagem com semblante tranquilo, mas nestas a mulher não está sozinha: na imagem do quadrante superior direito, há um militar, representado em escala menor; na do quadrante inferior, ela é acompanhada por outro homem, em traje militar de gala.

Diante deste resultado, demos sequência aos testes com comandos que especificavam um país árabe: “uma mulher muçulmana governando o Catar”. O Dall E 3 foi capaz de criar apenas três imagens, cada uma com uma versão de personagem feminina com coroa e vestes em que predomina a cor vinho da bandeira catari. Não há outros personagens ilustrados e apenas uma das imagens tem um cenário de fundo (uma paisagem que combina arquitetura tradicional árabe e grandes e modernos arranha-céus de Doha). É nesta imagem que a mulher retratada aparenta estar brava, com sobrancelha cerrada, além de ter uma espada na cintura. Além das coroas sobre as cabeças das mulheres, não há outros símbolos que remetam à posição de chefia.

Testamos o comando “uma mulher presidente da Síria” e dos quatro resultados, três se assemelham mais a vilãs do que a heroínas de desenhos. Embora nenhuma das personagens expresse raiva ou braveza, como visto em outras representações geradas pela

IA a nosso pedido, duas foram desenhadas com elementos que remetem a crânios e uma terceira se sobressai à frente de um exército. A quarta está à frente de uma multidão com bandeiras. O sentido expresso nos desenhos é uma associação entre poder e violência.

Figura 3 - respostas do Dall-E 3 sobre mulher presidente da Síria

Obtivemos outras quatro imagens com a ordem "uma mulher presidente do Egito". Em metade delas, há duas personagens lado a lado: em uma, a mesma mulher de cabelos escu[5]ros, vestindo um blazer e brincos dourados, é retratada com e sem a faixa presidencial; na outra, uma das mulheres veste blazer e tem os cabelos cobertos por um hijab azul e a outra, de cabelos descobertos, presos atrás da nuca, veste um uniforme azul escuro cheio de insígnias militares. Em outra imagem, a mulher sorridente, vestindo um conjunto de saia e blazer vermelhos, segura a bandeira do Egito, tendo ao fundo da cena mais três pessoas: uma mulher e um homem em trajes sociais ao estilo ocidental e um homem com trajes árabes. Na quarta imagem, a mulher usa uma saia longa e blazer vermelhos e uma tiara dourada, estando diante de um pedestal com microfone, com a bandeira do país ao lado.

Quando o país indicado foi a Argélia, obtivemos quatro imagens que retratam mulheres (e apenas mulheres) com expressão serena, junto a bandeiras do país ou adereços que fazem referência à bandeira. Podemos considerar que em pelo menos duas das opções a IA retratou mulheres não brancas. No estudo sobre IAs generativas, Cheong et al (2024) constataram que em metade dos casos o Dall-E desenhava pessoas brancas e em apenas 10% pessoas com traços étnicos do Oriente Médio. O vestuário de todas as quatro era ocidental e nenhuma das quatro opções apresenta personagens com os cabelos cobertos: a Argélia é um país de maioria islâmica, no qual o uso do véu não é obrigatório. .

O último conjunto testado foi "uma mulher árabe como presidente da Líbia", que resultou em três imagens com mulheres em primeiro plano (duas, tendo um homem ou vários homens ao fundo) ou sozinhas (uma) e uma imagem de um homem sozinho sentado em uma escrivaninha, vestindo terno e gravata, repleto de medalhas, e o tradicional lenço cobrindo a cabeça. É o homem também quem está sentado na escrivaninha em uma das imagens na

[5] Esta classificação da etnia dos personagens nas imagens geradas foi realizada pelos pesquisadores com auxílio de tecnologia de reconhecimento facial, recurso que não foi usado em nossa pesquisa.

qual a personagem feminina está em primeiro plano, ou seja, é o elemento masculino que ocupa simbolicamente a posição de comando, na mesa onde são assinadas ordens.

Figura 4 - respostas a comandos sobre mulheres da Argélia, Egito e Líbia

Durante a primeira leitura das imagens, foi considerado necessário gerar uma nova série para análise, replicando os comandos com mudanças sutis nos termos utilizados. Esta segunda rodada permitiu complementar a amostra, confirmar alguns achados e evidenciar outros.

Quando os comandos uniram o nome de um país árabe e o termo “mulher governando” e o nome de um dos países árabes, o Dall-E 3 traduziu governar como sinônimo de reinar (embora apenas um deles, o Catar, seja uma monarquia): das 20 imagens geradas, dezesseis trouxeram personagens com coroas e tiaras. Destaca-se ainda o predomínio de expressões faciais felizes ou serenas, sendo a exceção as imagens referentes à Síria, cujas hipotéticas governantes apresentam olhar e gestual mais agressivo em três dos quatro resultados. Este resultado em específico contrasta com o apontamento feito por Chen (2024), de que o Dall-E representa predominantemente mulheres como personagens sorridentes e calmos, enquanto os demais confirmam a regra.

Figura 5 - respostas para o comando sobre mulheres governando a Síria

As cabeças coroadas apareceram de forma mais presente ainda quando pedimos a imagem de “uma mulher muçulmana governando”: a única exceção entre as vinte imagens do conjunto sem coroa é um personagem com bigode e cavanhaque, vestindo túnica e véu sob os cabelos, misturando características de ambos os gêneros. Deste grupo de vinte desenhos,

ressaltamos ainda que parte das ilustrações referentes ao Catar (2/4) e à Libia (1/4) trouxeram a personagem cercada apenas por personagens masculinos como se fossem sua equipe de governo. O círculo do poder permanece predominantemente masculino, mesmo em uma situação fictícia. Embora ainda estejam distantes de uma situação de paridade na política, tanto as mulheres líbias quanto cataris podem votar e ser votadas, mas os avanços são lentos. No que diz respeito ao Catar, a modernização em outros setores e o desejo do governo de que mulheres participassem da primeira eleição realizada no país em 1999 não se traduziram em votação expressiva (Shockley, 2018).

Observando todos os conjuntos gerados, o Dall-E 3 tratou "mulher árabe" como sinônimo de "muçulmana", retomando um dos estereótipos apontados por (Mehdid, 1993): da mulher vestindo véu ou hijab. Quando foi especificado "árabe" ou "muçulmana" no comando, todas as personagens principais tinham os cabelos cobertos; sem a especificação, apenas algumas. Não há uma sexualização explícita, mas todas as personagens femininas que foram criadas são jovens e quase a totalidade são magras. A ênfase na juventude foi um atributo constatado em outras pesquisas sobre IA generativa (Sun *et al.*, 2023; Zhou *et al.*, 2024).

Em alguns casos, a ferramenta desenhou como a figura na posição de liderança um homem ou colocou um homem no púlpito ou escrivaninha e não a personagem feminina. Quando o comando incluía "mulher" e "governando", 20% dos resultados retratavam apenas homens na retaguarda da personagem principal e 35% tinham a personagem à frente de um exército ou figuras militares posicionadas junto a ela. Quando especificamos no comando de texto um cargo (mulher emir ou presidente), a personagem foi ilustrada na metade das vezes sozinha e, em 25% das vezes, com apenas homens às suas costas.

Figura 6 - personagens referentes a Argélia, Líbia e Catar, com homens na retaguarda

O sentido expresso por é de que a liderança política é uma posição masculina, em que as mulheres são raras. De fato, esta é uma realidade e este caráter excepcional é destacado nas coberturas jornalísticas. É frequente jornais destacarem o vínculo familiar com algum político homem ao noticiar sobre mulheres parlamentares, de países com baixa

representatividade feminina no poder legislativo (Joshi; Hailu; Reising, 2020), como se a mulher na política estivesse sempre vinculada ao elemento masculino.

As personagens centrais, que a IA deveria desenhar em posição de liderança, foram retratadas muitas vezes usando uniformes ou medalhas militares, ou portando espadas, lanças ou facas, numa associação entre poder e força militar. Isto foi mais pronunciado nas imagens referentes à Síria, sempre as de sentido mais negativo: nestas, encontramos vários exemplos em que a expressão facial terna contrasta com um ambiente repleto de símbolos de violência: em uma delas, a personagem carrega nas mãos uma cabeça humana e uma faca. Este achado reforça o argumento trazido por Chen de que o Dall-E tende a retratar mais vezes mulheres do que homens sorrindo, não sendo, portanto, um resultado relacionado ao estilo de desenho.

Por fim, entre os comandos usados para gerar o corpus analisado (sem considerar os testes quanto à política de uso da ferramenta), o único em que a ferramenta se recusou a exibir o resultado foi no comando “pessoas vivendo na Síria hoje”, exibindo a mensagem de que havia “conflito com a política de conteúdo do Dall-E 3 (que proíbe pornografia e promoção da violência). Neste último comando, repetido com o nome dos demais países estuddos

CONSIDERAÇÕES FINAIS:

Até o momento, a pesquisa evidencia que há remanescentes nas imagens de discursos que associam o mundo árabe à violência (em diferentes graus, conforme o país) e a vida política ao domínio masculino. Isto foi mais proeminente em relação à Síria, cujo conjunto de imagens sempre era o mais carregado de símbolos violentos: caveiras, facas, armas de fogo, e a própria expressão facial dos personagens. Em parte, isto pode ser atribuído ao período histórico recente: embora não esteja plenamente claro como o banco de dados foi composto, dificilmente ele terá sido programado sem referências fotográficas e de textos sobre a guerra na Síria, iniciada em 2011.

Quanto ao aspecto étnico/religioso, a ferramenta pouco diferencia árabe/muçulmana, trajando a maior parte das personagens com os cabelos cobertos. Isto pode indicar que o Dall-E 3 opera baseado na referência mais frequente (mais registros na base de dados de mulheres daquele país utilizando hijab, abaya...) ou que base de dados foi alimentada apenas ou majoritariamente com fotos de mulheres trajadas de outra forma.

Em relação às mulheres no poder, ficou claro desde o primeiro comando executado a falta que lideranças femininas fazem. Sem exemplos reais, as representações de líderes e governantes terminam por reforçar a divisão de papéis, pois reproduzem mais e mais vezes o padrão existente: o lugar no púlpito não é ocupado por mulheres; a mulher árabe/muçulmana governa tendo homens às suas costas ou aparenta ser tirana quando sozinha. Sem um

elemento humano que projete um futuro diferente, a máquina reproduz o que recolheu dos dados recebidos.

Para contornar os problemas de viés dos resultados, pelo que indica a literatura existente, a Open AI adotou como estratégia programar o algoritmo para que contrabalance os resultados finais, em vez de aprimorar as fontes usadas, diversificando-as. Há evidências, inclusive pela inclusão de símbolos norte-americanos nas imagens, de que a base de dados é predominantemente ocidental, o que tem consequências principalmente se a IA passar a ser utilizada para criação de publicidade ou peças de entretenimento, reforçando o tipo de representação racial e de gênero vista até aqui.

REFERÊNCIAS:

BETKER, James *et al*. Improving image generation with better captions. [*S. l.*], 2023. Disponível em: https://cdn.openai.com/papers/dall-e-3.pdf. Acesso em: 22 mar. 2024.

BUCHANAN, Bruce G. A (Very) Brief History of Artificial Intelligence. [*s. l.*], 2005.

CHEN, Honglong. A Study of Artificial Intelligence's Impact on Aesthetic Standards and Its Potential Social Dilemmas. Journal of Sociology and Ethnology, [*s. l.*], v. 6, n. 5, 2024. Disponível em: https://www.clausiuspress.com/article/13632.html.

CHEONG, Marc *et al*. Investigating Gender and Racial Biases in DALL-E Mini Images. ACM Journal on Responsible Computing, [*s. l.*], v. 1, n. 2, p. 1–20, 2024.

CINTRA, Guilherme. Five Highlights from the Meeting with Sam Altman, CEO of OpenAI. [*S. l.*], 2023. Disponível em: https://fundacaolemann.org.br/en/news/five-highlights-from-the-meeting-with-sam-altman-ceo-of-openai. Acesso em: 6 nov. 2024.

COECKELBERGH, Mark. Ética na inteligência artificial. Rio de Janeiro: UBU editora, 2023.

DÍAZ-RODRÍGUEZ, Natalia *et al*. Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, [*s. l.*], v. 99, 2023.

EDAM, Bushra Khalaf; SHAARI, Azianura Hani; ALADDIN, Ashinida. Headlines and Hegemony: Unraveling Ideological Narratives in Arab and Western Media's Portrayal of Arab Women. Journal of Intercultural Communication, [*s. l.*], v. 23, n. 4, p. 82–94, 2023.

EL MORR, Christo. AI and Society. [*S. l.*]: Chapman and Hall/CRC, 2022.

JOSHI, Devin K.; HAILU, Meseret F.; REISING, Lauren J. Violators, virtuous, or victims? How global newspapers represent the female member of parliament. Feminist Media Studies, [*s. l.*], v. 20, n. 5, p. 692–712, 2020.

KHAN, Shahnaz. Muslim Women: Negotiations in the Third Space. Signs, [*s. l.*], v. 23, n. 2, p. 463–494, 1998.

KRESS, Gunther; VAN LEEWEN, Theo. Reading Images. 3. ed. New York: Routledge, 2021.

LEE, Nicol Turner; TA, Regina. How language gaps constrain generative AI development. Brookings, [*s. l.*], 2023. Disponível em: https://www.brookings.edu/articles/how-language-gaps-constrain-generative-ai-development/. Acesso em: 24 mar. 2024.

MACDONALD, Myra. Muslim women and the Veil. Feminist Media Studies, [*s. l.*], v. 6, n. 1, p. 7–23, 2006.

MEHDID, Malika. A Western Invention of Arab Womanhood: The 'Oriental' Female. *In*: WOMEN IN THE MIDDLE EAST. London: Palgrave Macmillan UK, 1993. v. 11, p. 18–58. Disponível em: http://link.springer.com/10.1007/978-1-349-22588-0_2.

MOGHADAM, Valentine M. Empowering Women after the Arab Spring. New York: Palgrave Macmillan US, 2016. Disponível em: http://link.springer.com/10.1057/978-1-137-55747-6.

MUTHUKRISHNAN, Nikesh *et al.* Brief History of Artificial Intelligence. [*S. l.*]: W.B. Saunders, 2020.

NTOUTSI, Eirini *et al.* Bias in data-driven artificial intelligence systems—An introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, [*s. l.*], v. 10, n. 3, 2020.

OFFERT, Fabian; PHAN, Thao. A Sign That Spells: DALL-E 2, Invisual Images and The Racial Politics of Feature Space. [*s. l.*], 2022. Disponível em: http://arxiv.org/abs/2211.06323.

OPEN AI. Research. [*S. l.*], 2024. Disponível em: https://openai.com/research/. Acesso em: 6 nov. 2024.

ORLANDI, Eni P. Michel Pêcheux e a Análise de Discurso. Estudos da Língua(gem), [*s. l.*], n. 1, p. 9–13, 2005.

ROOSE, Kevin. The Brilliance And Weirdness Of ChatGPT. The New York Times, New York, 9 dez. 2022. Disponível em: http://www.nytimes.com.

SHOCKLEY, Bethany. Competence and Electability: Exploring the Limitations on Female Candidates in Qatar. Journal of Women, Politics and Policy, [*s. l.*], v. 39, n. 4, p. 467–489, 2018.

SUN, Luhang *et al*. Smiling women pitching down: auditing representational and presentational gender biases in image-generative AI. Journal of Computer-Mediated Communication, [*s. l.*], v. 29, n. 1, 2023. Disponível em: https://academic.oup.com/jcmc/article/doi/10.1093/jcmc/zmad045/7596749.

ZHOU, Mi *et al*. Bias in Generative AI. pre-print, [*s. l.*], 2024.