

UNIVERSIDADE FEDERAL DE PELOTAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ANTROPOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO PROGRAMA DE
PÓS-GRADUAÇÃO EM ARTES PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

**Acessibilidade e emoções mediadas por IA:
pesquisa interdisciplinar e tecnopoética**

Projeto de pesquisa apresentado à Universidade Federal de Pelotas no âmbito do Programa de Estímulo à Pesquisa Interdisciplinar na Pós-Graduação (PAPIn), com aporte de cotas de bolsas de Mestrado e Doutorado oriundas da chamada CNPQ N° 35/2023 (PIBPG), Edital 01/2024.

Angela Raffin Pohlmann (PPG Artes)
Adriane Borda (PPG Arquitetura e Urbanismo)
Brenda Salenave Santana (PPG Computação)
Claudia Turra Magni (PPG Antropologia)
Guilherme Ribeiro Corrêa (PPG Computação)
Luciane Kantorski (PPG Enfermagem)
Tatiana Bolivar Lebedeff (PPG Letras)

Pelotas, junho de 2024

SUMÁRIO

Introdução	2
Objetivos gerais	4
Objetivos específicos	4
Alinhamento com Objetivos de Desenvolvimento Sustentável	4
Princípios metodológicos	5
Equipe	7
Cronograma geral	8
Referências	9

Introdução

Partindo de reflexões ancoradas na linha de pesquisa "Máquinas, Corpos e Grafias" do Coletivo Antropoéticas¹, acerca dos processos de produção e recepção de imagens e narrativas visuais mediados pela IA, o presente projeto visa complexificar e expandir a discussão sobre as relações entre tecnologia, emoções e acessibilidade, envolvendo uma interlocução entre distintas áreas de conhecimento.

Compreendendo que a acessibilidade, na prática, ainda é tratada enquanto elemento excepcional voltado para um corpo divergente, que nem sempre é acolhido em suas singularidades, a pesquisa deve investigar as potencialidades e/ou entraves relacionados a alternativas tecnológicas, mediadas por IA que possibilitem a emissão e a recepção de imagens e sons, por parte de pessoas com deficiência visual e/ou auditiva, considerando: (1) a inclusão de emoções e ambiências sonoras (INGOLD, 2015) no processo de audiodescrição enquanto recurso de acessibilidade para a recepção de imagens por parte de pessoas com déficit ou sem visão, bem como (2) a produção de narrativas verbo-visuais (BRUNO, 2007) que considerem a experiência sonora para pessoas surdas no processo narrativo de conteúdos que partem da palavra falada.

Portanto, tendo presente o papel das emoções enquanto elementos imprescindíveis para a produção de sentidos, a pesquisa problematiza as relações entre os processos de leitura e recepção de imagens, assim como os processos de recepção de conteúdos verbais, considerando que os mesmos envolvem aspectos simbólicos e subjetivos, relacionados a repertórios individuais e de senso comum (socialmente compartilhados e culturalmente determinados).

Escobar (2000) sugere que as montagens, misturas, hibridizações e afetos entre corpos orgânicos e maquínicos são sempre mediadas por narrativas científicas e ficcionais, o que remete à necessidade de dissolução das fronteiras imaginadas entre natureza e artifício, ciência e política, tecnologia e sociedade, *online* e *offline*. Nesse sentido, considerando a relevância dos sentidos, das emoções e percepções para recursos de acessibilidade associados à IA é necessário uma abordagem integrada e complexa que envolva contribuições das áreas tecnológicas, humanísticas, artísticas e da saúde (PEDROSA, 1986; OITICICA, 2008; FERRAZ, 2013; PARÉ, 2010; MAGALHÃES e BEIGUELMAN, 2014; BILÁ, 2017; LOZANO-HEMMER, 2019).

As emoções, apreendidas e valoradas de acordo com o universo simbólico e cultural de cada pessoa (REZENDE; COELHO, 2010; LE BRETON, 2019), são representações fundamentais nos

¹ Trata-se de Grupo de Pesquisa cadastrado no Diretório do CNPq e vinculado ao Laboratório de Ensino, Pesquisa e Produção em Antropologia da Imagem e do Som (LEPPAIS) da Universidade Federal de Pelotas. Através do edital FAPERGS/CNPq 07/2022, o Coletivo foi contemplado pelo projeto "Estudo antropológico sobre percepções, emoções e inteligência artificial (IA)", do qual a presente proposta constitui desdobramento.

processos de significação do sujeito em sociedade. Neste sentido, a antropologia multimodal (LAPLANTINE, 2005) ao sublinhar a importância dos sentidos (HOWES e PINK, 2010; LE BRETON, 2016) e a articulação entre distintos modos de fazer, pôr em diálogo e comunicar os fenômenos estudados, pode atuar na promoção da dignidade das pessoas com deficiência.

Considerando-se a hegemonia dos sentidos da visão nas culturas ocidentais (LE BRETON, 2016), pensar formas de mediação dos processos de recepção e leitura de imagens para pessoas com deficiência visual torna-se um tema de extrema relevância no campo das tecnologias, mas também em termos antropológicos, uma vez que a deficiência surge enquanto barreira em resposta ao *ethos* da nossa cultura e, enquanto categoria social, no contexto das próprias relações. Da mesma forma, o desafio de criar imagens verbo-visuais que mediem a comunicação da palavra falada para pessoas surdas é de extrema relevância, seja em termos de acessibilidade, de inclusão social ou de direitos fundamentais do ser humano.

Portanto, a hipótese que mobiliza esta pesquisa é a de que as IAs podem ser eficazes na mediação dos processos de acessibilidade que viabilizem a recepção de imagens e sons. Mais do que isso, acreditamos que é possível treinar redes neurais para que a pessoa com deficiência possa receber, não apenas informações relacionadas ao conteúdo das imagens, mas também informações acerca de uma variedade de emoções que podem ser associadas às imagens com as quais iremos trabalhar. Se isso de fato se confirmar, avançaremos em termos da produção de estratégias que viabilizem a equidade social. Por fim, estaremos problematizando a articulação entre emoções, novas tecnologias e sentidos na produção de uma antropologia multimodal que se pretende inclusiva, podendo ser aplicada a diversos contextos de pesquisa.

Considerando as potencialidades da IA na contemporaneidade e suas múltiplas formas de aplicação, com impactos sobre a totalidade das relações humanas em termos de subjetividades, alteridades, modos de vivenciar e conceber o espaço, o tempo, o mundo e a existência, este projeto justifica-se pela proposta inovadora, relacionada ao treinamento de redes neurais que sejam capazes de mediar processos de comunicação visual e/ou sonora a partir de descrições que incluam nuances emocionais em seus respectivos processos de recepção das mensagens. Trata-se, em certa medida, de uma alfabetização da máquina para uma leitura emocional das imagens, de um lado, e da mediação da palavra falada através de visualidades produzidas por IA.

Para tanto, as competências da computação, da antropologia, das letras, das artes, da saúde e da arquitetura, por meio de profissionais que compõem a equipe deste projeto e atuam nos respectivos Programas de Pós-Graduação da Universidade Federal de Pelotas, são de suma importância, pois trata-se de expertises indispensáveis para a realização desta proposta interdisciplinar. O projeto aqui apresentado busca financiamento para **duas bolsas de mestrado** e

uma bolsa de doutorado. Poderão concorrer às bolsas estudantes matriculados nos PPGs com concentração nas áreas listadas.

Objetivos gerais

Contribuir para o estudo das potencialidades e limitações da IA no que concerne à multissensorialidade e às formas de acessibilidade nos processos de expressão e recepção de emoções através de imagens e narrativas visuais e sonoras, reduzindo desigualdades e promovendo equidade social de pessoas com deficiência em termos de acesso a formas de comunicação e apropriação espacial da cidade, bem como de conteúdos culturais e educacionais.

Problematizar teoricamente as relações entre humanos e não humanos, a partir de experimentações com tecnologias e mediação de IA para o acesso a aspectos subjetivos, que dão expressão humana (nuances emocionais) ao processo de descrição de imagens;

Objetivos específicos

- Investigar as potencialidades e/ou entraves da IA relacionados à inclusão de emoções no processo de audiodescrição e de comunicação imagética (como recurso de acessibilidade à recepção de imagens e de mensagens sonoras, respectivamente);
- Treinar uma rede neural para produção de audiodescrição para pessoas com deficiência visual considerando as nuances emocionais contidas nas imagens;
- Treinar uma rede neural que medie o processo de recepção de conteúdos orais por pessoas com deficiência auditiva a partir de recursos verbo-visuais mediados por IA.
- Tensionar teoricamente as potencialidades da IA no processo de trocas simbólicas associadas à comunicação de informação oral por meio da produção de imagens, bem como da produção de verbo-visualidades para pessoas com deficiência auditiva;

Alinhamento com Objetivos de Desenvolvimento Sustentável

O projeto articula-se ao **ODS 4 - Educação de Qualidade**, especialmente à meta 4.a, que trata de educação apropriada para crianças e adultos sensíveis às deficiências e ao gênero, proporcionando ambientes de aprendizagem seguros, não violentos, inclusivos e eficazes para todos, através de tecnologia para viabilizar a transmissão de conteúdos orais/visuais em processo formativo.

O projeto converge com o **ODS 10 - Redução das Desigualdades**, promovendo inclusão social, econômica e política, ao desenvolver tecnologia que facilita a recepção de conteúdos orais/visuais de pessoas com deficiência e contribuir para a melhoria do acesso à comunicação, à informação, à educação e ao mercado de trabalho.

Ao desenvolver uma rede neural que facilita a recepção de conteúdos orais/visuais mediada por inteligência artificial (IA), o projeto promove a inclusão e acessibilidade de pessoas com

deficiência e idosos em espaços públicos e comunitários, dialogando, portanto com o **ODS 11 - Cidades e Comunidades Sustentáveis**.

Princípios metodológicos

A pesquisa se configura enquanto uma pesquisa-ação que, além de seu escopo teórico no campo da antropologia visual, busca soluções tecnológicas para a produção de acessibilidade visual e sonora. Sob o ponto de vista tecnológico, o método a ser empregado para o estudo da relação entre as emoções evocadas pela recepção e a leitura de imagens será baseado na associação entre imagens e Inteligência Artificial (IA) e requer o treinamento de uma rede neural que traduza/incorpore nuances emocionais apreendidas das imagens no processo audiodescritivo. Já o método para acessibilidade sonora por meio da visualidade faz o caminho de transposição: texto/narrativa/som de fala para imagem.

A antropologia, por meio da observação participante, ocupa-se do registro do processo de desenvolvimento da pesquisa e dos desdobramentos oriundos das performances maquínicas, das relações entre humanos e IA, sobretudo no que diz respeito à mediação do processo de recepção de imagens por meio de audiodescrição que não se restrinja ao aspecto formal das imagens. Ademais, buscando uma relação dialógica durante todo o processo, busca a compreensão da experiência de recepção da imagem por meio da audiodescrição, primeiramente com pessoas que possuem a visão, confrontando as imagens mentais produzidas a partir da audiodescrição, com a imagem que foi mediada pela IA. E, em um segundo momento, com pessoas que se beneficiam da acessibilidade visual, buscando compreender o modo como vivenciaram a experiência de mediação pela IA.

Redes neurais são modelos computacionais inspirados no funcionamento do cérebro humano, compostas por unidades de processamento chamadas neurônios artificiais. O treinamento de uma rede neural envolve a apresentação de um conjunto de dados de entrada, chamados de dados de treinamento, juntamente com as saídas desejadas. Durante o treinamento, a rede neural ajusta os pesos de suas conexões para aprender a mapear corretamente os dados de entrada para as saídas desejadas. Por ser uma tarefa computacionalmente intensiva, especialmente para redes neurais profundas, o treinamento geralmente requer um poder de processamento adequado, tipicamente com o uso de GPUs (unidades de processamento gráfico).

Atualmente, existem várias arquiteturas de redes neurais para gerar imagens a partir de texto e vice-versa. Algumas das redes neurais mais comumente usadas são as *Generative Adversarial Networks* (GANs) (PAN, 2019). Recentemente, a arquitetura *AttnGAN* (XU, 2018) passou a ser utilizada para incluir mecanismos de atenção, possibilitando gerar imagens a partir de texto, com maior atenção a partes específicas da descrição durante a geração da imagem, melhorando assim a

correspondência entre o texto e a imagem gerada.

No contexto do projeto proposto, busca-se gerar imagens a partir de texto, levando em conta as nuances emocionais. Isso será alcançado por meio do uso de estratégias de análise de sentimentos, permitindo a extração das emoções presentes nos textos. O sentimento identificado será utilizado para filtrar as narrativas e descrições textuais das cenas, destacando palavras relacionadas ao sentimento em questão e reduzindo a importância das demais palavras. Em seguida, o novo texto será utilizado para gerar imagens com redes neurais já treinadas e disponíveis publicamente, como a *AttnGAN*². Durante o projeto, também poderão ser consideradas e investigadas outras arquiteturas, de acordo com os objetivos e resultados preliminares.

Para a geração de textos a partir de imagens, as arquiteturas mais consolidadas são a *Show and Tell* (CNN-RNN) (VINYALS, 2015) e a *Show, Attend and Tell* (SAT) (XU, 2015). Ambas as arquiteturas codificam a imagem em um vetor de características e geram uma sequência de palavras que compõem a descrição. No entanto, o SAT vai além, utilizando o mecanismo de atenção. Esse mecanismo permite que o modelo se concentre em partes diferentes da imagem durante o processo de geração da descrição, resultando em descrições mais detalhadas e relevantes. Essa capacidade de atenção do SAT torna-o uma escolha preferida quando se busca obter o máximo de qualidade nas descrições geradas a partir das imagens, inclusive incorporando nuances emocionais, como objetiva este projeto. Como a arquitetura SAT está publicamente disponível³, esta será utilizada como ponto de partida para as investigações conduzidas no projeto.

A criação de ambiências sonoras que serão associadas às imagens (desenhos, fotografias, colagens) será realizada por meio da justaposição e modelagem de sons obtidos em distintos ambientes e, também, de sons recuperados de bancos de efeitos sonoros públicos. A produção de narrativas imagéticas, além da produção das artistas envolvidas, contará com imagens do acervo público do Google imagens.

As tecnologias de acessibilidade elaboradas serão aplicadas, de modo piloto, em museus da UFPel e no site Antropoéticas, vinculado ao Laboratório de Ensino, Pesquisa e Produção em Antropologia da Imagem e do Som (LEPPAIS) com vistas à aplicabilidade em outros contextos. A partir dos resultados obtidos e após testes e adequações, será aplicada em outros contextos, como os de ensino-aprendizagem em escolas, e no Prêmio Pierre Verger, articulado ao Comitê de Antropologia Visual da Associação Brasileira de Antropologia CAV/ABA).

² <https://github.com/AaronCCWong/Show-Attend-and-Tell>

³ <https://github.com/AaronCCWong/Show-Attend-and-Tell>

Equipe

Claudia Turra Magni (Coordenadora) (CV: <http://lattes.cnpq.br/8774264386533161>) - Atua no PPG em Antropologia da UFPel. Coordena o Coletivo Antropoéticas (CNPq) do LEPPPAIS, onde se situa a linha de pesquisa Máquinas, Corpos e Grafias, dedicada às tecnopoéticas e processos comunicacionais em coletividades humanas.

Adriane Borda (CV: <https://lattes.cnpq.br/3860172079417937>) - Atua no PPG em Arquitetura e Urbanismo da UFPel. Coordena o GEGRADI (Cnpq), desenvolve pesquisas com tecnologias de representação aplicadas à produção de recursos assistivos para a interpretação de bens culturais.

Angela Raffin Pohlmann (CV: <http://lattes.cnpq.br/1188602959337493>) - Pesquisadora PQ 2 do CNPq, atua no PPG em Artes Visuais da UFPel. Trabalha com poéticas de modo geral e com a produção de desenhos, gravuras e estilos visuais distintos a serem mediados pelas tecnologias de acessibilidade ligadas às imagens e ao som.

Brenda Salenave Santana (CV: <http://lattes.cnpq.br/3048630293482453>) - Atua no PPG em Computação da UFPel, na linha de Sistemas Inteligentes. Sua pesquisa está voltada à mineração de dados, inteligência artificial, processamento de linguagem natural e análise de discursos de ódio.

Guilherme Ribeiro Corrêa (CV: <http://lattes.cnpq.br/1389878856201800>) - Pesquisador PQ 2 do CNPq, atua no PPG em Computação da UFPel, mais especificamente na linha de pesquisa Multimídia e Processamento de Sinais Visuais, com foco em processamento de imagens.

Luciane Kantorski (CV: <http://lattes.cnpq.br/3260989033020920>) - Pesquisadora PQ 1C do CNPq, atua no PPG em Enfermagem da UFPel. Suas atividades acadêmicas estão voltadas, principalmente, aos temas de saúde mental, reforma psiquiátrica e reabilitação psicossocial.

Tatiana Bolivar Lebedeff (CV: <http://lattes.cnpq.br/0081286250806899>) - Por meio de sua abordagem comunicativa e do estudo de objetos de aprendizagem para o ensino de línguas, contribui para o estudo e a elaboração de formas de comunicação inclusivas.

Cronograma geral

O cronograma de execução deste projeto foi organizado da forma apresentada na Tabela 1, dividida em oito semestres. As atividades previstas estão descritas abaixo.

Tabela 1: Cronograma de atividades a serem desenvolvidas ao longo do projeto.

Atividade	Semestre							
	1º	2º	3º	4º	5º	6º	7º	8º
1	x							
2	x							
3	x							
4		x	x	x				
5		x	x	x				

6		x		x				
7	x							
8	x							
9		x						
10					x			
11						x		
12							x	
13						x	x	x
14								x
15		x	x	x	x	x	x	
16	x	x	x	x	x	x	x	

1. Seleção dos bolsistas, elaboração de TCLE e submissão do projeto ao Comitê de Ética;
2. Levantamento e estudo sobre o estado da arte em arquiteturas de redes neurais artificiais para conversão de texto para imagem e de imagem para texto;
3. Levantamento e estudo de ferramentas e modelos pré-treinados de redes neurais artificiais publicamente disponíveis para conversão de texto para imagem e de imagem para texto;
4. Experimentação e testes com ferramentas e modelos pré-treinados de redes neurais artificiais publicamente disponíveis;
5. Treinamento de novos modelos e modificações arquiteturais em redes neurais artificiais para incorporar nuances emocionais à geração de textos e imagens;
6. Desenvolvimento de ferramenta (website ou aplicativo) para gerenciar e automatizar a avaliação subjetiva de imagens e textos gerados por indivíduos selecionados;
7. Condução de avaliações subjetivas de imagens e textos gerados por estudantes voluntários da UFPel, a partir de chamada pública;
8. Desenvolvimento de pesquisa bibliográfica;
9. Preparação para a aplicação das metodologias e práticas de campo previstas no projeto;
10. Aplicação, piloto, da tecnologia desenvolvida em um museu da Universidade Federal de Pelotas e em trabalhos expostos de modo permanente no site do Antropoéticas⁴;
11. Análise de resultados a partir da experiência de visitantes, mediadores e gestores do museu;
12. Elaboração de parâmetros e instrumentos que permitam aplicar a tecnologia desenvolvida em diversos contextos (museus, escolas, etc.);
13. Divulgação dos resultados em eventos científicos e publicações em revistas;
14. Criação de banco de imagens e efeitos sonoros produzidos pelos bolsistas;
15. Trabalho de campo e etnografia;
16. Relatório da pesquisa.

⁴ <https://www.antropoeticas.com/>

Referências

- BEIGUELMAN, Giselle. **Futuros Possíveis: arte, museus e arquivos digitais**. Editora Peirópolis: Edusp, 2014.
- BEIGUELMAN, Giselle; LA FERLA, Jorge. **Nomadismos tecnológicos**. Editora Senac, 2011.
- BEZERRA, Daniele Borges; ELIAS, Alexsânder Elias; MARTINS, Valéria de Paula; MOURA, Lisandro de Lima; PINHEIRO, Patrícia dos Santos; TAMAYO, Luís Carlos. Etnografias multissensoriais e mediações Antropoéticas. **Iuminuras**. Disponível em: <https://seer.ufrgs.br/index.php/iluminuras/article/view/130192>. Acesso em jun. de 2023.
- BRUNO, Fabiana. BRUNO, Fabiana. Cartografia verbo-visual da velhice: fotobiografias e montagens de memórias. **Revista Chilena de Antropologia Visual**. n. 10, Santiago, dez. 2007. 30-53 p. Disponível em http://www.rchav.cl/2007_10_art02_bruno_&_samain_por.html#Layer2. Acesso em abr. de 2023.
- BILÁ, Gabriela. **Teleport City**. Catálogo, 2017. Disponível em https://issuu.com/novoestudiobsb/docs/teleport_city_catalogo_issuu_01_e67054bfdaaa97. Acesso em jun. de 2023.
- HOWES, D. e PINK, S. “The Future of Sensory Anthropoogy/The Anthropology of the Senses. **Social Anthropology**. 2010, v. 18, n. 3: Pp.331-40.
- LAPLANTINE, F. **Le social et le sensible: introduction à une anthropologie modale**, Paris: Téraèdre. 2005.
- LE BRETON, David. **Antropologia dos sentidos**. Petrópolis, RJ: Vozes, 2016.
- LE BRETON, D. **Antropologia das emoções**. Petrópolis, RJ: Vozes, 2019.
- LOZANO-HEMMER, Rafael. **Linear atmosponia**. Disponível em: https://www.lozanohemmer.com/linear_atmosponia.php 2019. Acesso em jun. de 2023.
- PAN, Zhaoqing et al. **Recent progress on generative adversarial networks (GANs): A survey**. IEEE access, v. 7, p. 36322-36333, 2019.
- REZENDE, C.; COELHO, M. C. As emoções nas sociedades ocidentais modernas. In: **Antropologia das emoções**. Rio de Janeiro: FGV, 2010.
- VINYALS, Oriol et al. Show and tell: A neural image caption generator. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2015. p. 3156-3164.
- XU, Kelvin et al. **Show, attend and tell: Neural image caption generation with visual attention**. In: International conference on machine learning. PMLR, 2015. p. 2048-2057
- XU, Tao et al. **Attngan: Fine-grained text to image generation with attentional generative adversarial networks**. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 1316-1324.11