

UNIVERSIDADE DE LISBOA
FACULDADE DE BELAS-ARTES



IMAGEM INVISÍVEL

Estudo sobre imagem, acessibilidade e inteligência artificial

Ana Carla Batista Corrêa

Trabalho de Projeto

Mestrado em Design de Comunicação

Trabalho de Projeto orientado pelo Prof(a). Doutor(a) Sonia Isabel Ferreira dos Santos Rafael

2021

RESUMO

Este trabalho de projeto tem como ponto de partida o conceito de “imagem invisível” e procura investigar as possibilidades de criar melhores descrições de imagens digitais para expandir a acessibilidade na web através de tecnologias como a inteligência artificial. Essa pesquisa explora três principais pilares: imagem, acessibilidade e inteligência artificial. Primeiro, procura-se compreender de que modo uma imagem é construída baseada nos sentidos. Em seguida, analisa-se a acessibilidade na web hoje através de investigadores das áreas. Por fim, desenvolve-se uma investigação sobre inteligência artificial, traçando pontos sobre reconhecimento de imagem e visão computacional para entender como a tecnologia pode auxiliar a acessibilidade com o objetivo final de transformar a web em um ambiente mais inclusivo. Essa investigação resulta em um projeto digital experimental chamado Imagible (disponível em www.imagible.app), uma aplicação com habilidade de descrever imagens e fotografias em texto. No final, essa pesquisa e seu projeto resultam em uma validação conclusiva em que o projeto Imagible é capaz de auxiliar designers, criadores de conteúdo e desenvolvedores a escrever textos alternativos, assim como pode concluir que novas tecnologias podem beneficiar ferramentas de acessibilidade para construir uma web mais inclusiva.

Palavras-Chave:

Imagem; Acessibilidade; Inteligência Artificial; Visão Computacional; Reconhecimento de imagem

ABSTRACT

This project has the concept of the “invisible image” as a starting point and is looking to investigate possibilities to improve digital image descriptions and expand accessibility on the web through artificial intelligence technologies. This research explores three central pillars: image, accessibility, and artificial intelligence. The project starts studying image construction based on the senses. Then, analyze the web accessibility today over field researchers. Finally, it starts an investigation on artificial intelligence, raising points about image recognition and computer vision to realize how technology can help accessibility with the ultimate goal of making the web a more inclusive environment. This investigation results in an experimental digital project, Imagible (available at www.imagible.app), an application with the ability to describe images and pictures by text. In the end, this research and project result in a conclusive validation that Imagible can assist designers, content creators, and developers to write proper alternative text, and it also concludes that new technologies can benefit accessibility tools to build a more inclusive web.

Keywords:

Image; Accessibility; Artificial Intelligence; Computer Vision; Image Recognition

AGRADECIMENTOS

A presente dissertação de mestrado, toda investigação e o desenvolvimento de projeto se dá em agradecimento ao apoio de algumas pessoas em especial. Em primeiro lugar, não posso deixar de agradecer à minha orientadora, Sonia Isabel Ferreira dos Santos Rafael, por todo o auxílio e empenho com que sempre me orientou nesta jornada de investigação. Igualmente, desejo agradecer pelo grande esforço de Myrian Katto, programadora e parte da equipe de desenvolvimento do projeto, no qual sempre apoiou e se dedicou a tornar o projeto Imagible.app uma realidade, auxiliando na criação da aplicação e desenvolvendo as integrações de APIs com muito sucesso. E por fim, quero agradecer à Faculdade de Belas Artes que nos possibilitou a criação desse projeto.

ÍNDICE

Introdução	9
Áreas de estudo	10
Investigação Ativa: Imagible.app	13
Objetivo de estudo e questões de investigação	14
Enquadramento	15
Capítulo I: Sobre Imagem	19
Introdução	19
Construção e Percepção da Imagem de Aumont	20
Espaço e percepção da imagem	23
A função da imagem	27
Imagem e seu espectador	28
O imaginário de Sartre	29
Imagem e Memória	32
A imagem fotografada	33
A imagem como comunicação e linguagem	35
Capítulo II: Sobre Acessibilidade	38
Contexto	38
Foco de Acessibilidade	39
Design de Inclusão	40
Liberdade e independência	43
Tecnologias Assistivas	43
Screen Readers	47
Imagens invisíveis e alt-text	49
Audiodescrição	51
Regras de acessibilidade	53
Futuro da acessibilidade e evolução da tecnologia assistiva	55
Capítulo III: Sobre reconhecimento de imagem	57
O que é a imagem invisível?	57
Reconhecimento de Imagem	58
Gombrich: reconhecimento e rememoração	59
Reconhecimento de Imagem no Séc. XX e XXI	60
O Boom da Inteligência Artificial e do Machine Learning	63
Introdução ao computer vision	65
Reconhecimento e classificação de padrões em imagem	68
Reconhecimento e Descrição de Imagem	70

Computer vision em benefício da inclusão	71
O outro lado da inteligência artificial	71
Introdução à problemática	71
Algoritmo de exclusão e bias de classificação	74
Dados são produtos e o produto somos nós	79
A inclusão é o destino final	85
Investigação ativa: Projeto Imagible.app	89
Nota introdutória	89
O projeto	90
Metodologias	90
Fase: Exploração	92
Ideação e descobertas	92
Grupos, Personas & Foco	93
Tanya, 32, Product Designer	95
Marcus, 26, Designer	95
Andrei, 39, Desenvolvedor	96
Martina Prado, 32, Front-end developer	96
Estudos de caso e entrevistas	97
Estudo de caso: Seeing AI, Microsoft	99
Seeing AI User Story	102
Projetos de Referência	103
Fase: Criação	104
Branding	104
Regras de criação e desenvolvimento pré-estabelecidas	104
Arquitetura de informação	105
Wireframes	105
Protótipos	108
Desenvolvimento e Implementação	110
Conclusão e validação	112
Resultados de testes	113
Participantes do user testing	114
Requerimentos aplicados para o teste	114
Objetivo do produto a ser testado	114
Barreiras	114
Soluções	115
Motivações	115
Premissas e crenças iniciais	115
O que não está em teste?	115

Parâmetro de sucesso para o teste	115
Jornada de entrevista	115
Tarefas para performance	115
Principais resultados primários	116
Implementações futuras sobre testes	118
Roadmap	118
Notas finais sobre resultados	119
Bibliografia	121

Introdução

O que é a imagem invisível? Uma imagem pode existir sem o mundo visual? É possível a percepção de uma imagem sem aquilo que a torna tangível? Dentro destes questionamentos propõe-se aqui a investigação sobre a "Imagem sem imagem" – a imagem invisível, que não tem visibilidade e que não corresponde a uma realidade sensível.

Num primeiro momento, este trabalho explora as ideias de pensadores e filósofos sobre o modo como a imagem é construída com base nos sentidos. Em seguida, apresenta análises de investigadores científicos da área de acessibilidade e inteligência artificial, construindo assim uma investigação sobre a imagem relacionada à tecnologia e à acessibilidade, levantando pontos como o reconhecimento de imagem e a visão computacional para compreender como a tecnologia podem auxiliar a acessibilidade, com o objetivo final de tornar a web um ambiente mais inclusivo.

A investigação ativa resulta na criação de um projeto digital, uma aplicação capaz de descrever cenários cotidianos por meio de texto. A construção desta aplicação permitirá levantar hipóteses sobre a questão central da dissertação – em como a inteligência artificial pode auxiliar designers e desenvolvedores a construir uma web mais inclusiva e acessível?

A aplicação, chamada de Imagible (disponível em www.imagible.app), é um experimento no auxílio à inclusão, uma aplicação web capaz de descrever imagens com o intuito de auxiliar designers, desenvolvedores e criadores de conteúdo, na construção de textos alternativos mais eficazes.

1. Áreas de estudo



Figura 1. Gráfico com as dimensões das áreas de estudo. Descrição da imagem: A imagem contém três círculos que se entrelaçam nas cores rosa, azul e amarelo com os escritos "Acessibilidade, Imagem e Inteligência Artificial".

Atualmente ferramentas tecnológicas combinadas com inteligência artificial são elementos poderosos para acessibilidade, neste trabalho se explora exatamente essa relação entre a imagem acessível e a inteligência artificial. Como resultado de pesquisa, criou-se uma plataforma com a habilidade de descrever imagens cotidianas, ambientes e cenários, com o intuito de auxiliar designers e desenvolvedores na construção de legendas, descrição de imagens e textos alternativos adequados. A ideia por trás do projeto é explorar ferramentas como a visão computacional de forma ética, ajudando-a a se tornar um elemento chave para uma web mais inclusiva.

A investigação aqui se divide na teoria, sobre uma investigação intensa sobre os principais elementos dessa tese para assim construir a base e os pilares deste projeto.

No primeiro capítulo, falamos sobre a imagem, que se aprofunda nas análises de Aumont (2004) e Sartre (1996), onde na concepção sartriana a ideia de imagem é ligada à ideia de consciência e que os objetos só existem enquanto nós pensamos neles. Enquanto para Aumont (2004), vemos toda sua concepção construtivista, onde o sistema visual está equipado "por

construção" com instrumentos capazes de reconhecer os elementos visuais que constroem a percepção de objetos e espaços, como fendas, linhas e ângulos. Para Aumont, o olho não constrói percepções, mas sim coleta informações.

A partir desses conceitos, surgem os primeiros questionamentos deste projeto: seria possível uma máquina capturar, extrair e transmitir informações assim como o olho humano? Essa captura pode se dar de maneira neutra? Como uma máquina pode capturar uma percepção sem modificar sua interpretação? Também sobre essas ideias e conceitos de Sartre e Aumont, surge a questão: o que é a imagem invisível? E assim nos aprofundamos no questionamento, é possível capturar uma imagem não-visual? A imagem, por mais que remeta a uma ideia visual, é composta por outros sentidos. Como percebido por Aumont (1992) e Sartre (1996) o imaginário da imagem é criado também por ideias não-visuais.

Já no segundo capítulo, falamos sobre acessibilidade, principalmente a acessibilidade no mundo digital, as ideias de investigação sobre acessibilidade partem da visão de Sarah Horton e Whitney Quesenbery (2013), mas também inclui-se neste segundo capítulo a relação entre o design, a acessibilidade e a tecnologia, partindo das ideias de Laura Kalbag (2017).

Apresenta-se uma investigação sobre as principais tecnologias utilizadas na acessibilidade da web, chamadas de tecnologias assistivas, como o *screen reader*, que é uma das ferramentas que pessoas com deficiência visual utilizam em seus dispositivos.

As tecnologias assistivas podem dar às pessoas acesso à web que de outra forma não teriam. Essas ferramentas auxiliam na possibilidade de comunicação com o mundo digital, portanto, é importante considerar que tais tecnologias auxiliam a independência das pessoas.

Apóia-se a ideia de Kalbag (2017) que considera o design universal o melhor caminho para inclusão, onde o design projetado com acessibilidade é criado especificamente para incluir pessoas com deficiência e o design universal foca em projetar produtos para o mais amplo espectro de pessoas possível. Segundo a autora, é necessário o uso da empatia como o primeiro passo para a acessibilidade não ser apenas uma ferramenta extra, mas sim parte do todo. Kalbag enfatiza que é através desse fator que designers se tornam bons criadores de produtos, entendendo as diferentes barreiras a fim de criar melhores soluções.

Mas apenas isso não é suficiente, é necessário tomar consciência de que a acessibilidade faz parte do processo de design e precisa estar presente em todas as etapas de um projeto. A investigação ativa dessa dissertação explora a função de designers e desenvolvedores como os

grandes responsáveis por tornar a web um local mais inclusivo, e com isso sua indispensabilidade em conhecer ferramentas que auxiliem a criação desse espaço.

O capítulo foca nas ideias de Horton e Quesenbery (2013) para discutir a importância tanto da empatia quanto de ser flexível. Para as autoras, é necessário abraçar a empatia e a flexibilidade natural durante pesquisas e testes, para poder entender onde é necessário aplicar o conhecimento do design e da tecnologia a fim de criar a inclusão por meio da independência e da liberdade.

Por fim, se apresenta o terceiro capítulo, sobre reconhecimento de imagem, onde se procura novas tecnologias assistivas e as relaciona com os avanços tecnológicos a fim de explorar e apresentar recursos como descrições de imagem por áudio através do Computer Vision e da Inteligência Artificial. Que também abre o caminho para a investigação da relação entre o reconhecimento de imagem e a acessibilidade para deficientes visuais, possibilitando a transformação de textos, descrições e legendas através do seu uso.

Para isso, percorre-se os primeiros passos da inteligência artificial na visão de David Carmona (2019) até o seu renascimento no ano de 2015 - momento em que houve crescente sofisticação das técnicas na área de Machine Learning e Deep Learning.

Também explora-se a trajetória do computer vision desde a Revolução Digital até os mais recentes avanços tecnológicos e estudos, que foram os fatores que levaram à evolução da capacidade do computer vision na última década.

Aqui também levantam-se hipóteses e críticas às novas tecnologias, que seguem crescendo imensamente em áreas que nem sempre são neutras e trazem benefícios à sociedade como um todo. É notável hoje que a tecnologia pode tanto ser utilizada para inclusão, como também pode ser o ponto de partida para a exclusão e a manipulação.

Cada vez mais vê-se tecnologias de reconhecimento de imagens para fins militares, de vigilância, de monitoramento, para compras online etc. Assim como aponta-se o outro lado da inteligência artificial, que é carregada de bias e imparcialidade devido aos seus bancos de imagens. Destaca-se muito as pesquisas de Kate Crawford (2013), a qual aponta que vivemos em uma era em que dispositivos pessoais são vistos como servidores da necessidade pública, aumentando as desigualdades já existentes na sociedade.

Como resolução, essa investigação explora iniciativas e aplicativos como o Envision AI da Envision Technologies e o Seeing AI da Microsoft, que desejam criar um mundo mais

inclusivo através do uso da inteligência artificial. Apresentam-se projetos como o Center Human Technology¹, criado por Tristan Harris, que busca na tecnologia e na transformação de sistemas, a mudança na mentalidade a partir da qual os sistemas de tecnologia persuasiva são construídos, realinhando a tecnologia com a humanidade.

Muito ainda pode ser explorado no campo de computer vision, mas é necessário que exista um foco para que essas ferramentas sejam utilizadas em benefício da sociedade, pois ainda parece haver maior interesse no seu desenvolvimento no que diz respeito à conversão de lucros. É preciso criar novas ferramentas de maneira responsável e evoluir as que já existem para tornar o espaço digital um local cada vez mais inclusivo para todos.

2. Investigação Ativa: Imagible.app

Diante da temática de investigação, acessibilidade e reconhecimento de imagem, decide-se desenvolver um projeto experimental para criar ferramentas de auxílio às tecnologias acessíveis.

A criação do projeto utiliza métodos e processos de Design Thinking durante toda a descoberta e investigação dos temas, onde se busca entender as dificuldades do usuário e conhecer o caminho a ser explorado. Também se definem as jornadas e as problemáticas para criação de uma estratégia de soluções a partir da ideação que possuam valor e possam ser transformadas em ideias e projetos. Após essa etapa, parte-se para a criação do produto, passando pela arquitetura de informações, definições de tecnologias, criação de wireframes e a interface UI, culminando no início de um desenvolvimento experimental de *computer vision* em paralelo à testes de *user testing* e *rapid testing*, desenvolvendo melhorias e validações do produto.

Esse projeto utiliza a empatia em todas suas fases, utilizando principalmente a perspectiva de suas *personas*, a fim de escutar a voz de tais para criar soluções mais acessíveis e inclusivas. Em adição, esse projeto conta com várias técnicas experimentais, onde se procura sistematicamente evidências e soluções para apoiar as ideias presentes, mantendo em mente a complexidade para testar e desenvolver o mesmo.

Hoje em dia o texto alternativo necessita ser adicionado manualmente, e por isso esse projeto tem como objetivo auxiliar grupos de personas como **desenvolvedores, criadores de**

¹ Informação sobre Center Human Technology disponível em www.humanetech.com

conteúdo e designers, na criação de imagens acessíveis e de alt-text automáticos com revisão para otimização de processos. A ideia é auxiliar na construção de descrições apropriadas de imagens, construindo, ou ao menos se aproximando de, uma equidade na web para pessoas com deficiência visual.

Os métodos e experimentações desse projeto resultam, por fim, na criação de um *app web* capaz de descrever cenários cotidianos através de texto. A construção desta aplicação funciona através do reconhecimento de imagem, proporcionando reflexões sobre o que é a "imagem sem imagem". Esta aplicação, que ainda está em fase Beta e em testes, tem como foco auxiliar tecnologias assistivas de maneira a ser utilizada em benefício da inclusão. Os resultados alcançados são:

- a. Uma aplicação web open source, chamado [imagible.app](#), no qual pessoas podem utilizar da ferramenta reconhecimento de imagem para descrever imagens e gerar *alt-text* automaticamente, sob revisão, para seus projetos.

3. Objetivo de estudo e questões de investigação

A metodologia desta investigação parte da exploração do método científico, assim como a componente prática, para compreender e explorar o tema de estudo, no qual se tem como objetivo a criação de um projeto experimental que relacione o reconhecimento de imagem e a acessibilidade. Investiga-se aqui um futuro tecnológico em benefício da acessibilidade.

O interesse pelo objeto desse estudo emerge da necessidade de aplicar a acessibilidade aos passos da criação de produtos digitais. Com mais de oito anos de experiência como designer de produto, é evidente na minha perspectiva que se carece de mais estudos sobre a acessibilidade, maior conhecimento e mais ferramentas para auxiliar esse processo que é frequentemente ignorado por muitos criadores de produtos digitais, que envolve desde pessoas corporativas, criadores de conteúdos a desenvolvedores e designers.

O principal questionamento sobre o que é uma imagem invisível traz a reflexão do modo como a tecnologia pode se tornar uma aliada da acessibilidade. Neste contexto, este projeto procura explorar questões como:

- A visão computacional pode descrever imagens tanto quanto a visão humana?

- A acessibilidade está inclusa como parte do processo de design?
- Como a inteligência artificial, o computer vision, o reconhecimento de imagem pode auxiliar designers e desenvolvedores a construir uma web mais inclusiva e acessível?

Questões secundárias:

- O que é a imagem invisível?
- O que é uma imagem sem imagem?
- Como disponibilizar informação de imagens de uma maneira não visual?
- O que é uma descrição de imagem?
- Como as tecnologias assistivas suportam a descrição de imagens?
- Quais são as tecnologias assistivas para pessoas com deficiência visual?
- Como a tecnologia assistiva torna-se uma ferramenta de independência?
- Como o reconhecimento de imagem pode auxiliar o acesso à informação?
- Como mudar a inacessibilidade do mundo digital?

4. Enquadramento

Estima-se que 2,2 bilhões² de pessoas no mundo possuem algum tipo de deficiência visual e, destas, pelo menos 1 bilhão de pessoas têm deficiência visual que poderia ter sido evitada ou que ainda não foi resolvida, de acordo com o *World report on vision* publicado pelo World Health Organization em 2019. Dentre esses dados estima-se que pelo menos 39 milhões são pessoas com perda de visão total, um número crescente junto ao aumento da população mundial.

A sociedade foi construída por e para pessoas que possuem a capacidade de ver, onde a visão é o mais dominante dos sentidos humanos, e isso a torna vital em todas as fases da vida. O

² **Dados extras do atual cenário segundo Organização Mundial de Saúde - WHO:** A cada 5 segundos uma pessoa no mundo fica totalmente sem visão. 90% dos deficientes visuais estão em países em desenvolvimento. A Índia é o país com maior percentual de deficientes visuais no mundo. Em média 58% tem mais de 60 anos, 32% tem entre 45 e 49 e 11% tem menos de 44 anos. Apenas 10% das pessoas totalmente cegas sabem ler em braille. Globalmente mais de 1 bilhão de pessoas necessitam de tecnologias assistivas (não apenas visuais). Aproximadamente 200 milhões de pessoas com deficiência visual não tem acesso a tecnologia assistiva. Poucos países possuem a política programa nacional de tecnologia assistiva. Para uma pessoa com visão, 90% da informação que é processada por um cérebro é visual.

recém-nascido depende da visão para reconhecer e se relacionar com sua mãe; a criança, para dominar o equilíbrio e aprender a andar; o aluno, para ir a pé à escola, ler e aprender; a jovem para participar do mercado de trabalho; e a mulher mais velha, para manter sua independência. Porém sem a visão, as pessoas necessitam lutar para aprender a andar, ler, participar da escola e trabalhar. E essa luta é ainda mais difícil quando a sociedade não é inclusiva (WHO, 2019)

Muitas pessoas com deficiência visual enfrentam vidas sociais restritas, perspectivas de trabalho reduzidas e menor independência, devido a dificuldades na execução das tarefas diárias. Assim como grandes cidades e vilas, os sistemas educacionais, esportes, mídia e tantos outros aspectos da vida contemporânea são organizados em torno da visão (WHO, 2019).

Hoje há várias tecnologias emergentes relacionadas às barreiras da visão, incluindo o uso médico de aplicativos para avaliação da vista, para auxiliar cirurgia de catarata e tecnologias de inteligência artificial para a detecção de uma variedade de doenças oculares. Essas novas tecnologias oferecem mais esperança na melhoria ao acesso e a qualidade dos cuidados de saúde às comunidades mais negligenciadas. No entanto, mais pesquisas são necessárias em ambientes do mundo real antes da ampla adoção dessas tecnologias. O uso de análises de big data também tem o potencial de aumentar o conhecimento do serviço de vigilância e etiologia das condições oculares, monitorando os resultados médicos (WHO, 2020).

Os avanços da tecnologia, como o desenvolvimento de smartphones, reconhecimento de voz e recursos de acessibilidade em sistemas operacionais de computador aumentaram drasticamente o acesso à informação e a comunicação para indivíduos com deficiência visual e cegueira. Os livros de áudio digital estão amplamente disponíveis em números cada vez maiores para pessoas com dificuldade na leitura impressa. Indivíduos com deficiência visual podem navegar usando GPS e usar bengalas eletrônicas para auxiliar na detecção de obstáculos próximos. Alguns dos exemplos fornecidos aqui não são exaustivos e com o ritmo de inovação acelerado no campo dos cuidados com a visão, é provável que haja mais avanços tecnológicos nas próximas décadas (WHO, 2020).

Por esse motivo, pesquisas sobre tecnologias acessíveis são fundamentais para que o mundo se torne mais inclusivo, assim como as pesquisas sobre inteligências artificiais e visão computacional também o são para o benefício da inclusão. É evidente que as tecnologias baseadas em reconhecimento de imagem, por exemplo, têm um grande potencial de auxílio na quebra de obstáculos não só na prevenção médica, mas também para uso cotidiano.

Pessoas com deficiência visual utilizam *screen readers* para escutar o conteúdo de uma tela e essas tecnologias assistivas podem dar às pessoas acesso à web que de outra forma não teriam sido capazes de se envolver. Mas eles só podem auxiliar as pessoas até certo ponto. O leitor de tela facilita o acesso e a inclusão dessas pessoas em redes sociais e canais de notícias, por exemplo, mas o conteúdo visual dessas plataformas somente será totalmente acessível com descrições de imagem e *alt-text* apropriados. (Kalbag, 2017).

No mundo digital um dos maiores obstáculos para acessibilidade são os textos alternativos, também chamados de *alt-text*. Designers e desenvolvedores enfrentam uma grande barreira para descrever imagens da maneira correta, ao mesmo tempo que existe uma negligência evidente por parte de corporações no incentivo à criação de descrições de imagens acessíveis.

Essa negligência, mesmo estando em inconformidade com as legislações em alguns países, afeta toda a web. Por exemplo, quando um leitor de tela se depara com uma imagem sem o atributo `alt`, ele acaba lendo o nome do arquivo (ou seja, não transmite informação específica), ou ainda pior, quando o mal uso é atribuído em benefício do marketing, ele lê as palavras chaves designadas para mecanismos de busca (Horton, Quesenbery, 2013)

A tecnologia está em constante evolução e a uma velocidade impressionante. Novos dispositivos, sistemas operacionais, navegadores e tecnologias assistivas são lançados todos os dias e todos afetam de alguma maneira o desenvolvimento da acessibilidade.

A tecnologia está em constante evolução e a uma velocidade impressionante. Novos dispositivos, sistemas operacionais, navegadores e tecnologias assistivas são lançados todos os dias e todos afetam de alguma maneira o desenvolvimento da acessibilidade.

Entretanto, em contraste com o ritmo vertiginoso das mudanças tecnológicas, as melhores práticas de acessibilidade tendem a evoluir com muito menos frequência e em menos velocidade, e isso inclui a maneira de descrever conteúdos visuais. A acessibilidade não necessariamente se refere tanto a desenvolvimentos tecnológicos, mas principalmente em encontrar muitas maneiras de melhorar a tecnologia e as experiências atuais (Kalbag, p.147, 2017).

É necessário que os métodos aplicados acompanhem essa crescente e principalmente que utilize essas tecnologias, como a inteligência artificial e o reconhecimento de imagem. A visão computacional representa exatamente essa junção, tendo um grande potencial para auxiliar os meios de acessibilidade a trazerem a independência e confiança para essas pessoas. A

acessibilidade é a jornada, enquanto a inclusão é o destino final.

- 2.2 Bilhões de pessoas com algum tipo de deficiência visual;
- 95% da sociedade é construída sobre estímulo visual;
- Descrições de imagens são frequentemente negligenciadas na web;
- Muitos desenvolvedores, designers e criadores de conteúdos não são conscientes da importância de textos alternativos, ou não sabem como criá-los de maneira efetiva;
- Muitos designers e criadores de conteúdos não são conscientes da existência de alt-text;
- Tecnologias assistivas, assim como leitores de tela, possuem limitações e não estão evoluindo na mesma velocidade que poderiam.

Capítulo I: Sobre Imagem

1. Introdução

As derivações da palavra imagem e dos étimos latinos *imago*, *imagine* ou *imaginis* compõem um vasto vocabulário, e em linguagem científica, a coexistência de muitos nomes para designar uma coisa pode causar questionamentos (Bacelar, 2010). O uso de cada derivação se dá em diferentes contextos, podendo estes serem análogos ou não. Portanto, a fim de evitar ambiguidades, define-se aqui os conceitos de Imagem, Imaginação e Imaginário.

A palavra imagem significa a representação de uma pessoa, coisa ou espaço, que acontece através de uma expressão artística (pintura, escultura, fotografia etc.) ou sensorialmente graças à ausência da causa que a produziu. A definição de imaginação é a capacidade de criar ou representar pelo pensamento, é o meio e o fim do ato de imaginar. Por último, o imaginário é um processo de produção de imagens e símbolos que podem pertencer tanto a um indivíduo (Imaginário Pessoal) quanto a um coletivo (Imaginário Social), para expressar sua relação de alteridade com o mundo; ele existe unicamente na imaginação ou, ao menos, necessita desta para ser alcançado.

Acerca dos estudos sobre a imagem em relação à imaginação e à consciência, há diversos desdobramentos filosóficos possíveis, tornando inconcebível apontar uma definição unânime entre todos os pensadores. Seria simplista e errôneo ignorar pontos conflitantes entre os estudos, levando-se também em consideração que essa continua a ser uma temática muito pesquisada. As divergências são necessárias no processo de constante transformação a respeito do que se entende como "ideia" e "percepção" da imagem.

Para apresentar algumas das múltiplas linhas de pesquisa sobre o assunto, este trabalho aprofunda-se individualmente no que propõem pesquisadores como Donis A. Dondis & Camargo (2007), Gilbert Durand (1997), Gilles Deleuze (1989), Henri Bergson (1999), Jacques Aumont (2001), Jacques Le Goff (1994), Jean-Paul Sartre (1996), Platão (428-348 a.C.), Roland Barthes (1989), entre outros.

2. Construção e Percepção da Imagem de Aumont

É necessário investigar os processos de construção e de percepção da imagem, tal qual sua relação com o imaginário humano, apontados pelo teórico, escritor e professor francês Jacques Aumont (1942-).

Para Aumont (2001) o processo de construção da imagem inicia-se do impacto da luz sobre algo ou alguém, criando-se uma formação visual, uma superfície que pode ser observada. O olho age, então, como um dispositivo de captura que extrai para o cérebro infinitas possibilidades de interpretações humanas. Segundo a linha de pensamento do autor, o papel do olho não é de construir percepções ou "decodificar" *inputs*, ou seja, perceber de fato e dar significado ao que se vê, mas sim o de extrator da informação.

As unidades elementares que são responsáveis por essa extração e também por formar a percepção do objeto e do espaço, são constituídas, portanto, pelo próprio sistema visual do olho. Este, desenvolve no decorrer da vida instrumentos capazes de reconhecer um contorno visual e sua orientação, uma fenda, uma linha, um ângulo, um segmento. São essas as ferramentas que tornam a percepção uma atividade direta (Aumont, 2001).

Ainda assim, a percepção visual é um ato inconsciente do ser humano. Mesmo ao basear-se nestes instrumentos de reconhecimento que podem ser descritos, a complexidade acontece pelas inúmeras possibilidades de contextos em que eles podem ser aplicados. Como exemplo, citado por Aumont (2001), a noção individual de forma pode se tornar abstrata quando em relação às características físicas nas quais se materializa: ela pode mudar de tamanho, de espaço e até de elementos que a compõem, sem no fim ser verdadeiramente alterada como forma.

Devido a estes fatores passíveis de diferentes significados, é importante ressaltar o interesse do autor francês em tratar sobre a "forma total", que é construída pelas características de um objeto representado em uma imagem, trazendo com isso, a percepção da forma como uma unidade completa, na qual a configuração da estrutura de suas partes são formadas de maneira racional, não aleatória.

Por exemplo, para o reconhecimento da imagem de um carro não seria cabível identificar apenas partes separadas do todo, como apenas a roda, a porta ou o parabrisas, mas sim fazer a identificação da relação espacial entre todos os elementos que o compõem, formando-o enquanto um objeto completo, identificando a ele uma forma total. Pode-se dizer assim, que a identificação

de objetos realizada através da extração de informação de elementos da forma, deve ser feita como um todo, o que a torna parte importantíssima no processo de compreensão da imagem exercido pelo espectador.

É notável como a atividade direta de extração de informação na criação de uma imagem, pode ser comparada, de certa maneira, ao processo de captura de informações por uma máquina fotográfica:

[...] se o olho se assemelha até certo ponto a uma máquina fotográfica, se a retina é comparável a uma espécie de chapa sensível, o essencial da percepção visual realiza-se depois, através de um processo de tratamento da informação que, como todos os processos cerebrais, está mais próximo de modelos informáticos ou cibernéticos do que de modelos mecânicos ou ópticos. (Aumont, 2001, p.22).

A comparação feita pelo autor entre o olho e uma máquina fotográfica só se aplica, portanto, à ótica do processamento da luz, onde a percepção visual é a responsável pelo processamento da informação que nos chega e entra em nossos olhos em etapas sucessivas (Aumont, 2001). Essa teoria é apresentada, nos parágrafos seguintes, de maneira explicativa.

Como toda informação, a percepção é codificada — em um sentido que não é o mesmo da semiologia: os códigos são, aqui, regras de transformações naturais que determinam a atividade nervosa em função da quantidade e da qualidade de informações contida na luz; onde o sistema visual é capaz de localizar e de interpretar certas regularidades nos fenômenos luminosos que atingem nossos olhos (Aumont, 2001).

Assim como a compreensão da luminosidade provém das reações do sistema visual à luminância dos objetos, a compreensão da cor provém de suas reações ao comprimento de onda das luzes emitidas ou refletidas por esses objetos, por conseguinte, tanto a cor como a luminosidade não estão "nos objetos", mas sim na percepção visual do espectador. Nas palavras do autor, "a própria percepção não é um processo instantâneo; certos estágios da percepção são rápidos, outros muito mais lentos, mas o processamento da informação se faz sempre no tempo." (Aumont, 2001, p. 31).

Aumont (2001), citando Platão, afirma que as propriedades físicas do mundo não dependem do nosso olhar sobre ele. O mundo tem, a grosso modo, sempre a mesma aparência no dia a dia, ou pelo menos espera-se encontrar nele uma vasta quantidade de elementos invariáveis. É a percepção desses aspectos consideravelmente inalteráveis (o tamanho dos objetos, suas formas, a localização, orientações, propriedades das suas superfícies etc.) que se designa a noção de constância perceptiva: apesar da variedade das percepções, localiza-se as constantes.

Em uma imagem se reconhecerá a forma de um ser humano, não apenas se for possível identificar um rosto, um pescoço, um torso, braços etc. (tendo cada uma dessas unidades, aliás, uma forma característica), mas se certas relações espaciais entre esses elementos forem respeitadas. (Aumont, 2001, p. 68).

Além da abordagem analítica, é importante ressaltar que a abordagem sintética do autor consiste em assimilar a percepção feita do mundo visual em um estímulo único, onde a imagem que passa pela retina, mesmo com modificações no tempo, contém toda informação necessária à percepção dos objetos no espaço, já que o sistema visual está suficientemente equipado para processá-la nesse sentido.

Segundo Aumont (2001), essa abordagem é representada pela ideia de inatismo desde o séc. XIX. Já no início do século XX, os gestaltistas, ou teóricos da Forma, voltaram a insistir na capacidade inerente do cérebro em organizar o visual segundo leis universais e temporais. Mas foi somente com os trabalhos de J. J. Gibson, a partir de 1950, que essa abordagem surge sob o nome de Teoria Ecológica da Percepção Visual.

Existe também uma oposição nítida entre a teoria da forma Gestaltista, da percepção inata do todo, e a teoria construtivista, da aprendizagem. Na concepção construtivista, supõe-se que a distinção entre a figura e o fundo é adquirida, portanto as imagens não necessitam estabelecer uma convenção suplementar por sobre as imagens figurativas em geral, essa distinção entre elas é parte da própria noção de figuração. Em termos simples, "imagem *figurativa* é a que produz ou reproduz *figuras* e seus fundos" (Aumont, 2001, p. 71). Ainda segundo o escritor, a teoria construtivista segue quatro leis relacionadas a imagem figurativa:

- *Lei de proximidade*: elementos próximos são mais facilmente percebidos como pertencentes a uma forma comum do que elementos afastados.
- *Lei de similaridade*: elementos de mesma forma ou de mesmo tamanho são mais facilmente vistos como pertencentes a uma mesma forma de conjunto;
- *Lei de continuidade*: existe uma tendência "natural" para continuar de modo racional uma determinada forma, se ela estiver inacabada;
- *Lei de destino comum*: refere-se às figuras em movimento, e afirma que elementos que se deslocam ao mesmo tempo são percebidos como uma unidade e tendem a constituir uma forma única (Aumont, 2001, p. 71).

As leis descritas baseiam-se na noção central de Forma (*Gestalt*), apesar de não haver unanimidade entre pesquisadores sobre uma definição única, e implicam em uma relação mais profunda entre os elementos da figura como um todo e os elementos de maneira isolada.

Há décadas as tentativas para superar a teoria da Forma concentram-se em torno da noção de informação (*Forma e informação*). O principal apontamento desta última é de que, em uma certa figura, existem partes que fornecem muita informação e outras que fornecem pouca. Essas partes com pouca informação se dão por consequência de serem completamente previsíveis (Aumont, 2001).

O autor ressalta também, que a possibilidade de percepção de uma imagem de maneira separada à suas interpretações, é um processo árduo somente possível à espécie humana. O que traz um dos primeiros questionamentos deste projeto: seria possível uma máquina capturar, extrair e transmitir informações tal qual o olho humano? Essa captura pode se dar de maneira completamente neutra? Como uma máquina seria capaz de capturar a perspectiva de uma imagem sem, contudo, modificar sua interpretação?

a. Espaço e percepção da imagem

O espaço físico pode ser descrito com bastante exatidão por meio do modelo geométrico de três eixos de coordenadas. Tradicionalmente chamados de X, Y e Z, os eixos são cruzados perpendicularmente por linhas num espaço cartesiano. Esse modelo deriva da geometria "euclidiana", a qual se caracteriza, entre outras coisas, pela descrição do espaço através do

apontamento de três dimensões: a verticalidade, a horizontalidade e a profundidade (Aumont, 2001). Hoje em dia esse conceito é tão estabelecido socialmente, que o termo "3D" (abreviação para três dimensões) é presente na linguagem comum, utilizado principalmente na descrição de produtos imagéticos direcionados ao entretenimento, como o cinema, a televisão e os vídeo games.

Sobre esse pensamento, Aumont (2001) diz que a percepção de espaço não acontece cotidianamente apenas pelo sentido da visão, pois o sistema visual não possui um órgão especializado na percepção de distância. Desta forma, a ideia de espaço acontece através da percepção corporal e sua relação com o restante do ambiente: horizontalmente, ao se locomover e observar os corpos que também se movem no mesmo sentido, e verticalmente, ao se sentir preso ao chão, percebendo também que todos os corpos são condicionados nessa direção (em outras palavras, a percepção da gravidade). Esse tipo de "incapacidade" do sistema visual pode ser explorado através da arte, onde diversos artistas experimentam os limites desse sentido: Maurits Cornelis Escher (1898-1972) se destacou nas artes visuais ao brincar com as possibilidades de comunicação entre os olhos e o cérebro, e o cineasta Alfred Hitchcock (1899-1980) também abusou do poder do olhar em sua obra, chegando a trabalhar com o surrealista Salvador Dalí (1904-1989).

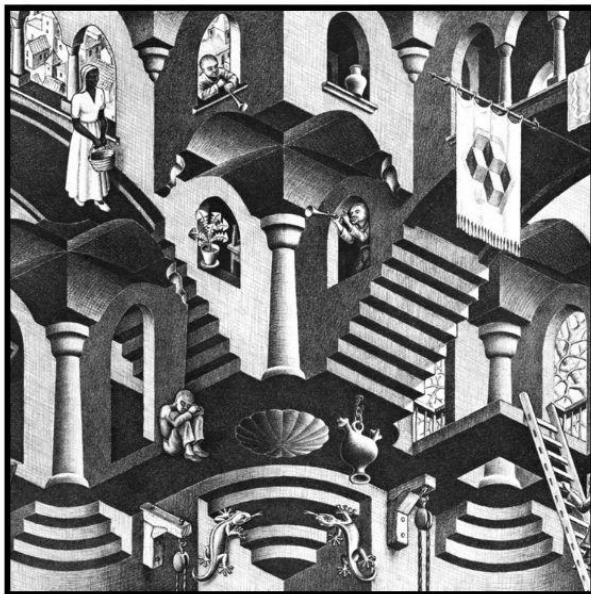


Figura 2. Côncavo e o Convexo (1955) Escher. Descrição da Imagem: A imagem é uma litografia que expressa a ambiguidade geométrica, contém elementos como colunas, janelas que podem ser interpretadas como convexas e curvas ao mesmo tempo diante do contraste dos elementos, a imagem contém diversos elementos como duas pessoas com trompetes, uma pessoa segurando uma cesta e uma pessoa sentado, assim como outros elementos como uma concha, vasos, escadas e bandeiras compõe uma cena ambígua de contrastes e formas.

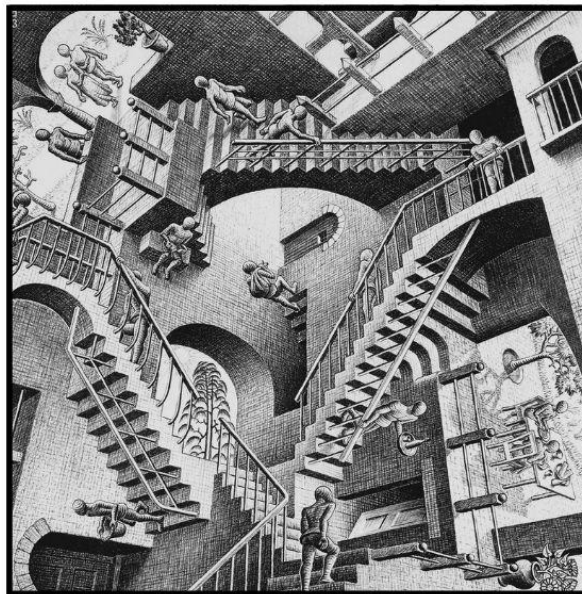


Figura 3. Relativity (1953) Escher. Descrição da Imagem: A imagem é uma litografia que retrata um universo onde as leis da gravidade não se aplicam, as formas geométricas existem em três fontes de gravidade, cada ser ortogonal aos dois outros, essas fontes foram a estrutura principal da imagem cp, sete escadas, e cada cena contém diversos elementos como dezesseis personagens, distribuídos entre cada fonte de gravidade, seis em cada um e cinco nos outros dois que forma um único elemento sobre a perspectiva e representação de imagens tridimensionais em uma imagem bidimensional.

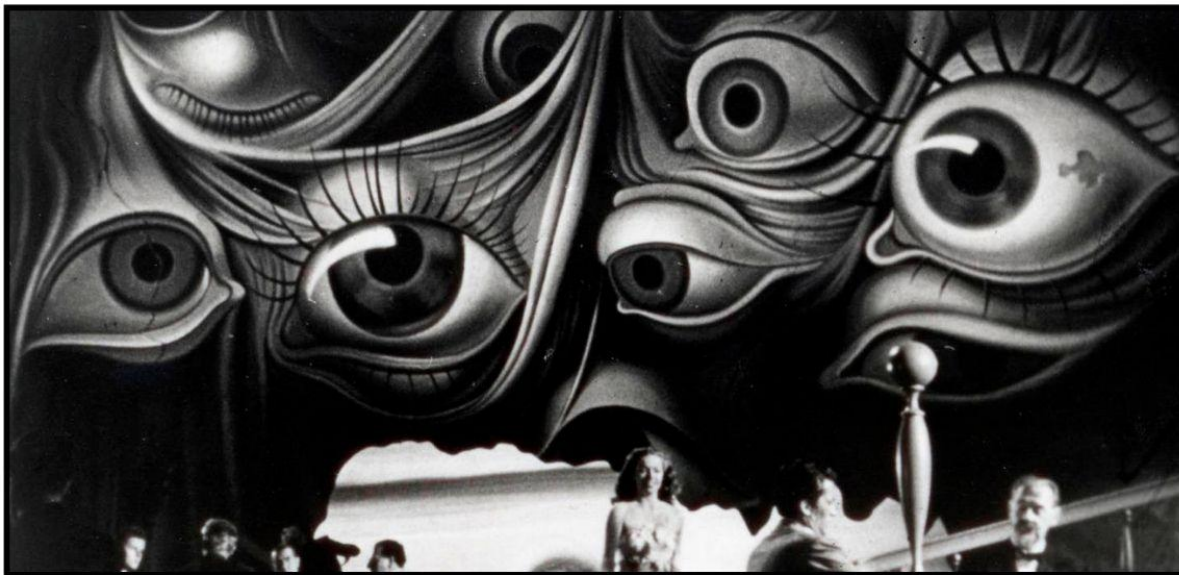


Figura 3. Cenografia de Salvador Dalí para o filme *Spellbound*, dirigido por Hitchcock (1945). Descrição da Imagem: A imagem é composta por um cenário onde tem-se como uma cortina de olhos, com oito olhos distribuídos em formas orgânicas dentro de um tecido, abaixo da cortina percebe-se alguns atores um total de sete pessoas distribuídas em um cenário lúdico e metafórico como se os olhos estivessem observando os personagens dentro de uma ilusão.

Levando em consideração o modelo geométrico apresentado, uma compreensão total imagética só é possível nessas três dimensões, e qualquer transformação que capture esse espaço em uma imagem de apenas duas dimensões, sofrerá perda de informação. As próprias categorias fundamentais de espaço e de tempo, tocam, inevitavelmente, na concepção que se tem do visível e do visual, além da conseqüente relação de um com o outro causada pela percepção. Conclui-se que o conceito de espaço é tanto de origem tátil, experiencial, quanto visual, inerte (Aumont, 2001).

Exemplos do reconhecimento de espaço por outro sentido sensorial, no caso a audição, podem ser percebidos através de animais como a coruja, o morcego e o golfinho, e mais recentemente, graças a estudos contemporâneos, até de seres humanos nascidos com algum tipo de deficiência visual. Esses animais, ainda que por vezes capacitados de uma ótima visão, possuem um mecanismo de ecolocalização, conseguindo emitir ondas sonoras que refletem em forma de eco, fazendo-os perceber o que existe ao redor. Certos pássaros como a coruja, criam

mapas sensoriais do espaço auditivo e conseguem projetar todos os pontos de informação memorizados. Os morcegos, por exemplo, possuem neurônios que detectam a velocidade dos objetos pela variação do som, medindo a distância pelo tempo de duração do retorno do som, e pelo ângulo de entrada do som refletido no seu aparelho auditivo. Entre os humanos essa habilidade é mais rara e menos eficiente, porém possível e comprovada cientificamente³.

Dentro desses exemplos levantados aplica-se a abordagem sintética, onde Aumont (2001) cita Gibson e sua teoria ecológica da percepção visual, que aponta as "variáveis complexas" como implicações nesse entendimento mais amplo do que é a visão.

b. A função da imagem

Em reflexão sobre as funções da imagem, Aumont (2001) questiona para que a imagem é utilizada. Os modelos descritos a seguir pelo autor, são constantes observadas em todas as produções feitas pelo ser humano ao querer se relacionar com o mundo. Estes *modos* vão além das particularidades a qual são submetidas quem visualiza a imagem, sejam elas temporais ou sociais.

1. O *modo simbólico*. Inicialmente as imagens serviram de símbolos; para ser mais exata, de símbolos religiosos, vistos como capazes de dar acesso à esfera do sagrado pela manifestação mais ou menos direta de uma presença divina. [...] Na verdade, são inúmeros os exemplos em que a iconografia religiosa, figurativa ou não, é vasta e ainda atual. [...] Os simbolismos não são apenas religiosos, e a função simbólica das imagens sobreviveu muito à laicização das sociedades ocidentais, quando mais não seja para veicular os novos valores (a Democracia, o Progresso, a Liberdade etc.) associados às novas formas políticas.

2. O *modo epistêmico*. A imagem traz informações (visuais) sobre o mundo, que pode assim ser conhecido, inclusive em alguns de seus aspectos não-visuais. A natureza dessa informação varia (um mapa rodoviário, um cartão postal ilustrado, uma carta de baralho, um cartão de banco são imagens cujo valor informativo não é o mesmo), mas essa função geral de *conhecimento* foi também muito cedo atribuída às imagens. [...] Essa função foi

³ Referência complementar de SANTORO, André. 2002 / Que animais enxergam por meio de sons. (Publicado em 30 de abril de 2002, web: <https://super.abril.com.br/ciencia/que-animais-enxergam-por-meio-de-sons-e-como-eles-conseguem-fazer-isso>)

consideravelmente desenvolvida e ampliada desde o início da era moderna, com o aparecimento de gêneros "documentários" como a paisagem e o retrato.

3. O *modo estético*. A imagem é destinada a agradar seu espectador, a oferecer-lhe sensações (*aisthesis*) específicas. Esse desígnio é sem dúvida também antigo, embora seja quase impossível pronunciar-se sobre o que pode ter sido o sentimento estético em épocas muito distantes. [...] Seja como for, essa função da imagem é hoje indissociável, ou quase, da noção de arte, a ponto de se confundirem as duas, e a ponto de uma imagem que visa obter um efeito estético pode se fazer passar por imagem artística (na publicidade essa confusão atinge seu auge).

Leva-se em consideração essas "funções" entre os modos simbólico, epistêmico e estético neste trabalho, explorando principalmente os dois primeiros conceitos. A escolha de não tratar diretamente sobre o modo estético é por entender que o mesmo passa por uma interpretação sensorial e ligeiramente abstrata, ao passo que o objetivo dessa tese é discernir formas de descrever a imagem de maneira sistêmica e lógica. A compreensão de símbolos e padrões visuais são os primeiros passos para tornar possível a análise de formas que constroem uma imagem, e assim, estabelecer sua relação com o mundo.

a. Imagem e seu espectador

Para o professor universitário Jacques Aumont (2001), as imagens são feitas para serem vistas, por isso a importância de destacar o sentido da visão na compreensão da construção da imagem. Foi graças a essa afirmação que ele pôde constatar que o olho não é um instrumento neutro, o órgão, ao contrário, é um dos principais pontos de encontro entre o cérebro e o mundo, sendo induzido pela interpretação automática do sujeito que o utiliza para perceber uma imagem. Esse sujeito é nomeado como *espectador* pelo autor, possibilitando assim a compreensão separada entre o olho e quem olha.

Ao nomear um capítulo de seu livro como "o espectador constrói a imagem, a imagem constrói o espectador" (Aumont, 2001, p. 81), o teórico expressa que a abordagem da relação espectador-imagem consiste, antes de tudo, em tratar o sujeito como parceiro ativo da imagem, e consequentemente tornando-o agente emocional e cognitivo de sua interpretação.

A definição do espectador por Aumont é complexa, não haveria como ser diferente ao tentar descrever precisamente um sujeito, pois tudo do que este é formado afeta a sua relação

com a imagem. Com o objetivo de fazê-lo, o autor levanta pontos determinantes a serem levados em consideração na formação de um expectador: o saber, os afetos e as crenças que são vinculadas às vivências; já o tempo e as diferentes culturas são fatores que costumam ir além da capacidade de autopercepção do sujeito, mas que podem ressignificar a interpretação de uma imagem. Ao se refletir sobre constantes históricas e interculturais, quais seriam os elementos de afeto identificáveis por um espectador em uma imagem? E seria possível identificar alguns destes mesmos elementos no criador dessa imagem? É possível que a captura de informações neutras sejam feitas através de uma fotografia?

Jacques Aumont (2001) quando aborda qualquer artefato social, diz que o espectador sempre carrega consigo um saber (*arché*) sobre sua gênese, e conclui que mesmo que a fotografia seja uma imagem atemporal, o conhecimento do espectador sobre o tempo e seus elementos permanece atuante, o que pode então ser resgatado no processo de leitura da imagem.

Na obra *L'Art* (1950), Ernst Hans Josef Gombrich (1909-2001), austríaco e historiador da arte, propõe a expressão "papel do espectador" para descrever o conjunto de atos perceptivos e psíquicos pelos quais o sujeito faz existir a imagem, ao percebê-la e compreendê-la. Essa abordagem de Gombrich usa da ideia de esquemas perceptivos, onde a atividade do espectador, diante da imagem, consiste em utilizar todas as capacidades do sistema visual em organizar a realidade (partindo do princípio de elementos invariáveis do cotidiano, como já tratado neste trabalho) e relacionar com dados armazenados na memória sob forma esquemática. (Aumont, 2001).

Em resumo, o "papel do espectador" é extremamente ativo. Para Gombrich, o espectador necessita do "reconhecimento" para a construção visual e da "rememoração" para o emprego dos esquemas e da junção de um com o outro, com isso conseguindo construir uma visão coerente do conjunto da imagem. "Compreende-se porque esse papel do espectador é tão central para toda a teoria de Gombrich: é ele quem *faz* a imagem" (Aumont, 2001, p. 90), conclui o autor.

3. O imaginário de Sartre

O filósofo, escritor e crítico francês Jean-Paul Sartre (1905-1980), aborda a questão de maneira diferente. Através de sua perspectiva existencialista, em que separa a imaginação da percepção, ele considera a ideia de imagem ligada à ideia de consciência. Em suas palavras, "o

ato de imaginar... é um ato mágico. É uma encantação destinada a fazer aparecer o objeto pensado, a coisa desejada, para podermos nos apossar deles." (Sartre, 1996, p.166).

Segundo a concepção sartriana, a imaginação é uma capacidade mental que possibilita a representação de objetos por meio de suas qualidades transmitidas à mente, estas adquiridas por meio dos sentidos. Na filosofia tais qualidades transpostas em sentidos são chamadas de "qualidades secundárias", que ocorrem quando a ação do subconsciente pronuncia-se na ação da consciência.

Sartre (1996) trata em seu estudo de experimentar fenomenologicamente ações da consciência, assim como de experimentar a fenomenologia da imagem. Seu primeiro passo foi tentar expulsar os seguintes conteúdos da consciência para tornar possível captá-la em sua essência como puro ato: a sensação e a imagem; sendo a imagem, nesse caso, considerada um elemento psicológico analítico. Essa fenomenologia se estabelece em refletir tais imagens, descrevê-las e tentar determinar e classificar seus traços, sejam eles distintos ou semelhantes.

Uma vez estabelecida a consciência como puro ato, definida pela espontaneidade sem conteúdos inertes, Sartre propõe entendê-la dentro do conceito de *in-der-Welt-sein* ("ser-no-mundo" ou "estar-no-mundo"), formulado pelo filósofo alemão Martin Heidegger (1889-1976). Ao usar esse termo, ele sugere a definição da consciência como sujeito, objeto, consciência e mundo. Heidegger coloca o homem e o mundo como correlatos, não existe homem fora do mundo e não existe mundo fora da concepção do homem. Partindo de tal pressuposto, Sartre primeiramente evoca o mesmo sentido fenomenológico do matemático alemão Edmund Husserl (1859-1938), ao afirmar que "[...] toda consciência é consciência de alguma coisa" (Sartre, 1996, p.123) e depois que "[...] a imagem é um certo tipo de consciência. A imagem é um ato e não uma coisa. A imagem é consciência de alguma coisa." (Ibid., p. 137).

O crítico francês escolhe usar a palavra alemã *Bewusstsein*, o equivalente de "consciência", para aproveitar o sentido mais amplo que sua tradução germânica carrega. Com isso, pode-se diferenciar o seu significado conforme as estruturas em que aparece, por exemplo: em consciência perceptiva e consciência da imagem, a palavra carrega sentidos diferentes. Segundo a terminologia alemã, na percepção ela torna-se uma "certa consciência", e na imagem mental, ou objeto da imagem, ela é parte dessa consciência. Perde-se assim, a interpretação única de consciência enquanto um conjunto de estruturas psíquicas (Sartre, 1996).

A imagem mental se dá no fato de ser uma relação da consciência para com o objeto, ela é a organização sintética total. Ainda através da fenomenologia, Sartre (1996) analisa que a consciência imaginante que se tem de uma pessoa não é a mesma consciência da imagem, pois mesmo que tal pessoa seja diretamente atingida, a atenção não é focada para uma imagem, mas sim para um objeto. Dessa forma, o objeto da imagem não é em si mesmo a imagem; a imagem é a organização sintética total. Para o filósofo, a palavra "imagem" designa nada mais do que a relação da consciência ao objeto.

Ao perceber um objeto, existe um tempo e um espaço que influenciam nas maneiras possíveis do mesmo ser compreendido. Essa percepção ocorre espaçadamente, pois é necessária que seja absorvida a partir de variados ângulos, variados olhares, não há como acontecer de uma vez e ela seguirá em constante busca por sua essência. Já o objeto da imagem forma-se em um único ato, a concepção total do mesmo acontece de uma vez e não se relaciona com outras possibilidades, é essencialmente a consciência que se tem do objeto. Nas palavras do filósofo Igor Silva Aves, em seu artigo sobre consciência imaginante, existe "uma radical diferença entre pensamento e percepção e, por isso mesmo, não se pode reduzir um ao outro, ou seja, não se pode perceber um pensamento nem pensar uma percepção." (Alves, 2015).

Segundo Sartre (1996), os objetos da imagem só existem enquanto nós pensamos neles. Sendo assim, nos dois mundos que se estabelece, o imaginário e o real, suas constituições são dadas de maneira semelhante, variando apenas as interpretações e os agrupamentos em que acontecem. O que vem a definir as diferenças do mundo imaginário do mundo real é uma atitude puramente da consciência.

Tal atitude consciente pode ser dita representativa, no sentido de que vai buscar seu objeto terreno da percepção e de que visa os elementos sensíveis que a constituem, assim como acontece na consciência perceptiva em relação ao objeto observado. Por outro lado, a consciência da imagem pode ser espontânea e criadora, mesmo que ancorada em padrões concretos. Assim, pode-se presumir que sem o imaginário, o ser humano ficaria asfixiado pelo mundo, em essência, sem a possibilidade de ultrapassar a própria existência.

4. Imagem e Memória

Henri Bergson (1859-1941), filósofo e premiado pelo Nobel de Literatura, reflete em seu livro *Matéria e Memória* (1999) sobre como a matéria é essencialmente formada por imagens e em como a memória pode ser o ponto de convergência entre a anatomia, a fisiologia e a psicologia, além de também apresentar os conceitos de imagem-movimento e imagem-lembrança.

Antes de avançar em explicações acerca da memória, cabe as seguintes definições do autor: “chamo de matéria o conjunto das imagens, e de percepção da matéria essas mesmas imagens relacionadas à ação possível de uma certa imagem determinada” (Berson, 1999, p. 17).

Para entender mais profundamente o que os conceitos de imagem significam para Bergson, podemos seguir a análise feita pelo filósofo francês Gilles Deleuze (1925-1995):

Por que essa palavra “imagem”? É muito simples... A imagem é o que aparece. Denomina-se imagem aquilo que aparece. A filosofia sempre tem dito “o que aparece é o fenômeno”. O fenômeno, a imagem é o que aparece. Bergson nos diz então, que o que aparece está em movimento [...] (Deleuze, 1981, p. 5).

Ao interpretar o conceito de Bergson, a imagem-movimento para Deleuze não apresenta um contraponto entre as duas palavras, colocando a imagem presente apenas na consciência e o movimento nas coisas, como até então se entendia essa relação. Desse modo, a imagem não é consequência do movimento, o que existe são apenas as imagens-movimento, onde a imagem é o próprio movimento ao mesmo tempo que o movimento é também imagem. A unidade da experiência dessa relação é a imagem-movimento (Deleuze, 1981).

Já a imagem-lembrança é um recorte onde se assume que há um grande período de tempo passado, e que no momento em que é evocada, torna-se atualizada num presente sempre em devir, podendo inclusive se confundir com a percepção do real. Segundo Bergson:

[...] nossa lembrança continua em estado virtual; dispomos-nos assim apenas a recebê-la adotando a atitude apropriada. Pouco a pouco aparece

como que uma nebulosidade que se condensa; de virtual ela (lembrança) passa ao estado atual; e, à medida que seus contornos se desenham e sua superfície se colore, ela tende a imitar a percepção (Bergson, 1999, p. 156).

Bergson (1999) defende que a memória é influenciada, sobretudo, pelo afeto, pois quanto mais força emocional está presente em uma lembrança, mais fixada ela será dentro do ser. Porém, ao não possuir ferramentas que saibam diferenciar as infinitas sensações possíveis a serem causadas afetivamente em uma memória, a percepção corporal é considerada vulnerável.

Nota-se assim que a imagem-lembrança, ponto a ser explorado nesse contexto, é construída também por cunhos afetivos, segundo os franceses Bergman e Deleuze. Nessa perspectiva, a memória grava momentos afetivos em um estado virtual que corresponde não apenas aos fatores visuais, mas também a todos os outros sentidos sensoriais. São as sensações gravadas na memória que tornam-se a porta de entrada para a lembrança espontânea.

5. A imagem fotografada

O sociólogo, filósofo e semiólogo francês Roland Barthes (1915-1980) define as três práticas da fotografia como "*operator*, *spectator* e *spectrum*". Aquele que faz a imagem é o *operator*; o que olha a fotografia é o *spectator* e o *spectrum* é aquilo que envolve a fotografia, a sua história, "o retorno do morto". Para ele, a fotografia é como um ditado "vejo, sinto, portanto noto, olho e penso" (Barthes, 1989, p.39).

É interessante pensar o *modus operandi* de Barthes sobre a fotografia de maneira atualizada, principalmente ao relacionarmos-o com o reconhecimento de imagem por inteligência artificial feito atualmente, adotando assim uma nova perspectiva sobre os três conceitos. O *operator*, referido como fotógrafo, pode ser na contemporaneidade uma câmera de vigilância; o *spectator*, até então humano, agora pode ser um sistema de reconhecimento de imagem; o *spectrum*, o objeto contido e a sua história, agora podem ser dados e informações processados e reconhecidos por códigos, onde é possível extrair seus elementos principais de informação e com isso lhe dar um significado descritivo.

Por definição do filósofo francês, a fotografia tende à anulação do fluxo do tempo e à mortificação de um momento no próprio ato de criação da imagem, ao tornar-se representação do que registra (Barthes, 1989). Hoje, poderia-se dizer que essa tendência também afeta a captura de dados, pois o que se captura em um frame em movimento não se repetirá.

Para Barthes (1989), a fotografia reproduz ao infinito o que só ocorre uma vez: ela repete mecanicamente o que nunca mais poderá repetir-se existencialmente. Em uma foto, o acontecimento jamais se sobrepassa a outra coisa, um objeto fotografado sempre terá sua essência protegida. O sociólogo afirma que um objeto torna-se redundante dentro de uma fotografia, não podendo transformar-se do que é em si, ou mesmo do que era no momento em que foi fotografado, "um cachimbo, nela, é sempre um cachimbo, intransigentemente" (Barthes, 1989, p. 15). A fotografia enquanto captura do momento, sempre trará consigo seus dados referentes.

Quando vê-se uma foto não é exatamente ela que se enxerga, mas sim a imagem de um objeto que passou e por algum motivo foi capturado. A imagem retirada de uma fotografia é uma imitação de si própria, pois ao fotografar um alguém, cria-se ele como algo que não é um sujeito nem um objeto, mas sim um sujeito que está se tornando um objeto por ação do tempo.

Nas décadas em que viveram a maioria dos autores citados nesse trabalho, o ato da captura fotográfica continha uma tendência consciente no que diz respeito à escolha e ajuste de um espaço adequado para atingir o objetivo do fotógrafo, fossem eles de cunho documental ou artístico. Porém, é evidente que nem todos os registros aconteceram da maneira exata com que foram elaborados e infinitos fatores ainda podem alterar a imagem capturada durante o processo. Enquanto hoje, observa-se uma extrema facilidade com que se registram imagens através de milhares possibilidades de câmeras, permitindo uma aleatoriedade quase completa em seu ato, o que pode alterar a maneira com que se entende a execução da fotografia, porém não seu resultado como imagem a ser observada. É por isso que a seguinte afirmação de Barthes continua cabível: "a fotografia fotografa o notável mas logo, por uma inversão conhecida, ela decreta notável aquilo a que ela fotografa." (Barthes, 1989, pg 57). Ou seja, ainda que não de maneira consciente, ao fotografar algo, evidencia-o.

Segundo Barthes (1989), não existe imagem fotográfica sem algo ou alguém. No fundo, mesmo sendo objeto, uma foto sempre será invisível. A imagem dentro da fotografia é tão invisível quanto a de uma não-fotográfica, que neste caso refere-se às imagens que remetem à lembrança e que não necessitam de estímulos visuais para se concretizarem.

A imagem no mundo físico é, portanto, sempre perceptível através dos sentidos, porém quando essa mesma imagem é retratada em um papel, uma tela ou um pixel, ela torna-se imperceptível para todos os sentidos aquém da visão. Em outras palavras, originalmente uma superfície apenas consegue tornar acessível uma imagem pelo sentido visual, e esta é a questão central desta dissertação, a possibilidade de se conseguir acessar imagens por outro sentido que não o visual, mais especificamente, a criação de uma inteligência artificial que possibilite o reconhecimento e a descrição de imagens.

6. A imagem como comunicação e linguagem

A escritora e professora norte-americana Donis A. Dondis (1924-1984), aponta que no início do desenvolvimento da linguagem escrita, as palavras eram representadas por imagens, e quando estas não possuíam um correspondente direto, criavam-se símbolos que cumprissem essa função. Atualmente, na linguagem escrita já altamente desenvolvida usada no ocidente, as imagens foram abandonadas, e os sons, no ato da leitura, são representados exclusivamente por símbolos, a que chamamos de letras. A professora aponta que ao contrário da reprodução através de imagens, a utilização dos símbolos exige muito menos habilidade em seu aprendizado, comparando como exemplo o alfabeto inglês, formado por vinte e seis símbolos, e o chinês, que continua possuindo milhares de imagens-ideogramas (Dondis & Camargo, 2007).

Segundo Dondis & Camargo (2007), os dados visuais são compostos em três níveis distintos: o *input* visual, o material visual representacional e a estrutura abstrata. O *input* visual consiste na cópia e no sistemas de símbolos; o material visual representacional é identificável dentro do espaço e ambiente, e é reproduzido por meio da arte; a estrutura abstrata consiste na forma de um todo que é visto, seja natural ou seja resultado de uma composição. Portanto, a ideia de associar uma estrutura verbal e uma visual é perfeitamente compreensível seguindo esses três níveis de dados.

Cada modo destes listados (simbólico, representacional e abstracional) é composto por características que não se antagonizam, mas sim se sobrepõem e interagem mutuamente às suas qualidades. A percepção desses três pontos encontra-se no receptor das mensagens visuais, enquanto a interação deles "constituem, em conjunto, a força mais importante de toda comunicação visual, a anatomia da mensagem visual." (Dondis & Camargo, 2007, p. 67).

A nível de comparação, cabe retomar que a imagem na literatura gestaltista é compreendida através da percepção da mesma pelo espectador, onde o sujeito percebe e entende a imagem pelas próprias estruturas mentais. Essa ideia da compreensão da imagem para os gestaltistas, os teóricos da forma, é relacionada à percepção do mundo dentro de um processo de organização, onde existe uma ordem dos dados sensoriais que seguem as grandes categorias e as "leis" inatas do cérebro (Aumont, 2001). (cf. capítulo 1, 3.2.4)

Para Aumont (2001) são as teorias gestaltistas, revistas nas obras de Rudolf Arnheim (1904-2007), historiador da arte e psicólogo alemão, que desdobram essa concepção de modo sistemático em duas noções sugestivas, a de pensamento visual e a de centramento subjetivo. Segundo o autor, o pensamento visual parte, em conceito:

Ao lado do pensamento verbalizado, formado e manifestado pela mediação deste artefato humano que se chama linguagem, há espaço, segundo ele, para um modo de pensamento mais imediato, que não passa ou, pelo menos, não passa inteiramente pela linguagem, mas que se organiza, ao contrário, diretamente a partir dos perceptos dos nossos órgãos dos sentidos: o pensamento sensorial. (Aumont, 2001, p. 93)

Ainda seguindo as ideias de Aumont (2001), a noção de "pensamento visual" foi bastante difundida entre os estudiosos da época, porém, mais tarde, esbarrou com um problema de interpretação ambíguo, já que ainda não há maneira de comprovar que a linguagem não seja um fator na construção do pensamento visual. "Trata-se, pois, mais de uma facilidade de expressão que permite designar fenômenos em que a intervenção da linguagem é discreta ou não-localizada" (Aumont, 2001, p. 94), afirma o autor francês ao deixar claro que não existe concordância ou validação suficiente sobre esse conceito.

Segundo Aumont (2001), a noção de informação permitiu reescrever os princípios gestaltistas de modo geral, e com isso o englobar no *princípio do mínimo*: onde se em uma figura existem duas informações possíveis a serem determinadas, a mais simples é que será percebida de instância, pois sua descrição contém menos informação.

Já o conceito de *centramento subjetivo* parte da reflexão sobre a imagem baseada na ideia de Arnheim, onde o espectador tem uma concepção subjetivo-centrada do espaço ao seu redor.

Arnheim propôs descrever o espaço representativo definido por um centro, e essa ideia resulta da mesma concepção "indutiva" da relação do espectador com a imagem.

Sobre outro ponto de vista, o autor francês cita o cineasta soviético Sergei Eisenstein (1898-1948) e seu estudo na compreensão da imagem enquanto organismo, colocando assim as partes que a formam com equivalente importância em relação ao todo. Em uso de analogia à linguagem artística, para Aumont (2001) "uma obra de arte — em geral uma produção da mente — pode ser dita orgânica se nela a relação entre as partes for tão importante quanto as próprias partes: se ela 'parecer' um organismo natural" (p. 94).

Segundo Aumont (2001) o conceito de que a *imagem é estruturada como linguagem interior*, define-se na importância da linguagem como elemento fundamental na criação de uma obra artística. Para ele, essa ideia aparece em Eisenstein de diversas formas, pois o cineasta exercita a busca de um pensamento anterior à lógica, mais primitivo, tentando se aproximar à associação livre de ideias. Na visão do autor, essa "linguagem cinematográfica pode ser compreensível como manifestação de uma linguagem interior, que nada mais é do que outro nome do próprio pensamento." (Aumont, 2001, p. 95).

Aumont (2001) conclui que desse apanhado de diferentes abordagens em relação à imagem e quem a observa, o modelo de espectador em questão varia essencialmente se o seu foco está na leitura de uma imagem ou na produção dela. No caso da leitura da imagem, a tendência é de desenvolvimento de teorias analíticas que trabalhem a intelectualidade do sujeito. Já na produção da imagem, o desenvolvimento tende a ser mais heurístico, no sentido de buscar modelos antropológicos como pontos de partida que possibilitem a análise da imagem de maneira coerente com uma concepção do mundo verdadeira.

Por fim, tendo apresentado um entendimento de imagem que abrange diversos pensamentos, passa por diferentes séculos e discorre por variados autores, a proposta desta dissertação é partir desse ponto para se aprofundar na possibilidade de criação de uma inteligência artificial que compreenda os valores visíveis, ou não, de uma imagem, e com isso consiga torná-la acessível por outros sentidos humanos através da descrição da mesma – ainda que necessite revisão humana para tal. Esse feito possibilitaria não só uma amplitude no alcance de uma imagem, como também demonstraria a potencialidade da tecnologia nessa discussão artística e filosófica.

Capítulo II: Sobre Acessibilidade

1. Contexto

O termo acessibilidade contém múltiplas interpretações e sua definição varia de acordo com o contexto em que é aplicado. Todas as relações de acesso que afetam pessoas podem ser consideradas meios de acessibilidade, sendo estes avaliados por um conjunto de condições de igualdade e segurança para todos. Hoje entendemos a acessibilidade também como um requisito básico de usabilidade⁴ na utilização de espaços físicos e digitais.

Segundo os cientistas da computação Jakob Nielsen e Hoa Loranger (2007), em seu livro Usabilidade na Web, a definição do termo usabilidade é um atributo relacionado à facilidade de uso de algo, à rapidez com que as pessoas podem aprender a usar um produto, à eficiência que possui, ao quão memoráveis são seus padrões, ao quão propenso a erros e ao quanto os usuários gostam de usá-lo. Em última instância, para Nielsen e Loranger (2007), "se as pessoas não puderem ou não usarem um recurso, ele também pode não existir" (p. XVI).

Laura Kalbag (2017), autora, designer e palestrante britânica, define a acessibilidade no mundo físico como o grau em que um espaço seja utilizável por tantas pessoas quanto possível. Assim como a acessibilidade no mundo digital, que também é medida pelo grau em que o espaço pode ser usado pela maioria das pessoas. Podemos pensar que a acessibilidade⁵ como um requisito de usabilidade é, portanto, uma porta para a inclusão.

Segundo Kat Holmes, ex diretora de Design Inclusivo na Microsoft e autora reconhecida internacionalmente por sua pesquisa que une design e tecnologia, argumenta que "*Designing for inclusion starts with recognizing exclusion.*"⁶ (Holmes, 2018). Fazendo um paralelo, essa frase poderia ser aplicada também em questões de inclusão para qualquer outro espaço ou sobre

⁴ Usabilidade classifica-se sobre a extensão na qual um sistema, produto ou serviço pode ser usado por usuários especificados para atingir metas especificadas com eficácia, eficiência e satisfação em um contexto especificado de uso. Adaptado da ISO 9241-11: 1998.

⁵ Acessibilidade classifica-se (em sistemas interativos) como uma usabilidade de um produto, serviço, ambiente ou instalação por pessoas com a mais ampla gama de recursos. Adaptado da ISO 9241-171.

⁶ Tradução nossa para "o design para inclusão inicia-se reconhecendo a exclusão" (Holmes, Kat [2018-09-20T22:58:59]. Mismatch [Simplicity: Design, Technology, Business, Life] . The MIT Press). Ebook Kindle.

qualquer outro grupo excluído, pois toda inclusão requer desconstrução do que está estabelecido, caso contrário, não haveria necessidade de se pensar em acessibilidade do que já é acessível.

Algumas das barreiras que a acessibilidade enfrenta são: a comunicação, a mobilidade, a compreensão e o uso. A professora brasileira e pós-doutora em Ciências Sociais, Carla Cristina Garcia (2008), em seu livro *Sociologia da Acessibilidade*, define tais barreiras de acesso como:

- **Comunicação:** um processo no qual há troca de informações entre duas ou mais pessoas, de maneira direta ou indireta, no emite-se ou recebe-se por meio de linguagens;
- **Mobilidade:** a ação que implica no deslocamento de uma pessoa a um ponto a outro, com a apreensão de objetos por meios próprios ou mediante ajuda extra ou meio de transporte;
- **Compreensão:** é a capacidade de entender informações que se recebe durante uma comunicação, é o processo no qual o receptor entende o significado da informação que foi transmitido pelo emissor. Vale ressaltar que comunicação não é sinônimo de compreensão, portanto, se uma mensagem é emitida porém não compreendida pelo receptor, a comunicação é parcial, onde a informação circula porém não pode ser utilizada;
- **Uso:** refere-se a utilização, manipulação e interação entre sujeitos e objetos ou elementos no espaço físico ou digital. A utilização eficiente de um espaço, produto ou serviço implica no fato de poder ser utilizado por todas as pessoas independentemente de suas características físicas ou sociais.

Tendo em mente estes quatro fatores apresentados, pode-se pensar na acessibilidade como um conceito que busca com que estes elementos estejam presentes nos espaços, tornando-os assim, locais à disposição de todos. Quando qualquer um desses aspectos é eliminado, é muito provável que sua ausência provocará a exclusão de certos grupos de pessoas.

2. Foco de Acessibilidade

*Se você não pode considerar todas as deficiências, considere a cegueira.
Cerca de 80% dos problemas de acessibilidade estão relacionados à
cegueira. —Regine Gilbert*

A designer e educadora Regine M. Gilbert (2019), em seu livro *Inclusive Design for a Digital World* (Design Inclusivo Para um Mundo Digital, tradução nossa), ressalta que ao abordar a acessibilidade no design é necessário ter em mente os campos de sentidos relacionados: visual, auditivo, cognitivo e motor.

Tem-se, neste trabalho, o foco em discutir e refletir a acessibilidade voltada para pessoas com deficiência visual, a qual abrange um enorme espectro de deficiências relacionadas à visão e gera uma ampla gama de diferentes necessidades. Não existe uma solução simples e única, é necessário levar-se em consideração, além do sentido afetado, as particularidades de cada deficiência, distúrbio ou condição desenvolvida.

Segundo Kalbag (2017), os problemas com a visão não se concentram apenas na função de nossos olhos, mas também em como nossos cérebros percebem o que nossos olhos enxergam; dando ênfase indiretamente, ao que Aumont já discorreu sobre o tema. Ela aponta que condições como miopia, hipermetropia e astigmatismo são muito comuns em pessoas de todas as idades e tendem a se agravarem com o passar do tempo. Se acompanhadas minuciosamente, observa-se que mesmo essas condições mais comuns podem variar ao longo de um mesmo dia e, ainda mais possivelmente, no decorrer de um dia para o outro. Esse fator é importante ao se considerar que uma deficiência, ou a ausência dela, não significa uma característica necessariamente imutável, ela pode se modificar e junto dela modificar também as necessidades pessoais do indivíduo.

3. Design de Inclusão

*O Design Inclusivo é uma metodologia, nascida de ambientes digitais,
que possibilita e se baseia em toda a diversidade humana. Mais
importante, isso significa incluir e aprender com pessoas com várias
perspectivas. —Microsoft*

A acessibilidade não pode ser confundida com usabilidade, a qual é focada em fazer com

que o uso de um produto, serviço ou espaço seja utilizado por usuários específicos para realizar tarefas com eficiência e satisfação dentro do contexto. A acessibilidade deve estar integralmente relacionada ao design universal, no qual o processo de utilização de um produto, serviço ou espaço possa ser operado e usado por todos. Essa relação entre a acessibilidade e o design universal coloca seus usuários em igualdade, independente de possuírem qualquer deficiência.

Ao se levar em consideração a estrutura estabelecida e pouco pensada para acessibilidade no design, para Holmes (2018) "é muito mais provável que o design seja a fonte de exclusão do que de inclusão". Vem daí a necessidade de repensar a maneira com que se cria nesse meio, pois ao mesmo tempo que o design possui a capacidade de englobar todas as pessoas, ele mesmo pode excluí-las.

Designers são constantemente influenciados pelo próprio *bias* ao criar uma solução que possam ver, tocar, entender ou ouvir. O designer ao criar uma solução sobre a própria perspectiva, sem uma pesquisa de usuário envolvida, tende a funcionar somente para pessoas com circunstâncias ou preferências semelhantes às que criaram o produto (Gilbert, 2019).

A grande necessidade de pesquisa relativa à acessibilidade e inclusão é o ponto primordial que afeta o design de alguns produtos e serviços ao excluir uma grande parcela de pessoas. Esse ponto é ainda mais evidente ao levar-se em consideração quem faz parte da parcela da população que tem acesso ao estudo formal, à possibilidade de trabalho profissional na área e, por consequência, detêm o controle da produção e comunicação em grande escala. Existe, de fato, preocupação partindo dessa parcela da população, com a acessibilidade?

Por exemplo, podemos lembrar situações em que o design é falho, como uma porta que não abre para o lado correto, um *mouse* de computador que não funciona para pessoas canhotas, um sistema de pagamento com touchscreen e sem comando de voz, máquinas de venda automáticas que não aceitam múltiplas possibilidades de pagamento, uma cadeira que não comporta pessoas obesas, um carro que exige uma altura mínima, um website em apenas uma linguagem e sem transcrição auditiva ou sem conexão com um screen reader, etc. Diariamente pessoas sentem-se excluídas por situações como essas e, quando designers e desenvolvedores não dão a devida atenção a estes "detalhes", isso acaba por comprovar na prática a maneira com que a sociedade constrói o mundo sem pensar na inclusão (Holmes, 2018).

É possível acreditar que a grande maioria das pessoas já se sentiu excluída em algum momento, e conforme um indivíduo mais se afasta do que é socialmente considerado o padrão

normativo, com maior regularidade e força essa exclusão acontecerá. Esse fato, pode ser observado a partir das condições financeiras de cada um, do recorte de raça, das pluralidades de expressão de gênero e das diferentes disposições físicas e psíquicas, que podem envolver algum tipo de deficiência. A inclusão aparece sobre vários formatos, mas exclusão tem apenas um significado para a autora:

Pergunte a cem pessoas o que significa inclusão e você receberá centenas de respostas diferentes. Pergunte às mesmas pessoas o que significa ser excluído e a resposta será claramente uniforme: é quando você fica de fora. (Holmes, 2018)

Para criar a acessibilidade através do design universal, a empatia é o primeiro passo. Segundo a designer Kalbag (2017), é isso que torna designers bons na criação de produtos para outras pessoas, pois pode-se entender melhor seus problemas e criar soluções que atendam às suas necessidades. Leva-se em consideração neste ponto, que o designer nem sempre será o público alvo do que produz.

Em um amplo aspecto, Kalbag (2017) considera que o design universal é o melhor caminho para inclusão – enquanto o design acessível cria produtos que podem ser utilizados por pessoas com deficiência – o design universal cria produtos para o público mais amplo possível, que inclui, mas não se limita, às pessoas com deficiência.

As autoras do livro *A Web For Everyone*, Sarah Horton e Whitney Quesenbery (2013), que trabalham com design interativo e como especialista em usabilidade, respectivamente, defendem que as decisões de design tomadas em nome da acessibilidade geralmente beneficiam a todos, pois toda tecnologia é auxiliar, e não substitutiva.

Portanto, a porta de entrada para a universalidade no design está na empatia com o outro e também na realização de pesquisas, mas sobretudo na flexibilização natural. As pesquisas e testes ajudam a entender onde é necessário utilizar o conhecimento do design e da tecnologia no aprimoramento da inclusão, gerando independência e liberdade, porém esses conhecimentos não são pilares exclusivos do processo, é necessário o constante aprimoramento, empatia e aproximação com os mais diversos grupos de pessoas.

4. Liberdade e independência

As pesquisas e entrevistas⁷ deste projeto tornaram evidente que a utilização da palavra "ajudar" é inapropriada para este contexto – o termo escolhido mais adequado é "auxiliar", "dar liberdade" e criar "acesso à independência". Após os *feedbacks*, é notável que a persona a que se destina esse projeto não necessita de ajuda por meio da acessibilidade e de tecnologias assistivas, mas sim de acesso ao mundo através dessas possibilidades e estruturas que servem como meio auxiliar.

Em entrevista autoral feita com Karthik Mahadevan, CEO da Envision⁸, ele compartilha uma história que ocorreu enquanto ministrava um workshop de design para crianças na Índia, em uma escola para pessoas com deficiência visual. Ao explicar que designer é alguém que resolve problemas, Mahadevan sugeriu a seguinte pergunta aos estudantes: "se vocês se tornassem designers amanhã, qual seria o problema que gostariam de resolver?". A maioria das crianças então demonstraram seguir uma linha de raciocínio parecida, ao relacionar a resolução de um problema com uma ideia de liberdade. Elas deram respostas como: "quero ser mais independente; quero ir para a escola de maneira diferente; quero sair com meus amigos; quero ler um livro sozinha" (Mahadevan, 2019).

Mahadevan (idem) defende que a liberdade e a independência significam, principalmente, igualdade de acesso à informação, e que se for possível chegar ao momento em que todas as pessoas alcancem o mesmo nível de acesso, isso representaria oportunidades iguais de verdade.

5. Tecnologias Assistivas

Tecnologia assistiva são recursos e serviços que visam facilitar o desenvolvimento de atividades diárias por pessoas com deficiência. Procuram aumentar as capacidades funcionais e assim promover a independência e a autonomia de quem as utiliza.
—Deise Tallarico Pupo, Amanda Meincke Melo, Sofia Pérez Ferrés.

⁷ Pesquisas e entrevistas disponíveis em: <https://www.imagible.app/resources>

⁸ Envision é um aplicativo para smartphones que empodera usuários com deficiência visual para serem independentes reconhecendo mundo visual ao seu redor. O aplicativo utiliza o reconhecimento de imagem para descrever imagens capturadas nas mais diversas situações cotidianas. Disponível em <https://www.letsenvision.com/>

A tecnologia assistiva, ou abreviadamente definida pela sigla TA, é um termo criado em 1986 pela legislação americana, fazendo parte de um conjunto de leis que compõem a *ADA – American with Disabilities Act* (Lei dos Americanos com Deficiência), a qual visa regular os direitos dos cidadãos portadores de deficiência nos Estados Unidos da América.

Regine Gilbert (2019) apresenta a tecnologia assistiva como um termo amplo que inclui dispositivos de assistência, adaptação e reabilitação para pessoas com deficiência. Segundo a autora norte-americana, essa tecnologia é um fator de grande proporção para atingir uma maior independência para as pessoas, ao serem possibilitadas de executar tarefas que anteriormente eram impedidas ou tinham grande dificuldade em realizar. Os produtos de TA são projetados com métodos aprimorados, ou alternativos, para interação facilitada com as pessoas e oferecem acessibilidade adicional a indivíduos com capacidades reduzidas ou deficiências.

A tecnologia assistiva não é restrita apenas ao mundo digital, como computadores e smartphones, mas também é encontrada em produtos e espaços físicos. Em alguns casos esses produtos e espaços passam apenas por pequenas modificações, porém em outros necessitam da TA para se tornarem realmente acessíveis.

Gilbert (2019) salienta que existem muitas pessoas que não percebem, ou sequer possuem a consciência, de que utilizam tecnologias assistivas dentro do seu próprio cotidiano, como por exemplo, ao usar óculos e lentes de contato, que são considerados produtos assistivos tanto quanto aparelhos auditivos, cadeiras de rodas, bengalas, mapas táteis etc.

Os serviços disponibilizados através de TA não estão apenas relacionados ao uso restrito de pessoas com deficiências, eles estão atrelados ao benefício de todos, como acontece no uso de legendas e áudio descrições, que hoje em dia são opções básicas exigidas em qualquer produto audiovisual.

Outros *softwares* específicos para pessoas com baixa visão e deficiência visual, como *Screen Readers* (que serão apresentados no próximo subcapítulo) e ampliadores de telas, ajudam a criar um mundo digital cada vez mais acessível.

[...]Sara Hendren nos lembra que “todas as pessoas, ao longo de suas vidas, trafegam entre tempos de relativa independência e dependência.” E Hendren pergunta: 'Que tecnologia você está usando que não ajuda?'.

Nossos teclados e mouses nos ajudam a nos comunicar com um computador. Nossos fones de ouvido nos permitem ouvir áudio em nossos próprios espaços, sem perturbar os que estão ao nosso redor. Nossos telefones nos dão o conhecimento de toda a web em nossos bolsos. A tecnologia permite a todos nós e pode nos proporcionar uma melhor experiência do mundo ao nosso redor. (Horton & Quesenbery, 2013).

Ao citarem Sara Hendren, artista e pesquisadora em design, as autoras Sarah Horton e Whitney Quesenbery (2013) apresentam os pontos que devem ser focados para melhorar a experiência de todos no que diz respeito à usabilidade digital. Eles são divididos em quatro parâmetros amplos: o primeiro é o visual, que consiste em facilitar a visualização; o segundo é o auditivo, sobre facilitar a audição; o terceiro é o motor, para facilitar a interação; e o quarto é o parâmetro cognitivo, sobre facilitar o entendimento.

Referindo-se às tecnologias assistivas, o conceito de design acessível foca em tornar espaços acessíveis para pessoas com deficiência, ou tornar o acesso possível através de tecnologias acessíveis, o qual pode trazer benefícios para todos.

Ainda segundo as autoras, para um designer, a possibilidade de criar experiências acessíveis pode ser responsável por abrir sua mente ao confrontar-se com situações que não saberia como resolver anteriormente. Os princípios de acessibilidade e usabilidade juntos à tecnologia assistiva, dão às pessoas a capacidade de acessar informações e destituir barreiras que impedem o benefício de todos igualmente, facilitando assim o uso de produtos, serviços e espaços. A tecnologia é um direito de todos, e o papel do designer é projetá-la com inclusão e acessibilidade em mente.

No livro *Application of Assistive Technology*, os professores brasileiros Maria Lucia Miyake Okumura, Marcelo Rudek e Osiris Canciglieri Jr. (2013) falam sobre o processo de criação das tecnologias assistivas e a maneira como se devem compreender os aspectos de design universal, ergonomia e usabilidade, para assim atender o maior número de pessoas, levando em consideração o tipo de persona, atividade e cenário envolvido.

Hoje a classificação de produtos de TA divide-se por uso personalizado, uso individualizado, uso em grupo e uso na diversidade. Sendo o uso personalizado exemplificado como uma prótese, o uso individualizado como um aparelho celular ou um andador, a utilização

em grupo como uma bola com guizo e uso na diversidade como material didático (Okumura, Rudek, Canciglieri, 2013).

Portanto, é visível que os recursos de TA são muito variados, podendo ser simples artefatos e objetos já mencionados, como também podem ter uma elevada complexidade, como um software especial de leitura.

Atualmente estima-se centenas de tecnologias assistivas registradas e projetadas, nas quais incluem desde diferentes cadeiras de rodas até inúmeras ferramentas de alta tecnologia. Todos recursos de TA são organizados ou classificados de acordo com os objetivos funcionais a que se destinam. Ao longo do tempo torna-se cada vez mais evidente a importância social de projetar e desenvolver tecnologias de inclusão e acessibilidade, como também a valorização de mercado que possuem.

Pode-se observar avanços na área da informação e da comunicação que representam um grande passo no processo da acessibilidade nos dias atuais. Essa evolução da tecnologia tende a crescer exponencialmente, tanto quanto a área da comunicação – o que deve ser altamente levado em consideração, pois cada vez mais elaboraram-se possibilidades tecnológicas como instrumentos de inclusão e interação entre o planeta inteiro.

É evidente que o acesso à tecnologia vem-se difundindo rapidamente nos países ocidentais, em contraste com poucas décadas atrás onde ele era completamente restrito somente àqueles com grande poder monetário. Tais tecnologias estão hoje presentes no dia a dia, podendo serem utilizadas nos mais variados espaços públicos e privados, aumentando a capacidade de cada vez mais usuários com deficiência incluírem-se em diferentes esferas do cotidiano. O objetivo de criar tecnologias inclusivas é também conseguir abranger todas as classes sociais, pois mesmo nesta crescente evolução, os produtos mais modernos ainda são inacessíveis à grande maioria.

As tecnologias assistivas podem dar às pessoas acesso ao mundo digital que de outra forma não o conseguiriam, mas elas só podem cumprir esse papel até certo ponto. A partir do seguinte exemplo, Kalbag (2017) explica que é possível facilitar o acesso aos reprodutores de vídeo usando tecnologias assistivas, mas o conteúdo do vídeo só poderá ser totalmente acessível para pessoas com deficiência auditiva, se ele estiver equipado com legendas ou transcrições e, para pessoas com deficiência visual, com descrição de áudio. Ou seja, a inclusão só funcionará

completamente se partir das empresas e produtoras juntamente com seus produtores de conteúdo. TAs são importantes tanto quanto o esforço coletivo de inclusão.

Hoje a tecnologia assistiva está no meio da comunicação de muitos dispositivos e inclui softwares como *speech-to-text*, que traduzem a voz em áudio para texto descritivo; *screen readers*, que auxiliam pessoas com deficiência visual a conseguirem navegar em computadores e smartphones; assim como *pointing light*, *touchpads* e muitos outros.

Em Portugal, a definição das TAs é dada de acordo com o conceito proposto pelo Comitê de Ajudas Técnicas (CAT) da Secretaria de Direitos Humanos da Presidência da República:

Tecnologia assistiva é uma área do conhecimento, de característica interdisciplinar, que engloba produtos, recursos, metodologias, estratégias, práticas e serviços que objetivam promover a funcionalidade, relacionada à atividade e participação de pessoas com deficiência, incapacidades ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social. (CAT, Ata da Reunião VII, SDH/PR, 2007).

6. Screen Readers

Os Screen Readers são o foco principal das tecnologias assistivas para pessoas com deficiência visual e com baixa visão. Segundo a inglesa Laura Kalbag (2017), o símbolo da acessibilidade na web para pessoas com deficiência visual são *screen readers* como o *VoiceOver*⁹, que funcionam como uma peça de TA que lê o conteúdo de uma tela e o transmite uma áudio descrição por saída de áudio ou por dispositivo de saída em braille.

Regine Gilbert (2019) conta que para a Federação Americana para Cegos (AFB), em uníssono ao que apresenta Kalbag, os *screen readers* são programas de *software* que permitem que usuários com deficiência visual escutem o texto exibido na tela através de um sintetizador de

⁹ O leitor de tela VoiceOver da Apple foi revolucionário para usuários de computadores e dispositivos Apple. Antes do Mac OS X 10.4 introduzir esse recurso, os leitores de tela eram geralmente softwares independentes que podiam ser instalados em um computador. VoiceOver, no entanto, inclui um leitor de tela como uma parte essencial do sistema operacional em cada dispositivo. (KALBAG, 2017).

voz ou display em braille. Um *screen reader* é a interface entre o sistema operacional do computador, seus aplicativos e o usuário.

Os leitores de tela funcionam de forma que um comando pode instruir o sintetizador a ler ou soletrar uma palavra, uma linha ou o texto inteiro que aparece na tela, encontrar uma sequência de texto, anunciar a localização do cursor ou do item que está em foco no computador e assim por diante. Além disso, ele permite que os usuários executem funções mais avançadas como localizar um texto exibido em uma determinada cor, ler partes pré-designadas da tela, ler textos destacados e identificar a opção ativa em um menu. Os usuários também têm a possibilidade de usar o verificador ortográfico em um processador de texto e ler células específicas de um *spreadsheet*.

Segundo Gilbert (2019), algumas tecnologias assistivas são incorporadas aos dispositivos, enquanto outras são tecnologias independentes que funcionam junto com os navegadores web, que podem, opcionalmente, incorporar a acessibilidade. Três dos mais populares Screen Readers atualmente são o *JAWS*, o *VoiceOver* e o *NVDA*.

O software *JAWS* (Job Access With Speech) é o leitor de tela mais popular do mundo, desenvolvido para usuários de computadores cuja perda de visão os impede de ver o conteúdo da tela ou navegar com o mouse. Já o *VoiceOver*, criado pela Apple, foi revolucionário para usuários com deficiência visual e pode ser utilizado em computadores e dispositivos como smartphones. Antes do Mac OS X 10.4 apresentar esse recurso, os leitores de tela eram geralmente softwares independentes que poderiam ser instalados em um computador. O *VoiceOver*, no entanto, incluiu um leitor de tela como parte principal do sistema operacional e está presente em todos os dispositivos da marca. A Microsoft possui um leitor de tela chamado *Narrator*, semelhante ao *VoiceOver* da Apple, que recentemente foi aprimorado principalmente na usabilidade e no tempo de resposta de uso.

Nos últimos anos surgiram também alguns leitores de tela gratuitos e de código aberto, como o *NVDA* (NonVisual Desktop Access) para *Windows*, e também o *LSR* (Linux Screen Reader) e o *Orca* para *Linux*. Todos os aplicativos listados podem ser ativados de maneira rápida e eficaz, fornecendo acesso instantâneo a qualquer pessoa que precise de um leitor de tela (Gilbert, 2019).

7. Imagens invisíveis e *alt-text*

Para se entender o que é uma imagem invisível na web, é necessário entender como as imagens são traduzidas em informações digitais. Como já apresentado anteriormente neste trabalho, o filósofo Aumont (2001) defende que (no "ambiente cotidiano") as imagens que vemos são informações visuais unicamente extraídas do globo ocular, que ainda não possuem decodificação ou construção de percepções a seu respeito.

Portanto, com base na ideia de Aumont sobre a informação que é extraída da imagem, podemos criar a questão: o que é a imagem invisível? Para refletir sobre esse conceito, é necessário separar os dois principais cenários em que ele se aplica, o cotidiano e o digital.

Para pessoas com visão 90% de tudo que é processado pelo cérebro são informações visuais¹⁰, mas para uma pessoa com deficiência visual essa porcentagem não se aplica, pois a imagem cotidiana de uma pessoa com baixa, ou nenhuma visão, é meramente não-visual. Porém, o fato de uma imagem ser não-visual, não significa que ela é invisível, sua informação ainda pode ser extraída e compreendida de diferentes maneiras, usando de ferramentas sensoriais e palpáveis. A imagem do cotidiano só se torna invisível neste caso, quando tais pessoas com deficiências visuais também não conseguem acessar as informações através de outros sentidos.

Enquanto que no cenário digital não se possui a possibilidade do palpável, uma imagem torna-se visível para este determinado grupo de pessoas apenas se houver a presença de um texto alternativo (*Alt-text*¹¹). Caso contrário, todo conteúdo exclusivamente visual torna-se invisível. Conclui-se então, que a imagem invisível no meio digital surge quando não há qualquer descrição de texto ou inteligência descritiva, com ou sem revisão humana, de imagens e de objetos visuais.

O texto alternativo é usado no meio digital por pessoas que utilizam *Screen Readers*, sendo o *alt-text* essencial para a compreensão de imagens visuais por estas pessoas que possuem uma deficiência visual, baixa visão ou até mesmo pessoas que não conseguem acessar o carregamento de página web por completo. As imagens transmitem informações importantes e normalmente não possuem descrição apropriada nas páginas da web, fazendo assim com que as pessoas que não tem acesso a elas sejam frequentemente excluídas da comunicação transmitida.

¹⁰ Estatística retirada de: <https://www.t-sciences.com/news/humans-process-visual-data-better>

¹¹ O símbolo **Alt** é uma descrição de imagem que consiste em várias palavras, lidas pelo rastreador quando a página é indexada. Se o usuário desativou a função de suporte gráfico no navegador, então o conteúdo da **tag** será exibido ao invés da imagem.

Essa compreensão apresenta que o desenvolvimento de textos alternativos (*alt-text*) e as descrições de imagens no meio digital são extremamente necessários para cumprir os requisitos de acessibilidade em quaisquer espaços virtuais.

De acordo com a descrição feita em 2019 pela W3C, escrita por Shadi Abou Zahra enquanto representante da *Web Accessibility Initiative – Ageing Education and Harmonization* (WAI-AGE), as propriedades de *alt-text* são equivalentes ao conteúdo não textual, elas podem, por exemplo, ser criadas para imagens tanto quanto para ícones, botões e gráficos. Neste último caso, são ainda capazes de descrever os dados representados, diagramas e ilustrações. Tais propriedades podem incluir descrições de arquivos de áudio e vídeo, como também servem para ajudar no controle de formulários, entradas e diversos outros componentes da interface do usuário.

O objetivo do *alt-text*, ou texto alternativo, é transmitir claramente as informações de uma imagem ou função, para assim fornecer uma experiência de usuário equivalente a quem não necessita deles. Por exemplo, uma possibilidade de texto apropriado para um botão de pesquisa seria "pesquisar", ao invés de "lente de aumento" que geralmente é o símbolo utilizado na representação desta função. Normalmente, não se vê o texto alternativo explícito em uma página web, ele costuma ficar presente apenas no código da mesma, como demonstrado no exemplo a seguir:

```
<img src = "imagem.png" alt="Aqui é onde fica o texto alternativo">
```

Toda imagem e conteúdo visual inclusivo deve possuir a propriedade *alt-text*, que pode ser apresentada de várias maneiras. É ela que torna possível a leitura de programas assistivos como o *VoiceOver* ou a exibição em dispositivos braille, como explicado anteriormente.

Os textos alternativos servem como guias no auxílio à navegação geral, à navegação do teclado e ao reconhecimento de voz. Eles também atuam como rótulos para identificar a localização de áudios, vídeos e arquivos em outros formatos, bem como de aplicativos incorporados como parte de um site (*W3C World Wide Web Consortium: WCAG2.0 HTML Techniques*).

Quando um leitor de tela se depara com uma imagem que não possui atributo *alt-text*, na tentativa de fornecer informações úteis, ele em alguns casos lê o nome do arquivo. Não é de

surpreender que quando as imagens têm nomes de arquivos digitais sem especificidade, como “photo.jpg” ou “PIC0013.png”, elas não sejam compreendidas. Fornecer alternativas de nomenclaturas para arquivos de imagens pode ajudar alguém a entender o contexto da mesma sem necessariamente vê-la (Kalbag, 2017).

Kalbag (idem) explica que para a criação do texto alternativo não é necessário utilizar descrições excessivas, mas elas também não podem ser banais. Um *alt-text* inadequado para uma imagem pode ser algo como “um carro azul”, enquanto descrever a imagem como “um carro azul estacionado em um parque ao lado de uma árvore” daria muito mais contexto e significado à imagem. É necessário ter consciência da informação e do conteúdo que se extrai de uma imagem visual para sua descrição, fazendo assim com que todos possam entendê-la.

8. Audiodescrição

Quando se fala em descrição de texto ou textos alternativos, é imprescindível não abordar o tema da audiodescrição. A audiodescrição não está presente apenas no mundo digital, como também engloba multifaces da comunicação e da cultura, presente no cinema, na televisão, na dança, em óperas, nas artes visuais e outros.

Hoje a audiodescrição nestes locais é feita de uma maneira muito humanizada, onde há um narrador durante a cena que descreve o que está acontecendo no momento durante as pausas naturais do áudio de um filme, por exemplo. A descrição feita por um narrador é uma forma de tradução, que torna o conteúdo acessível a milhões de pessoas.

Ainda que hoje existam tecnologias e inteligências artificiais para criação de legenda, toda audiodescrição deve ser revisada por humanos, seguindo a ideia de Kalbag (2017) que é necessário ter a consciência da informação e do conteúdo que se extrai de uma imagem visual para sua descrição. Conclui-se que a intervenção ou pelo menos a revisão de tais descrições devem ser feitas por uma pessoa, esse passo é de extrema importância para garantir que a audiodescrição seja neutra e correta.

A audiodescrição é uma profissão relativamente nova, no mundo iniciou-se pelos anos 70 e no Brasil só após os anos 2000, e em entrevista autoral (2020) com a audiodescritora Lívia

Motta, criadora do projeto Ver com Palavras¹², vê a tecnologia como uma aliada para área de audiodescrição. Segundo Livia, a revolução tecnológica trouxe mais inclusão e independência para as pessoas com deficiência visual, e os leitores de tela tiveram papel fundamental nesse momento.

A audiodescrição é um recurso de acessibilidade que amplia o entendimento das pessoas com deficiência visual em eventos culturais, gravados ou ao vivo, como: peças de teatro, programas de TV, exposições, mostras, musicais, óperas, desfiles e espetáculos de dança; eventos turísticos, esportivos, pedagógicos e científicos tais como aulas, seminários, congressos, palestras, feiras e outros, por meio de informação sonora. (Livia Motta, 2020)

Livia afirma que, nos dias de hoje, por causa da evolução tecnológica os profissionais da área podem focar nas suas habilidades essenciais, que são as escolhas tradutórias, e deixar a máquina fazer o trabalho de máquina - auxiliando na substituição de palavras, de memória, otimizando o tempo e o trabalho dos audiodescritores. (Entrevista autoral, Livia Motta, 2020)

Livia (2020) também afirma que ferramentas de reconhecimento de imagem podem ajudar principalmente a pesquisa de audiodescritores, no qual a inteligência pode auxiliar a diminuir os trabalhos manuais, como o reconhecimento de imagens textuais como de letras e palavras hoje já consegue auxiliar a digitalização de documentos e na tradução dos mesmos. Segundo ela, o reconhecimento de objetos pode chegar lá também, para isso é necessário bancos com muitas imagens, assim como qualquer pesquisa necessita de um *corpus* amplo suficiente para poder definir um padrão de itens, pois um banco limitado ou um *corpus* de pesquisa pequeno gera um resultado inconclusivo e também pode limitar a identificação de um objeto. E isso também aplica para as tecnologias de inclusão.

O impacto que a tecnologia pode trazer é trazer a amplitude dos conhecimentos, do pensamento, para poder tomar decisões mais ativas sobre como as coisas funcionam - e menos do

¹² Projeto Ver com Palavras é um espaço virtual que pretende colaborar com publicação de artigos, notícias, posts, informações sobre audiodescrição no Brasil. A empresa do projeto oferece serviços de audiodescrição para todos os tipos de espetáculos, eventos acadêmicos, corporativos e sociais, produtos audiovisuais, audioguias para museus, audiodescrição de imagens estáticas para publicações e sites, além de cursos de formação de audiodescritores. Acessado e retirado de <http://vercompalavras.com.br/>

trabalho manual.

Isso muda a humanidade, a tecnologia muda a humanidade e pode deixar a gente mais focado em fazer mais funções intelectuais, de pensamento e até mesmo de inclusão e empatia. (Livia Motta, 2020)

9. Regras de acessibilidade

Para o desenvolvimento da acessibilidade em tecnologias web e tecnologia assistiva, é necessário cumprir requisitos legislativos. Em muitos países surgiram, a partir dessa necessidade, diversas regras, regulamentos e iniciativas para fornecer o acesso universal à web e ao espaço digital, além de sistemas de comunicação que atendem a todos.

Na internet as regras mais exigidas para acessibilidades são ADA e WCAG 2.1. Hoje nos Estados Unidos da América, as regras da ADA são impostas como leis e exigem que as pessoas com deficiência possam ter acesso igualitário aos serviços públicos como restaurantes, cinemas e escolas. No país, o Departamento de Justiça já concluiu que a falta de acessibilidade para sites pode ser uma violação desta lei. Embora a ADA ainda não aborde explicitamente a questão da acessibilidade digital e à web, a legislação é ampla o suficiente para poder ser interpretada de maneira a abranger a esfera da comunicação.

Enquanto nos EUA as regras já são impostas como leis, no momento de 2020 na Europa, o assunto está em pauta de agenda pública para legislação que torna obrigatória a adesão às diretrizes de acessibilidade para conteúdo web (WCAG). Já no Brasil, existe a Lei Brasileira de Inclusão da Pessoa com Deficiência, lei que foi considerada um grande avanço em relação aos direitos da pessoa com deficiência, também aborda assuntos ao acesso à informação e à comunicação é um artigo que aborda a acessibilidade pelos sites da Internet no Brasil, porém mantém a obrigatoriedade apenas, segundo o artigo 63, sobre a acessibilidade nos sítios da internet mantidos por empresas com sede ou representação comercial no País ou por órgãos de governo.

A prova da influência de leis de inclusão são os números de ações judiciais que visam a acessibilidade da web que estão crescendo rapidamente. Em 2019 e 2018, houveram

aproximadamente 2000 processos por ano sobre esse assunto nos EUA, quase triplicando dos 814 em 2017. (UsableNet, 2019)

Na busca de evitar cometer discriminação, à medida que a www e outras tecnologias digitais se tornam cada vez mais predominantes na vida cotidiana, as empresas devem entender o quanto antes que não é difícil para um usuário argumentar que a acessibilidade digital é um direito civil para todos, e com isso buscar a inclusão também por motivos legais.

Hoje existem padrões restritos para acessibilidade na web que designers e desenvolvedores seguem na construção de produtos. O modelo mais utilizado é formulado pela WAI-AGE que fornece um conjunto de diretrizes internacionalmente reconhecidas como o padrão para acessibilidade na web, como o WCAG 2.1, que são diretrizes de acessibilidade de conteúdo da Web.

O WCAG implementa as diretrizes de acessibilidade de conteúdo da web, sua documentação explica como pode-se construir um conteúdo digital de maneira acessível para todos. As diretrizes cobrem padrões para conteúdos como a informação natural: texto, imagens e sons; definições de estrutura de página; código; apresentações etc. (W3C WAI, 2005).

Seguir tais recomendações de acessibilidade tende a beneficiar a usabilidade dos produtos digitais de todas as áreas possíveis. Por exemplo, a criação de textos alternativos para imagens e objetos pode auxiliar não somente as pessoas com deficiência visual a entender a finalidade do conteúdo, mas também abrangem pessoas com conexões lentas à internet no qual o conteúdo não é carregado e ainda prevê possíveis problemas que o próprio site de hospedagem possa apresentar. É com esse objetivo em mente, que se afirma que todas as informações precisam ser compreensíveis pelas tecnologias assistivas.

Portanto, as imagens e conteúdos sem textos alternativos são imagens invisíveis para muitas pessoas, e é extremamente importante que todo o conteúdo digital seja claro e compreensível para todos usuários.

Os padrões desempenham um papel vital na definição dos requisitos de acessibilidade para cada um dos componentes de conteúdo. De fato, é necessário que hajam leis para atender as necessidades básicas de acessibilidade, mas essencialmente para os designers e desenvolvedores envolvidos na criação de produtos acessíveis, é preciso ter um entendimento aprofundado de como as pessoas com deficiência utilizam ferramentas digitais e assim conseguir implementá-las de maneira mais eficaz.

10. Futuro da acessibilidade e evolução da tecnologia assistiva

Com a evolução da tecnologia como um todo, a utilização de tecnologias assistivas no ambiente físico por pessoas com deficiência, segundo Laura Kalbag (2017), não se limita mais apenas ao uso de cadeira de rodas ou óculos, assim como a acessibilidade digital não é apenas sobre leitores de tela. Pessoas em ambientes diversos e com habilidades variadas podem beneficiar de toda acessibilidade bem aplicada.

Ainda assim, cabe apontar que mesmo os objetos físicos já estabelecidos no cotidiano da sociedade foram importantíssimos no evoluir das questões a respeito da acessibilidade. Estes apontados anteriormente continuam a seguir em constante evolução e adaptação para atender às diferentes necessidades de cada usuário.

Dentro do universo digital, para cada tipo específico de dispositivo existem muitas maneiras diferentes de inserir, enviar e trocar informações com a web. A tecnologia assistiva é apenas mais uma ponte entre o usuário e o dispositivo.

Um recurso bastante utilizado hoje, é o reconhecimento de voz que funciona através de um software ou interface de usuário de voz (*Voice User Interface – VUI*), recebendo assim informações externas pelo uso de um microfone, possibilitando que o computador ou smartphone seja operado com comandos orais.

A tecnologia de reconhecimento de voz melhorou muito nos últimos anos, atualmente existem famosas VUI difundidas no mercado que servem como assistentes do cotidiano: a Apple tem Siri, a Microsoft tem Cortana, a Amazon tem Alexa e o Google tem Google Assistant. Essas VUI podem trabalhar com qualquer voz, embora geralmente não sejam tão capazes quanto interfaces de usuário de voz dedicadas ou softwares que transcrevem da fala para o texto.

Tais assistentes de voz fornecem habilidades para lidar com o clima, fazer ligações e experienciar aplicativos de entretenimento. Os assistentes de voz trouxeram para todas as pessoas, mas principalmente para as que possuem deficiência, novas oportunidades de otimizar o tempo e realizar tarefas de forma mais acessível.

Existem softwares mais avançados que funcionam com um treinamento em *machine learning* (aprendizado de máquina), possibilitando com que através da leitura feita pelo usuário de um texto pré-definido, o programa se adapte para sua voz especificamente. Isso permite que o software reconheça nuances e sotaques na fala do usuário e melhore sua precisão.

Para Régine Gilbert (2019), este é um momento crucial para novos desenvolvimentos em tecnologia assistiva. Não apenas os produtos de TA existentes são atualizados regularmente, mas a tecnologia nova está caminhando para melhorar cada vez mais a acessibilidade das pessoas com deficiência.

A autora desafia designers a olhar para trás e reconhecer inovações passadas, para que estas possam servir de inspiração na criação de um futuro melhor para a acessibilidade. Gilbert (2019) destaca, mais uma vez, que é evidente que essas tecnologias beneficiaram muito mais do que exclusivamente a comunidade com deficiência.

Porém nem sempre a acessibilidade consegue acompanhar esse crescimento exponencial, segundo Laura Kalbag (2017) a tecnologia está em constante evolução e a uma velocidade impressionante: novos dispositivos, sistemas operacionais, navegadores e tecnologias assistivas são lançados todos os dias e todos afetam o trabalho de acessibilidade. A autora aconselha que hoje, ao tentar acompanhar esse cenário, deve-se procurar constantemente novas TAs, alterações no suporte à acessibilidade em sistemas operacionais e avanços nas tecnologias *front-end*.

Kalbag (2017) também pontua que em contraste com o ritmo acelerado das mudanças tecnológicas, as melhores práticas de acessibilidade tendem a evoluir com muito menos frequência. Ela defende que a acessibilidade que se conhece hoje não está ligada diretamente ao desenvolvimento tecnológico, mas sim às maneiras em que se pode melhorar a tecnologia que já existe e as experiências que elas proporcionam.

A autora Regine Gilbert (2019) comenta que sempre que desenvolvia novos projetos de experiência do usuário (UX) e acessibilidade, ela não estava pensando em como criar e desenvolver a partir do uso de inteligência artificial, mas apenas procurando fazer de tal experiência a mais acessível possível para não deixar ninguém de fora. Hoje, é notável que criar experiências inclusivas requer conscientização e flexibilidade.

Conclui-se que, de fato, as experiências atuais com tecnologia assistiva e acessibilidade precisam evoluir junto à tecnologia. Atualmente deparamos-nos com tecnologias de AI, reconhecimento de imagem, reconhecimento de voz, reconhecimento motor e machine learning que abrem muitas possibilidades para a criação de TAs novas, modernas e que acompanhem essa mudança, por mais que as mesmas necessitem de revisão humana. Em comparação com o nível de produção geral, são poucos os projetos voltados à acessibilidade que exploram esses novos avanços. Este trabalho irá, portanto, explorar projetos que visam o futuro da acessibilidade.

Capítulo III: Sobre reconhecimento de imagem

1. O que é a imagem invisível?

Sobre a ideia de imagem atemporal de Aumont (1942), podemos criar a questão: o que é a imagem invisível? Para esclarecer o conceito, apresenta-se dois cenários gerais em que ele se aplica: o cotidiano e o digital. Resumindo o que já foi apresentado:

- **Cotidiano:** para um deficiente visual a imagem cotidiana pode ser sensorial e palpável, formando-se por um meio não-visual. Ao ser impossibilitado do alcance sensorial, entretanto, a imagem torna-se invisível.
- **Digital:** as imagens são descritas por leitores de tela e tecnologias assistivas. Porém a imagem digital é apenas visível na presença de *alt-text*, tornando-se invisível quando não há qualquer descrição ou inteligência descritiva da imagem.

A imagem, por mais que remeta a uma ideia visual, é composta por todos os sentidos humanos. Como percebido por Aumont (1992) e Sartre (1996), o imaginário da imagem é criado também por ideias não-visuais. Ao levar esse fato em consideração, o objetivo deste projeto é dissertar sobre a importância do desenvolvimento dos outros sentidos, pois para as pessoas que possuem deficiência visual, as imagens digitais sem descrições são invisíveis.

Seguindo essa linha de pensamento onde a construção da imagem dá-se por um imaginário sensorial, o áudio, por exemplo, é um caminho para a investigação da relação entre o reconhecimento de imagem e a acessibilidade para deficientes visuais, possibilitando a transformação de textos, descrições e legendas através do seu uso.

Para essa investigação é necessário, portanto, compreender a ideia de como o reconhecimento da imagem acontece para então entender como descrevê-la. Aumont (2001) e Gombrich (1965) descrevem múltiplos conceitos de como o reconhecimento de imagem acontece no cérebro, o que pode nos auxiliar a entender como o reconhecimento da imagem digital e suas descrições acontecem hoje em dia.

2. Reconhecimento de Imagem

As imagens enquanto objetos visuais são paradoxais por excelência: possuem duas dimensões mas permitem que nelas se vejam objetos em três dimensões. Esse caráter paradoxal acontece pois está ligado ao fato das imagens mostrarem objetos ausentes, pois ao serem obtidas por uma projeção da realidade tridimensional para uma reprodução bidimensional, implica imediatamente a perda de informação por "compressão". (Aumont, 2001).

Na explicação de Aumont (2001), do ponto de vista geométrico uma imagem em perspectiva pode ser a imagem de uma infinidade de objetos que têm a mesma projeção: logo, sempre haverá ambiguidades quanto à percepção da profundidade. Para o reconhecimento de objetos representados em imagens, ou pelo menos de suas formas, é notável que existe um padrão entre os diferentes formatos geométricos possíveis e o cérebro destaca o reconhecimento mais provável dentro do contexto (Aumont, 2001).

Estudos mostram que a performance humana é capaz de reconhecer um objeto com uma taxa de erro de 5.1% (Russakovsky et al, 2015). Essa porcentagem está relacionada ao fato das representações das imagens não serem suficientemente informativas quando possuem características contraditórias, como no caso das figuras ambíguas e paradoxais apresentadas nas imagens abaixo:

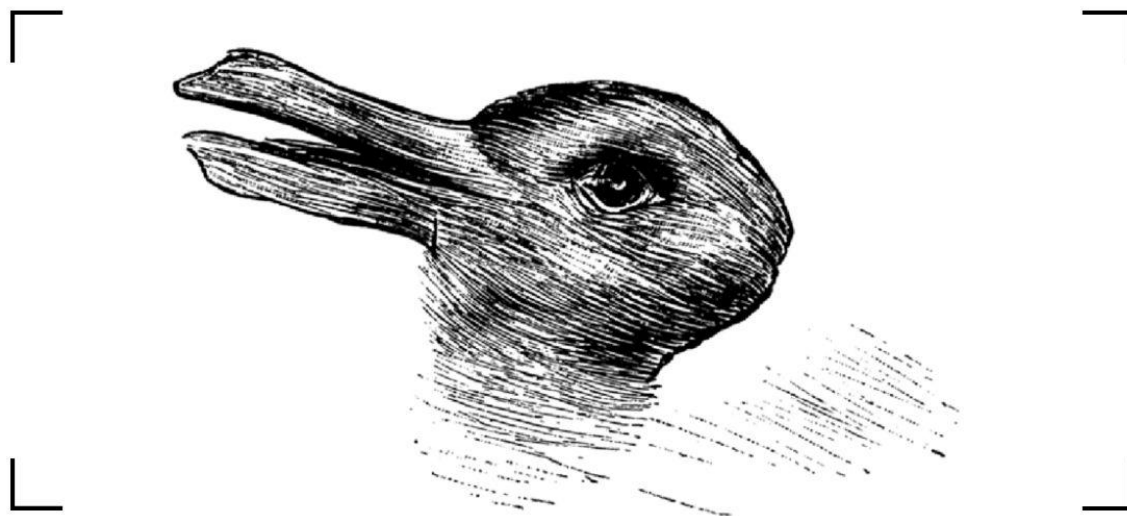


Figura 4. Kaninchen und Ente, Fliegende Blätter (1892). Descrição da imagem: A imagem é um

desenho em hachuras e mostra um contraste ambíguo entre um coelho e o pato dependendo do ponto de foco pode ser mais perceptível um animal ou outro, questionando “quais animais são mais parecidos um com o outro?”.

Muitos fatores podem influenciar a maneira com que cada pessoa identifica a imagem central: as cores, a ordem em que é colocada, o ângulo etc. Apesar da semelhança formal que cria uma maior probabilidade de ambiguidade perceptiva, para entender a probabilidade de seleção de objetos é necessário investigar a cognição (Aumont, 2001).

a. Gombrich: reconhecimento e memorização

Para iniciar essa pesquisa, Aumont (2001) aborda de frente a questão: "Por que se olha uma imagem?" (p.78), e aponta que toda relação com a imagem procede sob um conjunto de informações do universo simbólico. Para o autor, essencialmente, a resposta está contida na cognição, em transpor a imagem em termos psicológicos.

No pensamento de E. H. Gombrich, segundo Aumont (2001), existe a hipótese de que a imagem tem por função primeira garantir, reforçar, reafirmar e explicitar relação do ser com o mundo visual: ela desempenha papel de *descoberta do visual*.

Ainda de acordo com Aumont (2001), Gombrich ao estudar as imagens artísticas, cria em oposição dois principais modos de investigação psicológica sobre imagem: o reconhecimento e a memorização. Explica-se esses dois termos observando que a dicotomia coincide com a distinção entre função representativa e função simbólica, de certa maneira é uma espécie de tradução em termos psicológicos; uma puxa mais para a memória, logo para o intelecto e para as funções do raciocínio, e a outra para a apreensão do visível, para as funções mais diretamente sensoriais (Aumont, 2001).

O conceito de reconhecimento em uma imagem é definido em identificar o conteúdo da mesma. Para Aumont (2001), o processo de "reconhecer" elementos de uma imagem emprega as propriedades do sistema visual, tratando de reconhecer características do mundo real que encontram-se nas imagens, como contornos, suas cores, os gradientes de tamanho e de textura etc. Pode-se dizer, de maneira generalizada, que a noção de *constância perceptiva* está na base da

compreensão do mundo visual e permite atribuir qualidades constantes aos objetos e ao espaço (Aumont, 2001).

Em paralelo, o conceito de rememoração é, de modo geral, relativo à uma série de esquemas para a composição da imagem, onde tais esquemas resultam em codificação de fórmulas e símbolos para representação do mundo, tais como nos esquemas figurativos presentes na antiga arte egípcia ou na iconografia cristã bizantina (Aumont, 2001).

Na leitura feita por Aumont (2001), Gombrich afirma que o processo relacionado à capacidade mental apoia-se na memória, ou seja, são processados elementos memorizados em formas de objetos e de arranjos espaciais. Já a *constância perceptiva*, trata da comparação entre o que se vê no agora e o que já se viu no passado. Podendo-se assim definir que existe a possibilidade não só de se identificar os objetos contidos em uma imagem (apesar das eventuais distorções da reprodução), mas também de se reconhecer o conteúdo de uma imagem.

Um exemplo, citado por Aumont (2001), é o reconhecimento do rosto: o fato de ser possível reconhecer com facilidade o modelo de um retrato fotográfico ou de uma pintura, se dá graças à constância perceptiva. Também, o ato de reconhecimento da imagem em uma caricatura é somente possível através do uso de características *invariantes* retratadas, pois elas poderão desempenhar o papel de índices de reconhecimento, segundo Gombrich.

Ou seja, em entendimento amplo, o trabalho de reconhecimento aciona não só as propriedades básicas do sistema visual, mas também as capacidades de codificação abstratas do consciente: o ato de reconhecer não verifica exclusivamente as semelhanças entre pontos, mas, juntamente, acha invariantes através da visão, estas já estruturadas como espécies de grandes formas.

A produção de imagens jamais é gratuita, e, desde sempre, as imagens foram fabricadas para determinados usos, individuais ou coletivos. (Aumont, 2001, p.78)

b. Reconhecimento de Imagem no Séc. XX e XXI

Muito antes da criação e desenvolvimento da tecnologia como conhecida atualmente, surgiram diversas ideias de filósofos e matemáticos que são apontadas como fundamentais no que viria a se tornar a inteligência artificial. Aristóteles (384 a.C - 322 a.C), por exemplo, demonstra que esse conceito não é recente ao defender que a mente poderia ser semelhante a uma máquina segundo a ideia de Vannevar Bush (Bittencourt, 2001).

Segundo consta, em um dos primeiros livros brasileiros publicados sobre o tema, "Inteligência Artificial - ferramentas e teorias", do pesquisador e professor na área de informática Guilherme Bittencourt (2001), há séculos a inteligência artificial já é utilizada em equipamentos que marcavam o tempo e simulavam o comportamento de animais. Ao longo do tempo, o pesquisador reafirma, criaram-se aparatos matemáticos e técnicas para cálculos como o ábaco, e medidores mecânicos do tempo como os relógios. Essa evolução seguiu continuamente até a construção dos computadores, que passaram a ser desenvolvidos vigorosamente a partir de 1940.

Em 1943 os pesquisadores Warren McCulloch (1898-1969) e Walter Pitts (1923-1969) publicaram um artigo¹³ contando da experiência que fizeram: ao criar uma simulação de uma rede artificial, que funcionava com base nos números 0 e 1, (on e off, sim e não) esse modelo demonstrou que as funções poderiam ser calculadas a partir das redes neurais artificiais e que também era possível, ainda em teoria, aplicar o aprendizado de máquina (*Machine Learning*) segundo a teoria de Alan Turing.

Após a Segunda Guerra Mundial (1939-1945) o desenvolvimento da inteligência artificial evoluiu de uma forma limitada, sendo que os primeiros computadores eram vistos como caixas capazes apenas de efetuar operações e cálculos aritméticos. Na época, as ferramentas de programação não eram estimadas à capacidade independente de inteligência.

Em 1956, em um seminário da Faculdade de Dartmouth¹⁴, o termo "Inteligência artificial" foi criado e utilizado pela primeira vez por John McCarthy (1927-2011), professor, matemático, cientista e também o pai da linguagem de programação LISP - List Processing. O desenvolvimento dessa linguagem foi também uma célebre conquista de McCarthy, que a apresentou em 1958 e serviu como facilitadora no desenvolvimento da inteligência artificial. A

¹³ MCCULLOCK, Warren. PITTS, Walter. (1943) A logical calculus of the ideas immanent in nervous activity. Originally published in Bulletin of Mathematical Biophysics, vol. 5, 1943, p. 115-133.

¹⁴ McCarthy, J., Minsky, M., Rochester, N., Shannon, C.E., (Agosto, 1955) A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence., Recuperado em 2020, de <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>

linguagem é uma das mais antigas e foi usada pela primeira vez ao colocar um computador para jogar xadrez contra um adversário humano, ainda hoje ela permanece em uso.

Nesta primeira fase da Inteligência Artificial, entre os anos 1952 e 1969, houve um grande entusiasmo e uma crescente demanda de sistemas inteligentes, assim surgiram também grandes experimentos criados por cientistas como Herbert Simon (1916-2001) e Allen Newell (1927-1992), fundadores do primeiro laboratório de inteligência artificial, criado na Universidade Carnegie Mellon. Um dos projetos do laboratório foi a criação do GPS - General Problem Solver (Solucionador de Programas Gerais, tradução nossa) em 1957, o primeiro programa projetado para imitar protocolos humanos de resolução de problemas.

Durante e após a Revolução Digital (1950-1970), também conhecida como a Terceira Revolução Industrial, o reconhecimento de padrões, área envolta da inteligência artificial e do aprendizado de máquina, era formado apenas por teorias estatísticas. No final da década 1940 e início de 1950, Norbert Wiener, matemático conhecido como fundador da cibernética, integrou o grupo The Macy Conferences, grupo que configurou uma série de dez conferências interdisciplinares, contribuindo para a consolidação da teoria cibernética¹⁵ junto com outros cientistas. Foi somente após os anos 60 que surgiu, graças à grande demanda por criação de sistemas com inteligência artificial, a procura por sistemas capazes de reconhecer padrões. Um dos fatores que impulsionaram os estudos e o desenvolvimento desses sistemas, foi o conceito cada vez mais popular de produção em massa.

Um sistema de reconhecimento de padrões consiste de um sensor ou câmera que obtém observações a serem classificadas ou descritas; um mecanismo de extração de características que computa informações numéricas ou simbólicas das observações; e um esquema de classificação das observações, que depende das características extraídas. O reconhecimento de padrões tem como objetivo a classificação de objetos dentro de um número de classes. As categorias de estudo variam entre imagens, ondas de som, luz ou qualquer tipo de medida que necessite ser classificada.

Foi após a década de 1980 que se iniciou o desenvolvimento de sistemas com inteligência artificial em demanda por grandes empresas, o que apesar de ter feito aumentar o investimento e

¹⁵ Norbert Wiener definiu a cibernética em 1948 como "o estudo científico do controle e da comunicação no animal e na máquina". Em outras palavras, é o estudo científico de como humanos, animais e máquinas controlam e se comunicam. Retirado de: Cybernetics: Or Control and Communication in the Animal and the Machine. 28 de abril de 2020.

as pesquisas na área, ainda continha sistemas muito limitados e que não eram capazes de resolver todos os problemas solicitados. Apenas, no final da década, a inteligência artificial passou a ser considerada como uma ciência, resultando em pesquisas mais aprofundadas para uma melhor compreensão de sua complexidade.

Woody Bledsoe (1921-1995) foi uma das primeiras pessoas a fazer contribuições no reconhecimento de padrões e no raciocínio automatizado, por isso é considerado um dos fundadores da AI. Nos anos de 1960, o cientista e matemático foi também um dos primeiros a realmente começar a pensar em segmentar imagens de partes do rosto humano, tentando descobrir curvas, formas e detalhes em um conjunto de números para reconhecimento de padrões, o que pode-se dizer que foi a origem do reconhecimento facial. Até hoje estudos do cientista continuam a influenciar obras de artistas e escritores da área.

c. O Boom da Inteligência Artificial e do Machine Learning

David Carmona (2019) descreve em seu livro *The AI Organization*, que recentemente algo lhe chamou atenção sobre o que antes parecia um beco sem saída no desenvolvimento de software: os avanços nas redes neurais da comunidade de pesquisa e o vasto poder de computação disponível na nuvem estavam fornecendo resultados surpreendentes.

No ano de 2015 houve o renascimento da inteligência artificial. Anteriormente o termo descrevia "apenas" ocasiões em que uma máquina fosse capaz de executar tarefas normalmente associadas à inteligência humana, como prever uma jogada de xadrez, e por mais poderosas que fossem as análises nesse momento, a tecnologia definitivamente não estava nem perto do nível que alcançaria nos próximos anos (Carmona, 2019).

Neste ano, uma equipe da Microsoft Research desenvolveu uma rede neural tão profunda que foi capaz de identificar imagens melhor do que um ser humano, e havia energia de computação suficiente para executar a rede neural instantaneamente. Em questão de meses, os resultados record tornaram-se comuns nas notícias: características humanas foram alcançadas em tarefas como reconhecimento de fala, leitura de máquina e até tradução.

A partir de 2015 tudo mudou, muitas empresas passaram a utilizar tanto as tecnologias disponíveis quanto o próprio termo inteligência artificial. A principal razão para essa mudança foi

a crescente sofisticação das técnicas disponíveis para construir esses sistemas, particularmente na área de Machine Learning.

As técnicas de aprendizado de máquina já vinham sendo usadas no passado para construir sistemas de inteligência capazes de criar modelos matemáticos baseados em dados de treinamento e também fazer previsões de forma autônoma, mesmo que sem serem explicitamente programados para cumprir tal função (Carmona, 2019).

Hoje, a inteligência artificial junto ao machine learning são tópicos muito discutidos na área de tecnologia e possibilitam uma grande visão para o futuro das tecnologias assistivas e da acessibilidade. Para entender como essas áreas se conectam e podem se ajudar precisa-se primeiro ter clara as possibilidades que podem ser alcançadas com a relação entre essas tecnologias avançadas.

A capacidade da inteligência artificial é ampla e complexa, ainda assim existem diversas habilidades cognitivas específicas dos seres humanos que não fazem parte do que as máquinas podem alcançar, e ainda pode levar algum tempo até que elas cheguem lá.

A combinação de todas essas habilidades cognitivas humanas em uma máquina que poderia ser utilizada para qualquer cenário, é chamada de *General Artificial Intelligence*. As técnicas atuais estão mostrando grande sucesso na execução de tarefas mais restritas, como a tecnologia nomeada de *Narrow AI* que se refere principalmente a três recursos: Aprendizado, Percepção e Cognição.

O aprendizado, ou aprendizado de máquina, é a capacidade que o programa tem de aprender com o tempo de processamento, sem a necessidade de programação explícita. A percepção é uma área que por décadas tenta imitar a capacidade humana de perceber o mundo ao redor, até hoje com sucesso limitado, porém em constante evolução. Por fim, a cognição define-se pela capacidade de adquirir e processar conhecimento, envolvendo construções de alto nível que as inteligências usam para raciocinar, entender, resolver problemas, planejar e tomar decisões (Carmona, 2019).

Além dos elementos base de habilidades das AIs, existem vários algoritmos que podem ser usados para o Machine Learning, mas há um em particular que está por trás da explosão do uso do termo inteligência artificial: redes neurais artificiais (RNAs). Esse algoritmo é baseado em como os neurônios funcionam no cérebro humano, onde considera-se que cada neurônio é uma função de ativação simples, ligada a outros neurônios por meio de conexões ponderadas. Nos

últimos anos vê-se uma enorme quantidade de inovação relacionada a essa técnica com redes constituídas por dezenas de camadas, também podendo ser chamada de *Deep Learning* (Carmona, 2019).

Por mais que a máquina possa levar algum tempo para alcançar as habilidades cognitivas humanas, vale a ressalva de que não deve demorar muito para ela chegar lá. Pode-se supor essa afirmação devido as recentes evoluções divulgadas, como o lançamento em junho de 2020 do GTP-3¹⁶, um modelo inteligente que combina técnicas de *machine learning* e *deep learning*, criado pelo laboratório de pesquisa em inteligência artificial OpenAI.

O *Deep Learning* está por trás da maioria das descobertas recentes de AI relatadas. Com essa técnica, máquinas podem reconhecer elementos visuais por computer vision, reconhecer objetos em imagens, transcrever fala, responder perguntas sobre um texto ou traduzir textos de um idioma para outro, tudo com uma taxa de precisão melhor do que um humano pode alcançar (Carmona, 2019).

Essas três habilidades de Aprendizado, Compreensão e Cognição, combinadas ao Machine Learning e ao Deep Learning, formam milhares de possibilidades e recursos para serem utilizados e explorados na tecnologia. Essa dissertação deseja destacar principalmente os recursos tecnológicos utilizados para a inclusão, como as ferramentas de acessibilidade que usando desta combinação assertiva de forma consciente e ética, pode criar soluções cada vez mais abrangentes na resolução de espaços não inclusivos.

3. Introdução ao computer vision

*Se queremos que as máquinas pensem,
precisamos ensiná-las a ver.*

—Fei Fei Li

¹⁶ GTP-3 (Generative Pre-trained Transformer 3) é um modelo de linguagem autorregressivo que usa *deep learning* para produzir texto traduzido em vídeo com aparência humana. É o modelo de previsão de linguagem de terceira geração da série GPT-n criada pela OpenAI, um laboratório de pesquisa de inteligência artificial com sede em San Francisco. <https://openai.com/>

O nascimento do computer vision se dá após a Revolução Digital, como resultado do aprofundamento das tecnologias que pretendiam, de maneira a imitar a capacidade de cognição da visão humana, trabalhar o reconhecimento de padrões em imagens.

Através de uma câmera, a visão computacional compreende de forma relativa ambientes visuais e seus contextos. Devido a progressiva evolução do computer vision, hoje é possível perceber padrões humano-máquina para o reconhecimento de imagens, ambientes, objetos e faces por computadores e smartphones.

Um dos projetos iniciais mais ambiciosos foi o *The Summer Vision Project*¹⁷ em 1966, por A. Seymour Papert (1928-2016), que tinha a proposta de descrever imagens a partir da visão proporcionada ao anexar uma câmera ao computador. Um dos primeiros objetivos era construir um sistema que dividiria a imagem em partes, diferenciando os objetos do fundo.

Em 1982, Dana H. Ballard e Christopher M. Brown, em seu livro "Computer Vision"¹⁸, definiram o estudo sobre a visão computacional como uma construção de descrições significativas e explícitas de objetos físicos a partir de imagens. Depois de 16 anos, Emanuele Trucco e Alessandro Verri, em seu livro "Introductory Techniques for 3-D computer vision" (1998), a definiram como propriedades computacionais do mundo 3D a partir de uma ou mais imagens digitais. Em 2001, George Stockman e Linda G. Shapiro em seu livro também intitulado "Computer Vision", consideraram esta inteligência uma ferramenta capaz de tomar decisões úteis sobre objetos físicos reais e cenas baseadas em imagens detectadas.

A maioria dos sistemas de inteligência artificial utilizam inteligências cognitivas, as quais não estão focadas em perceber o mundo ao nosso redor e na realidade visam abstraí-lo. Mesmo nesses casos, a cognição é intrínseca à percepção. O computer vision é, por exemplo, um recurso que utiliza cognição e percepção com o potencial de auxiliar as pessoas a entenderem o mundo ao seu redor de maneira mais exata. Por exemplo, um sistema de computer vision que analisa exames de Raio X consegue identificar anomalias na imagem com maior velocidade e precisão do que um humano, melhorando ainda mais a longo prazo conforme for coletando mais informações.

Em citação à cientista e professora chinesa Fei-Fei Li (2015), atualmente diretora do

¹⁷ PAPERT, A. Seymour. 1966 The Summer Vision Project, artigo disponível em: <https://dspace.mit.edu/handle/1721.1/6125>

¹⁸ BALLARD, Dana. BROWN, Chris. 1984 Computer Vision. Livro restaurado por British Machine Vision Association e pelo EU's ECVision Network on Cognitivecomputer vision disponível em: <http://homepages.inf.ed.ac.uk/rbf/BOOKS/BANDB/bandb.htm>

Laboratório de Inteligência Artificial da Universidade de Stanford, "pouco a pouco, estamos dando visão às máquinas. Primeiro, as ensinamos a ver, então elas nos ajudam a ver melhor. Pela primeira vez, os olhos humanos não serão os únicos ponderando e explorando nosso mundo". Nessa fala, Li explicita a maneira com que a relação entre pessoas e AIs acontece, onde o criador, ao alimentar sua criação, é também estimulado a crescer.

Os recentes avanços tecnológicos e estudos sobre Deep Learning foram os fatores que levaram à evolução da capacidade do computer vision na última década, algo que hoje é visto em constante aprimoramento e em velocidade cada vez maior. Juntas, essas duas tecnologias tornaram-se uma grande ferramenta capaz de explorar diversas técnicas, algumas das mais utilizadas podem ser descritas como:

- **Classificação de Imagem:** Classificar o conteúdo de uma imagem;
- **Reconhecimento de Objetos:** Reconhecer os objetos em uma imagem e identificar os limites de cada um deles;
- **Reconhecimento de Ambientes:** Reconhecer cenas, atividades e comportamentos em imagens;
- **Reconhecimento Facial:** Detectar rostos, reconhecê-los e identificar emoções através de suas expressões faciais;
- **Reconhecimento de Texto:** Reconhecer texto escrito, incluindo texto manuscrito (reconhecimento óptico de caracteres);
- **Reconhecimento Específico:** Identificar um tipo conteúdo específico através de um modelo treinado em imagens e vídeos;
- **Reconhecimento de Comportamento:** Identificar padrões, comportamentos e traços de uma pessoa pela utilização de produtos digitais.

Existem também muitas ferramentas derivadas do computer vision que já estão inseridas no dia a dia de milhares de pessoas, como o restauro de imagens, jogos online, vigilância, biometria e carros autônomos. Algumas empresas oferecem o serviço de APIs¹⁹ integradas ao computer vision para criação de softwares de classificação e reconhecimento de imagem, como

¹⁹ API, *Application Programming Interface*, é um conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do software, mas apenas usar seus serviços.

Google Cloud Vision API, Microsoft Azure Computer Vision API, AWS Rekognition, IBM Watson Visual Recognition, CloudSight API entre outros.

O computer vision é um recurso utilizado em quase todas redes sociais, está em aplicativos como Instagram, principalmente com o reconhecimento facial para a utilização de filtros, no Facebook para criar marcações de pessoas, assim como no Google Lens é possível analisar uma imagem para identificá-la como semelhante. Um dos mercados em constante adaptação a essas novas tecnologias é o da segurança, tanto pública quanto privada. Vê-se cada dia mais câmeras especializadas espalhadas, utilizadas na identificação e detecção de comportamentos de maneira cada vez mais precisa. Na China já existem softwares capazes de fazer reconhecimento de emoções em softwares de vigilância policial²⁰, enquanto no Brasil já são aplicadas câmeras de alta qualidade capazes de fazer reconhecimento facial e de realizar relatórios de assiduidade e comportamento de pessoas dentro de grandes igrejas²¹.

a. Reconhecimento e classificação de padrões em imagem

Com a evolução do *machine learning*, pesquisadores da área de *computer vision* foram possibilitados de construir uma abordagem orientada a dados que tem como objetivo resolver e criar um algoritmo que classifica imagens em categorias distintas.

A classificação na linguagem de programação é a capacidade de identificar a classe ou categoria de itens pertencentes à uma imagem com base em seus recursos; por exemplo, ao processar uma imagem de trânsito é possível classificar os elementos incluídos na captura, como carros e semáforos.

Para classificar uma imagem é necessário obter um conjunto de outras imagens rotuladas a uma categoria, assim o algoritmo consegue analisar cada grupo de imagens e resultar com precisão a classificação de um objeto ou cenário. A exatidão do resultado é consequência direta de quantas vezes um mesmo objeto foi processado por imagens diferentes. Por exemplo, um processamento em um banco de imagens com cerca de mil fotos de cadeiras, seria mais apto a classificar o objeto com precisão do que se tivesse apenas cem imagens, levando em consideração

²⁰ Emotion recognition is China's new surveillance craze, 2020, Financial Times, artigo disponível em: <https://www.ft.com/content/68155560-fbd1-11e9-a354-36acbbb0d9b6>

²¹ Empresas lançam serviço de reconhecimento facial para igrejas no Brasil, 2019, A Publica, artigo disponível em:

<https://apublica.org/2019/11/empresas-lancam-servico-de-reconhecimento-facial-para-igrejas-no-brasil/>

a infinita possibilidade de ângulos, cenários e design.

Além da classificação por imagem existem outros sistemas que utilizam técnicas de classificação constantes, como os sistemas de recomendação que são capazes de encontrar semelhanças entre itens como filmes, livros e músicas, ou até mesmo definir o comportamento de compra através de um site *e-commerce*. Um tema bastante discutido contemporaneamente, representado em documentários como "O Dilema das Redes" (2020), dirigido por Jeff Orlowski e que entrevista grandes especialistas da privacidade de dados e bem-estar na web como Tristan Harris e Shoshana Zuboff, é a falta de legislação e também de compreensão dos consumidores sobre a maneira como utilizam esses produtos sem a percepção que eles mesmo podem tornar-se mercadoria no que diz respeito à compra e venda de informações, sendo direcionados e restringidos ao mesmo tempo.

Segundo a documentação da Microsoft Azure Computer Vision API²², os resultados de suas classificações de imagens são baseados em milhares de objetos reconhecíveis, seres vivos, cenários e ações. Quando as classificações são ambíguas ou pertencem a um conhecimento não comum, a resposta da API fornece 'dicas' que tentam esclarecer o significado dentro de um contexto num cenário conhecido. Atualmente esse tipo de classificação não está limitada ao assunto principal, como uma pessoa em primeiro plano, mas também inclui o cenário (interno ou externo), móveis, ferramentas, plantas, animais, acessórios, gadgets etc.

Apesar das recentes evoluções, ainda existem vários desafios associados a essa tarefa, incluindo variação de ponto de vista, variação de escala, variação de espécies, deformação de imagem, oclusão de imagem, condições de iluminação, confusão de fundo, entre outros. O fluxo resumido de classificação de imagens pode ser exemplificado da seguinte maneira:

²²Documentação da API do Microsoft Azure computer vision disponível em: <https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/>

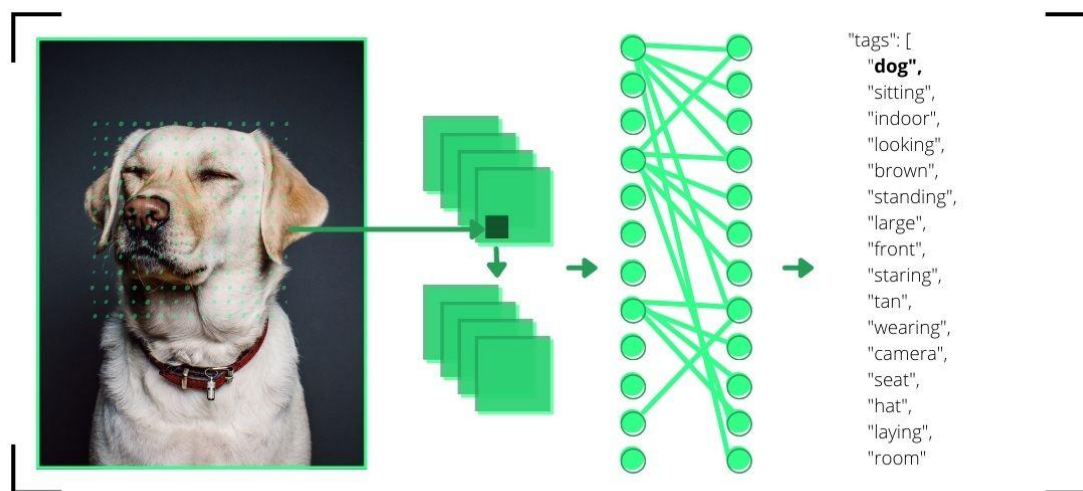


Figura 5. Esquema de classificação de imagem, do processamento à classificação. Descrição da imagem: Imagem contém a figura de um cachorro labrador bege com uma coleira vermelha e fundo escuro à esquerda, no meio uma representação de como o processamento computacional funciona baseado em pixels e quadrados que são identificados e classificados como elementos que retornam por fim mostrado à direita uma lista de tags possíveis para classificação dessa mesma imagem, na ordem de mais precisão, como “dog”, “sitting”, “indoor”, ”looking”, ”brown”, ”standing”, ”large”, ”front”, ”staring”, “tan”, “wearing”, “camera”, “seat”, “hat”, “laying” e “room”.

A técnica de classificação fornece ao computador muitos exemplos de cada grupo de imagem e ainda leva-se em consideração que é possível criar algoritmos de *machine learning* que analisem esses exemplos e aprendam progressivamente sobre a aparência visual de cada classe. Com essa técnica de aprendizado é possível criar modelos de classificação de imagens muito específicos, podendo identificar carros, pessoas, expressões faciais, cachorros, pássaros, objetos e qualquer tipo de classe visual. (Bishop, C. M., 2006)

b. Reconhecimento e Descrição de Imagem

Para o processamento de reconhecimento de imagem é necessário que existam bancos de

imagens com classificações em larga escala. Existem hoje alguns bancos de dados visuais que contêm milhares de classificações e modelos pré-treinados que tornam os processos possíveis. Alguns bancos de dados conhecidos são *xView*, *ImageNet*, *Open Image*, *IMDB-Wiki*, *Color Detection* e *Stanford Dogs dataset*.

c. Computer vision em benefício da inclusão

Nos dias atuais a utilização do computer vision segue crescendo exponencialmente em áreas que trabalham com o reconhecimento de imagens para fins de vigilância, monitoramento, reconhecimento de pragas na agricultura, reconhecimento de padrões para compras online e até mesmo identificação de anomalias em imagens médicas laboratoriais. Existem também iniciativas e aplicativos como o Envision AI da Envision Technologies e o Seeing AI da Microsoft, que desejam criar um mundo mais inclusivo através do uso da inteligência artificial, porém mesmo essas tecnologias, talvez pelo fato de serem muito recentes, ainda possuem pouco espaço dentro da acessibilidade digital.

Muito ainda pode ser explorado no campo de computer vision, mas é necessário que exista um foco para que essas ferramentas sejam utilizadas em benefício da sociedade, pois ainda parece haver maior interesse no seu desenvolvimento no que diz à conversão de lucros. É preciso criar ferramentas e desenvolver as que já existem para tornar o espaço digital um local cada vez mais inclusivo para todos.

4. O outro lado da inteligência artificial

a. Introdução à problemática

O uso da inteligência artificial é cada vez maior e com sua ascensão podemos ver os dois lados da moeda cada vez mais claros: a tecnologia pode tanto ser utilizada para inclusão, como também pode ser o ponto de partida para exclusão e manipulação.

Desde os anos 60 muito mudou na AI, incluindo a troca de foco de sistemas simbólicos para técnicas de aprendizado de máquina. Tendo em vista apenas a última década, pode-se notar uma grande expansão da inteligência artificial no mundo acadêmico e na indústria, além de ser perceptível como apenas um pequeno número de poderosas corporações tecnológicas aplicam sistemas de AI em escala internacional hoje em dia (Campolo & Crawford, 2020).

Segundo Alexander Campolo e Kate Crawford (idem), autores e pesquisadores na área, as técnicas de *deep learning* estão cada vez mais populares dentro da prática de AI. As técnicas de aprendizado, que identificam padrões em conjuntos de dados de grande escala para poder fazer classificações e previsões, são hoje mais precisas que quando feitas por humanos. Porém, sendo essa habilidade muito recente, existe uma carência de teorias para explicar porque as técnicas de *deep learning* funcionam com tanta precisão para detectar e prever padrões. Pesquisadores contemporâneos e líderes de campo passaram a descrever técnicas de deep learning como se fosse mágica (Campolo & Crawford, 2020).

Em seu artigo, os autores analisam o significado do *deep learning* de maneira historicamente específica, questionando quais são as características dos sistemas contemporâneos, as aplicações sociais que o levaram a ser assim e quais os efeitos que essas declarações produzem (idem).

Os problemas relacionais surgem quando os efeitos de predição e categorização social são desconhecidos pelos próprios criadores da AI, sendo assim inexplicavelmente injustos para quem é prejudicado e excluído pela forma que a tecnologia foi construída. Segundo Citron e Pasquale (2014), citados por Campolo e Crawford (ibidem), sem uma pesquisa e entendimento, aplicações que usam essas técnicas são uma ameaça não apenas devido ao processo legal que pode vir a acontecer, mas também uma ameaça às formas mais expansivas de contestação política e social, ao mesmo tempo que ela distancia os designers de inteligência artificial e as corporações que os empregam da responsabilidade ética e legal (Campolo & Crawford, ibidem).

Segundo Campolo e Crawford (ibidem), a contestação de injustiça começa com a compreensão de como a discriminação social afeta os diferentes grupos de pessoas. Se as diferenças são ocultadas por mecanismos de classificação, os quais não são possíveis de se interpretar, que são justificados por atingir um desempenho melhor do que o humano, é difícil saber como fazer demanda por justiça.

Atualmente vê-se que as experimentações dentro do mundo acadêmico e da indústria, em larga escala e com dados, estão assumindo um papel significativo para moldar tanto o comportamento científico como muitas coisas do dia-a-dia. Desde plataformas de redes sociais ao trânsito de estradas, dados são colhidos a cada segundo, conduzindo experimentos em público. Ainda assim poucas pesquisas são publicadas a respeito dos riscos de se identificar, rastrear e mitigar informações sobre as pessoas, que muitas vezes estão conscientes de estarem sendo monitoradas (Crawford & Metcalf, 2016).

Kate Crawford e Trevor Paglen (2019) argumentam em seu artigo *Excavating AI*, que a inteligência artificial não será necessariamente pior do que os sistemas operados por humanos em fazer previsões e orientar decisões. Pelo contrário, engenheiros de todo o mundo estão otimistas de que a AI pode ajudar a detectar e reduzir a exclusão e o preconceito humano. Porém, em alguns contextos atuais, estudos indicam que alguns sistemas AI afetam desproporcionalmente os grupos que já estão em desvantagem por fatores como raça, gênero e histórico socioeconômico (Crawford & Paglen, 2019).

Os sistemas contemporâneos de aprendizado de máquina são construídos com bases de modelos de treinamento compostos por conjuntos de dados, sendo fundamentais para sistemas de inteligência funcionarem, reconhecerem e interpretar o mundo. Os modelos de dados são as fronteiras que governam como cada sistema de AI opera, e, portanto, é essencial compreendê-los para entender questões sociais relacionadas ao AI (idem).

Crawford e Paglen (ibidem) explicam que as imagens não se descrevem. Este é um recurso que muitos artistas exploram há séculos, como o surrealista René Magritte que pintou a imagem de cachimbo com as palavras "isto não é um cachimbo", dando à imagem uma maneira diferente de ser interpretada ao rotulá-la. A imagem e suas rotulações são flexíveis e podem ser reconstruídas de várias maneiras, assim como podem mudar com o tempo e conforme o contexto cultural, podendo vir a significar coisas diferentes dependendo do espectador e sua localização no mundo. Ou seja, imagens estão constantemente abertas à interpretação e reinterpretação (Crawford & Paglen, 2019).

É evidente que existe uma complexidade cultural para descrição de imagens, ainda mais levando em conta a ressignificação das imagens ao longo do tempo, mas isso não significa que imagens não devem ser descritas. A importância da descrição da imagem, como já foi levantado aqui, é essencial para acessibilidade e inclusão digital. A interpretação e o *bias* devem ser

deixados de lado para a criação da descrição de imagem, é essencial que as descrições não carreguem bagagens tendenciosas e tenham como objetivo ser o mais claras possível.

Percebe-se atualmente a propagação de sistemas com reconhecimento facial e de emoções dentro das áreas de educação, saúde e segurança. Sistemas como esses já fazem parte das verificações "de segurança" em aeroportos e em entrevistas de grandes empresas. O fato de não ser público a base dos modelos em que os AI são treinados e também seus métodos para entender como funcionam refletem em graves consequências (idem). Logo no início do seu artigo *Excavating AI*, Crawford e Paglen (ibidem) questionam: qual é o papel das imagens em sistemas de AI? O que os algoritmos deveriam reconhecer em uma imagem e qual é a problemática reconhecida ou invisível nessa área?

Mergulha-se hoje em uma era na qual dispositivos pessoais são vistos como *proxies*²³ para a necessidade pública, e com isso aumenta-se o risco de que as desigualdades já existentes sejam ainda mais impregnadas. A cada conjunto de *big data* construído, é necessário questionar quais pessoas podem estar sendo excluídas, quais lugares são menos visíveis e o que acontece com essas pessoas ao viver na sombra de grandes conjuntos de dados (Crawford, 2013).

b. Algoritmo de exclusão e *bias* de classificação

Segundo Aumont (2001), não há imagem sem percepção de uma imagem. O estudo das grandes características da imagem, mesmo que rápido, evita muitos erros na compreensão da percepção, simplesmente pelo fato de que se a imagem é arbitrária e inventada, ou plenamente cultural, sua visão é quase imediata. O estudo intercultural da percepção visual demonstrou que pessoas que nunca passaram por esse teste antes têm capacidade inata de perceber tanto os objetos figurados em uma imagem, quanto sua organização de conjunto – contanto que lhes dêem os meios para empregarem essa capacidade, ao explicar-lhes o que é uma imagem (ibidem).

Para o autor francês, ao se fixar os olhos em uma imagem inicia-se um processo de varredura sobre várias experiências vividas. Também é importante ressaltar uma vez mais que, em princípio, a *percepção* das imagens, contanto que se consiga separá-la de sua *interpretação*, é um processo próprio à espécie humana, apenas mais aprimorado por certas sociedades (ibidem).

²³ Proxies, plural de Proxy. Proxy é um servidor (um sistema de computador ou uma aplicação) que age como um intermediário para requisições de clientes solicitando recursos de outros servidores.

Atualmente o maior banco de dados visuais disponível é o ImageNet, contendo mais de 14 milhões de imagens de 20 mil tipos de classificações e modelos para reconhecimento de objetos, sendo também o recurso mais popular para classificação de imagem desde 2010²⁴.

O projeto do ImageNet foi iniciado pela escritora e cientista Fei-Fei Li com a ajuda de voluntários, contando também com o Amazon Mechanical Turk para tirar fotos e classificar os objetos das imagens²⁵. Até hoje participantes de todo o mundo enviam algoritmos ajudando na criação de modelos de reconhecimento mais precisos. No início do projeto o ImageNet tinha uma taxa de erro em torno de 25%, já em 2012, o universitário Alex Krizhevsky apresentou como solução uma rede neural composta por oito camadas chamada AlexNet, alcançando uma taxa de erro de 15,3%. Nos anos seguintes a técnica introduzida foi aprimorada, em 2014 surgiu a GoogLeNet com 22 camadas, atingindo uma taxa de erro de 6,7% (Krizhevsky, Sutskever, Hilton, 2017).

Em 2015, o ano crucial para evolução da AI, uma equipe da Microsoft Research criou uma nova técnica de rede neural residual com uma profundidade de 152 camadas, alcançando uma taxa de erro de apenas 3,57%. Este momento consta como um marco importante, dado que o desempenho humano em reconhecer objetos em imagens cobre em média uma taxa de erro de 5.1%, sendo assim pela primeira vez a máquina conseguiu superá-lo (He, Zhang, Ren, Sun, 2015).

A maioria das técnicas de classificação de imagens são treinadas em modelos do ImageNet, alguns dos métodos mais eficazes de *computer vision* existentes foram testados neste conjunto de dados.

Nas palavras de sua co-criadora, a professora Fei-Fei Li, a ideia por trás do ImageNet era "mapear o mundo inteiro dos objetos"²⁶. Ao longo de vários anos de desenvolvimento, o crescimento do ImageNet foi enorme: a equipe pegou uma coleção de milhões de imagens da internet e rapidamente se tornou o maior usuário acadêmico do mundo do Amazon Mechanical Turk²⁷, usando uma grande equipe de trabalhadores (pessoas contratadas pela Amazon Turk)

²⁴Dados disponíveis em www.image-net.org

²⁵Dados retirados do artigo, publicado 11/03/2018, pela Wired, escrito por Jessi Hempel "Fei-Fei Li's Quest to Make AI Better for Humanity".

²⁶ Citação de Fei-Fei Li encontrada em Dave Gershgorin, "The Data That Transformed AI Research—and Possibly the World," Publicado por Quartz, em 26 de julho de 2017 <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/>. Emphasis added.

²⁷ Amazon Mechanical Turk é um site de crowdsourcing para que as empresas contratem trabalhadores localizados remotamente para realizar tarefas discretas sob demanda que os computadores atualmente não conseguem fazer. Disponível em www.mturk.com

fragmentados para classificar uma média de 50 imagens por minuto em milhares de categorias. Por uma década, o ImageNet tem sido o colosso do reconhecimento de objetos para o aprendizado de máquina e uma referência de grande importância para o campo (Crawford & Paglen, 2019).

Os modelos de treinamento de imagens categorizadas são a essência do *computer vision*, hoje eles são onipresentes nas tecnologias de reconhecimento de imagem, porém modelos de alguns bancos como o do ImageNet são construídos em base de suposições sobre a representação, rótulos e categorizações da natureza dessas imagens. Vale também citar que essas suposições remetem a abordagens históricas em que as pessoas eram rotuladas e classificadas visualmente por ferramentas que não só preservavam, como também reafirmavam condições da opressão racial (Crawford & Paglen, 2019).

Atualmente a problemática desse banco de dados visual, como explicam Kate Crawford e Trevor Paglen (2019), é de que:

Você abre um banco de dados de fotos usadas para treinar sistemas de inteligência artificial. No início, as coisas parecem simples. Você se depara com milhares de imagens: maçãs e laranjas, pássaros, cães, cavalos, montanhas, nuvens, casas e placas de rua. Mas à medida que você investiga o conjunto de dados, as pessoas começam a aparecer: líderes de torcida, mergulhadores, soldados, escoteiros, caminhantes do fogo e floristas. As coisas ficam estranhas: uma fotografia de uma mulher sorrindo de biquíni é rotulada como 'desleixada, vagabunda, mulher desleixada'. Um jovem que bebe cerveja é classificado como 'alcoólatra, dipsomaníaco, bebedor de álcool, exuberante, de molho, maria'. Uma criança que usa óculos escuros é classificada como uma 'pessoa fracassada, perdedora, insolvente e mal sucedida'. [...] Algo está errado com esta imagem. De onde vieram essas imagens? Por que as pessoas nas fotos foram rotuladas dessa forma? Que tipo de política ocorre quando as imagens são combinadas com rótulos e quais são as implicações quando são usadas para treinar sistemas técnicos? Resumindo, como chegamos aqui? (Crawford & Paglen, 2019).

Os autores argumentam que o ImageNet contém 2.833 subcategorias na categoria “Pessoa”, onde se classificam as imagens incluindo raça, nacionalidade, profissão, status econômico, comportamento, caráter e até mesmo questões que abrangem a moralidade, como: pessoa má, garota de programa, viciado em drogas, condenado, louco, fracassado, filho da puta, hipócrita, Jezabel, cleptomaniaco, perdedor, melancólico, não pessoa, pervertido, esquizofrênico, solteirona etc. Para além desses exemplos, existem ainda muitos outros insultos racistas e termos misóginos (Crawford & Paglen, 2019).

Kate Crawford comenta que no ano de 2019, subcategorias de "pessoas" do ImageNet começaram a desaparecer e com isso 1,2 milhões de fotos ficaram indisponíveis nos servidores da Universidade de Stanford. Segundo a autora:

As fotos de líderes de torcida, mergulhadores, soldados, coroinhas, aposentados e pilotos se foram. A foto de um homem bebendo cerveja caracterizado como “alcoólatra” desapareceu, assim como as fotos de uma mulher de biquíni apelidada de “desleixada” e um menino classificado como “perdedor”. A imagem de um homem comendo um sanduíche (rotulado de “pessoa egoísta”) teve o mesmo destino. Quando você busca por essas imagens, o site ImageNet responde com uma declaração de que está em manutenção, e apenas as categorias usadas na competição ImageNet ainda estão incluídas nos resultados da pesquisa (Crawford, 2019).

Curiosamente, Campolo e Crawford afirmam que a Amazon Mechanical Turk também tomou uma atitude parecida:

Em 2018, a Amazon Mechanical Turk contratou profissionais para remover todas as raças do seu conjunto de dados - essa limpeza de dados pode ser considerada como exclusão racial (2018, 248). A questão da generalização torna-se política quando os conjuntos de dados de treinamento são considerados representativos de nossos mundos sociais,

quando os proponentes da IA usam a precisão como justificativa para colocar limites na discricção, julgamento e agência humanos. Questões técnicas de precisão e desempenho são permeadas por escolhas políticas sobre categorias e normas (Benthall e Haynes 2019). Mas eles raramente são reconhecidos como tais." (Campolo & Crawford, 2020).

Os pesquisadores Joy Buolamwini e Timnit Gebru (2019) do MIT, trabalharam no Microsoft Research e mostraram que mesmo alguns dos softwares de reconhecimento mais avançados erram ao identificar mulheres de pele escura em 35% das vezes, em contraste com a mesma análise resultante em 1% para homens brancos. De igual forma, muitos estudiosos apontam há anos os riscos técnicos e sociais do reconhecimento facial, considerando ser necessário uma proteção legal que garanta os direitos de civilidade, justiça e responsabilidade para a circulação desses *softwares* em locais públicos, caso contrário essa tecnologia tem um grande potencial de exclusão e de acabar tornando as pessoas menos livres (Crawford & Paglen, 2019).

Em setembro de 2019, o site do ImageNet publicou um relatório dizendo que por ser um conjunto de dados de pesquisa muito influente, sua plataforma deveria ser examinada criticamente. O relatório reconhece que, por exemplo, sistemas de reconhecimento facial demonstraram possuir taxas de erro desproporcionais entre os grupos raciais. A plataforma reconhece ter 2.832 categorias de classificações de pessoas e que esse uso pode ser problemático, levantando questões importantes sobre justiça e representação. Nesse mesmo relatório são descritos esforços de pesquisa ao longo de 2018 para identificar e solucionar esses problemas, no qual reportam: a construção de um nova versão do ImageNet com remoção de classificações ofensivas ou pejorativas, remoção de má representatividade e de termos tendenciosos, e adição de novas imagens assegurando a diversidade de acervo esse trabalho manual ainda constava em andamento até o momento da publicação do relatório (Yang, et al 2019).

Segundo Crawford e Paglen (2019), a IBM criou um conjunto de dados chamado "Diversity in Faces", como uma resposta aos críticos que mostraram que o software de reconhecimento facial da empresa muitas vezes não reconhecia os rostos de pessoas com pele mais escura. A IBM prometeu publicamente melhorar seus conjuntos de dados de reconhecimento facial para torná-los mais "representativos". O Diversity in Faces foi construído

para ser um modelo e banco de imagens para garantir precisão e equidade do reconhecimento facial, montado especificamente para alcançar paridade estatística entre as categorias de tom de pele, estrutura facial, idade e gênero.

Kate Crawford (2013) critica o ex-editor-chefe da revista *Wired*, Chris Anderson, em seu artigo *The Hidden Biases in Big Data*. Anderson afirmou que "com dados suficientes, os números falam por si", onde Crawford questiona até onde o *big data* poderia cumprir essa promessa, pois os dados e conjuntos de dados não são objetivos e neutros como dá-se a entender, mas sim uma criação de design do ser humano. A autora comenta que ao dar voz aos números, o ser humano sofre interferência de si próprio, definindo os seus significados a partir de sua interpretação pessoal. A coleta, a interpretação e a percepção dos dados é humana, o que resulta na responsabilidade de quem o cria.

Para Crawford e Paglen (2019) os conjuntos de dados não são apenas matérias-primas que alimentam algoritmos, mas também modos de intervenções políticas. Todo o esforço em coletar imagens, categorizá-las e rotulá-las é em si uma forma política, pois é relacionada a elas que surgem questões sobre quem decide o que as imagens significam e que tipo de trabalho social essas representações prestam. Bias e algoritmos tendenciosos são enraizados com a visão de mundo de quem os constrói.

c. Dados são produtos e o produto somos nós

Quem tem o poder de construir e se beneficiar desses sistemas?

—Kate Crawford

Existem muitas vantagens na utilização de novas tecnologias para a evolução do mundo e da maneira que se entende, porém a forte tendência na captura e venda de dados por essas mesmas tecnologias cria uma dualidade a ser discutida contemporaneamente. Uma das problemáticas mais importantes a respeito é: quem é beneficiado por ambos lados?

Hoje em dia grandes empresas detentoras de sistemas de inteligência, os utilizam para criar produtos que possam vender mais. A autora Shoshana Zuboff (2019) conta que durante a revolução industrial, o empresário automobilístico Henry Ford conseguiu reduzir o valor de um automóvel em 60% através de um processo de alto volume de fabricação e baixo custo de

produção. Ford nomeou esse processo de produção em massa, momento em que tornou notória a seguinte frase: “o cliente pode ter o carro da cor que quiser, contanto que seja preto”.

Essa forma de produzir passou a ser aplicada a qualquer tipo de produto, modificando a estrutura de como a produção acontecia à medida que se juntava aos valores da economia. Esse processo estabeleceu o domínio de um novo capitalismo que tinha como base a produção em massa, objetivando um maior giro monetário e a ampliação de riquezas no século XX (Zuboff, 2019).

O capitalismo, dessa forma, evoluiu como o resultado das necessidades das pessoas em um determinado tempo e lugar. Henry Ford foi claro neste ponto: “A produção em massa começa na percepção de uma necessidade pública” (Zuboff, 2019). A autora explica que ainda assim:

Essa nova sociedade industrial manteve muitos dos motivos hierárquicos do antigo mundo feudal em seus padrões de afiliação com base em classe, raça, ocupação, religião, etnia, sexo e os leviatãs da sociedade de massa: suas corporações, locais de trabalho, sindicatos, igrejas, partidos políticos, grupos cívicos e sistemas escolares" (Zuboff, 2019).

Em outras palavras, Zuboff demonstra que mesmo através de diferentes sistemas econômicos, as estruturas que mantêm a desigualdade como base persistem até os dias atuais. Por consequência, a criação e manutenção das tecnologias é também usada como mecanismo de sustentação dessa máquina excludente.

Nas últimas décadas pode-se observar que assim como Ford explorou uma nova maneira de produção e consumo, a Apple foi uma das primeiras empresas a ter uma explosão de sucesso comercial ao valer-se de uma nova sociedade que demanda por consumo individualizado. No entendimento de Zuboff (2019), atualmente as tecnologias de comunicação, informação e dados, são mais propagadas do que a eletricidade, alcançando quase metade da população mundial.

A autora explica em seu livro *A Era do Capitalismo de Vigilância* (2019), que um dos primeiros passos para entender o seu conceito de capitalismo de vigilância embutido no dia a dia, é saber diferenciá-lo das tecnologias que o envolve e qual a lógica infiltrada por trás dessas tecnologias, onde se reivindica a experiência humana como matéria-prima gratuita para a interpretação de dados.

Dessa matéria-prima coletada, Zuboff acredita que alguns dados são aplicados para melhoria do produto ou serviço, enquanto o restante alimenta algoritmos que fabricam produtos de previsão que antecipam o que o usuário fará no agora, em breve e depois. A autora diferencia que quando uma empresa coleta dados comportamentais com permissão e apenas como um meio para melhorar o produto ou serviço, ela está comprometendo o capitalismo, mas não o capitalismo de vigilância. Cada uma das cinco maiores empresas de tecnologia pratica o capitalismo, evidentemente, mas a autora entende que nem todas são capitalistas de pura vigilância, pelo menos ainda. (Zuboff, 2019).

Os produtos de previsão modernos serão negociados em futuros mercados comportamentais que vão além dos anúncios online, sendo direcionados para muitos outros setores, incluindo seguros, varejo, finanças e uma gama cada vez maior de empresas e serviços determinados a participar desses novos e lucrativos mercados (Zuboff, 2019).

Sistemas de *machine learning* aprendem com os padrões que estão presentes nos dados nos quais são treinados, neste caso existem dois modelos: o padrão e o experimental. Enquanto o padrão pode ser feito com supervisão e, por consequência, potencialmente limitado na capacidade de aprender padrões que violem a privacidade, o experimental, em contraste, é autônomo e não se restringe a aprender com dados históricos já observados. Estes sistemas realizam experimentos para obter novos dados sobre as pessoas nas mais diversas situações, ou seja, são capazes de violar a privacidade e revelar dados que podem ser confidenciais (Barocas, et all 2016).

Shoshana Zuboff (2019) argumenta que grande parte dos aplicativos para *smartphones* exigem o acesso à localização mesmo quando a funcionalidade não necessita dessa informação. Segundo a autora, "varejistas usam "geofencing" para demarcar uma área geográfica e enviar alertas para smartphones dentro desses parâmetros: 'Venha aqui agora!', 'Compre isto aqui!', 'Uma oferta, só para você!'." (Zuboff, 2019).

Solon Barocas, et all (2016), explicam que esses sistemas podem descobrir combinações específicas de um anúncio para aumentar a probabilidade de que o usuário faça uma compra. Essas combinações em uma sequência de técnicas experimentais, montam um perfil de compra e comportamento para cada usuário específico e podem gerar percepções causais que excedem seu nível de consentimento.

Segundo uma reportagem feita por Kashmir Hill (2020), em 2019 mais de 600 organizações de aplicação de lei começaram a usar o Clearview AI, um aplicativo de

reconhecimento facial inovador desenvolvido por uma pequena empresa. Ao tirar uma foto de uma pessoa, o aplicativo retorna todas as fotos públicas da mesma, incluindo a fonte dessas imagens. A base desse sistema é composta por um banco de dados de mais de três bilhões de imagens que o Clearview afirma ter copiado do Facebook, YouTube, Vimeo e milhões de outros sites.

Essa AI tem o potencial de acabar com a privacidade e anonimato em locais públicos, levando em consideração que a ferramenta fornece milhares de dados dos cidadãos norte-americanos para centenas de agências, desde o F.B.I até os Departamentos de Segurança Interna. O código do aplicativo, analisado pelo The New York Times, inclui a possibilidade de combinar seus dados com óculos de realidade aumentada; ou seja, quem o usar seria potencialmente capaz de identificar cada pessoa que ver, automaticamente. A ferramenta poderia identificar ativistas em um protesto ou um estranho no metrô, revelando não apenas seus nomes, mas também onde moram, o que fazem e quem conhecem (Hill, 2020).

Na visão de Zuboff (2019), empresas insistem em seu direito de usar sistemas de reconhecimento facial para serem capazes de identificar um “estranho na rua” mesmo sem obter o consentimento do indivíduo. Atualmente, estudiosos de leis de privacidade respondem que não existe um direito legalmente estabelecido para tal funcionamento.

Em 2016, a NTIA²⁸ publicou suas “Recomendações de práticas de privacidade para uso comercial de reconhecimento facial”, onde questiona-se a quem as diretrizes apresentadas realmente beneficiam. Segundo essas recomendações, as empresas de tecnologia são “encorajadas” a tornar suas políticas de reconhecimento facial “disponíveis aos consumidores”. Pergunta-se então, quem realmente lê políticas de privacidade? Nas palavras de Zuboff (2019), “cada indivíduo lutando com as inúmeras complexidades de sua própria proteção de dados não será páreo para as assombrosas assimetrias de conhecimento e poder do capitalismo de vigilância”, e continua, “se as últimas duas décadas nos ensinaram alguma coisa, é que o indivíduo sozinho não pode suportar o peso dessa luta na nova fronteira do poder”. Na web, o simples ato de navegar em um site obriga o usuário a cumprir seus termos de serviço, mesmo que este não seja explícito. A falta de legislação traz cada vez mais a sensação de sociedade vigiada, a

²⁸ NTIA, sigla de National Telecommunications and Information Administration, órgão do Poder Executivo norte-americano responsável principalmente por assessorar o presidente em questões de telecomunicações e políticas de informação. Disponível em www.ntia.doc.gov

realidade se torna uma metáfora à sociedade de controle escrita por Foucault e Deleuze. A cada clique que se dá, uma nova pegada é rastreada.

Shoshana Zuboff (2019) aclama artistas que lutam contra a discrepância dessa falta de legislação, haja visto que atualmente muitos dedicam suas habilidades às perspectivas da invisibilidade humana. A arte aqui toma um papel político e crítico, assumindo muitas formas: capas de *smartphones* bloqueadoras de sinal, próteses de impressão digital falsa, visores de privacidade de LED que impedem a leitura de câmeras, camisetas cobertas de rostos de celebridades para confundir *software* de reconhecimento facial, casacos que bloqueiam ondas de rádio e dispositivos de rastreamento, um difusor de perfume que libera uma fragrância metálica quando um site ou rede desprotegida é detectada em qualquer um de seus dispositivos, dentre outros vários exemplos. Leo Selvaggio, um artista de Chicago, produziu máscaras em resina impressas em 3D para confundir o reconhecimento facial, chamando seu trabalho de “uma intervenção artística organizada”.

O designer brasileiro Pedro G. C. Oliveira, pesquisador residente no Interactive Telecommunications Program da NYU, e a designer e educadora sediada no Brooklyn, Xuedi Chen (2015), criaram um kit de ferramentas chamado Backslash. Este projeto foi criado como parte de uma dissertação de mestrado e reflete perfeitamente a disputa pela terceira modernidade que esta geração enfrenta. O kit contém uma variedade de dispositivos funcionais projetados para protestos e tumultos do futuro, nos quais incluem-se uma bandana inteligente para incorporar mensagens ocultas, dispositivos vestíveis em rede independente, dispositivos pessoais de caixa preta para registrar abusos de aplicação da lei e roteadores de implantação rápida para comunicação fora da rede. Os criadores escrevem que para jovens manifestantes digitalmente nativos, “a conectividade é um direito humano básico, e ainda assim, o futuro da tecnologia em protestos parece sombrio”, isso causado, principalmente, pela vigilância esmagadora (Zuboff, 2019).

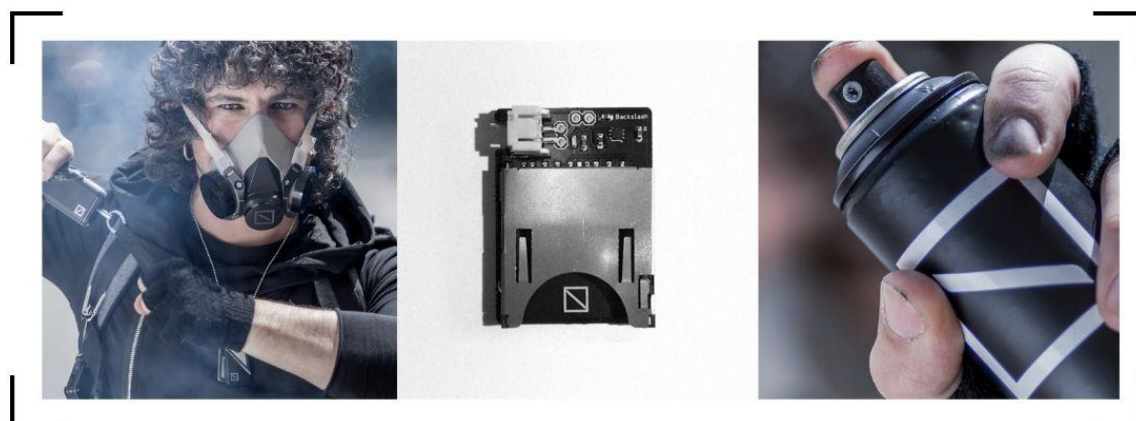


Figura 6. Backslash, projeto de Pedro G. C. Oliveira e Xuedi Chen (2015). Descrição da imagem: A imagem é composta por três fotografias do produto Backslash, onde a primeira fotografia mostra uma pessoa com cabelos escuros, olhos escuros, pele branca, com uma máscara de gás, vestindo um casaco preto, luvas pretas e segura um dispositivo roteador para acesso durante protestos, ao redor da pessoa existem uma névoa branca como uma fumaça. A segunda imagem é de um dispositivo chamado Drive, que possui a capacidade de fotos e vídeos anonimamente, o dispositivo é retangular e feito de metal e plástico. A terceira imagem é composta por uma mão segurando um spray de tinta, com um tubo preto e um símbolo prata que representa que a área está sob vigilância.

Já o artista radicado em Berlim, Adam Harvey, é apontado por Zuboff (2019) por possuir um trabalho voltado diretamente à problemática da vigilância. Harvey possui um projeto chamado “Stealth Wear”, uma série de peças de moda vestíveis destinadas a oprimir, confundir e evitar a vigilância de drones e, de forma mais ampla, softwares de reconhecimento facial (Zuboff, 2019).

Existe um número crescente de artistas que direcionam seus trabalhos para as áreas de vigilância e resistência, propondo maneiras práticas de unir arte com ativismo. Inclusive, desde os anos de 1990 o termo "ativismo" é utilizado por pessoas que usufruem desse encontro como maneira de se manifestar. É como um grito sufocado de "acorde", é necessário tomar atitudes objetivas que mudem o percurso dessa história. Zuboff comenta:

Cada passo nesse caminho ocorre como se estivesse na névoa da guerra: fragmentos espalhados e incidentes que aparecem abruptamente e muitas vezes na obscuridade. Há pouco espaço para perceber o padrão, muito menos suas origens e significado. No entanto, cada exclusão da possibilidade de um mundo digital melhor deixa um vazio que é preenchido de forma contínua e silenciosa pelas novas condições do poder instrumentário (Zuboff, 2019).

Na sociedade contemporânea, a tecnologia é também uma expressão permeada por ideais e objetivos capitalistas, haja visto que a conexão digital e a coleta de dados já se tornaram um mercado. As pessoas, por fim, são tratadas como seres irracionais e invisíveis, decifráveis exclusivamente por meios de 0 e 1. Zuboff escreve:

O que está em jogo é o princípio dominante da ordem social em uma civilização da informação e nossos direitos como indivíduos e sociedades para responder às perguntas: Quem sabe? Quem decide? Quem decide quem decide? O fato de o capitalismo de vigilância ter usurpado tantos de nossos direitos nesses domínios é um abuso escandaloso das capacidades digitais e sua outrora grande promessa de democratizar o conhecimento e atender às nossas frustradas necessidades de vida efetiva. Que haja um futuro digital, mas que seja primeiro um futuro humano (Zuboff, 2019).

5. A inclusão é o destino final

Como você orienta eticamente pensamentos e ações de duas bilhões de pessoas todos os dias?
—Tristan Harris

A intenção dessa dissertação é relatar as e questionar as práticas que garantem o direito de privacidade, de ética e de inclusão, principalmente no que diz respeito à criação de sistemas que

utilizem reconhecimento de padrões e inteligência artificial.

Está estabelecido o início de uma nova era de experimentação. A utilização do *computer vision* vem crescendo exponencialmente: de um lado para o reconhecimento de imagens com fins principalmente lucrativos, através da vigilância, monitoramento e análise de padrões para compras online; enquanto do outro com motivos um pouco mais benéficos socialmente, porém sem fugir completamente do ideal capitalista, para combater pragas que afetam a agricultura, na identificação de anomalias em imagens médicas laboratoriais, na identificação de notícias falsas e na criação de aplicações mais inclusivas e acessíveis.

Kate Crawford (2019) afirma que os engenheiros da área estão otimistas de que a AI pode ajudar a detectar e reduzir o preconceito humano, porém ainda são necessários muitos estudos que apontem as problemáticas dos sistemas de AI que afetam desproporcionalmente os grupos de minoria. Junto disso é também essencial que se criem leis regulamentadoras sobre como esses sistemas devem ser construídos seguindo preceitos de ética.

Kate Crawford e Jason Schultz (2019) argumentam que a maioria das soluções para essas preocupações foram focadas na supervisão tecnológica ou regulatória, entretanto diante do cenário atual, poucas dessas abordagens conseguiram fornecer estruturas de responsabilização adequadas, seja por não conseguirem abordar os aspectos sociais e estruturais mais amplos dos problemas, seja por falta de vontade política para implementá-las. Os autores concluem que é hora de considerar novos paradigmas para a responsabilização, especialmente para potenciais violações constitucionais. A amplitude necessária em questões legislativas é complexa e vasta, pois ainda que uma AI seja criada em um país com certo tipo de regulamentação, em todos os outros que for utilizada estará sujeita a diferentes leis, tornando essas preocupações ainda mais alarmantes no que diz respeito às relações internacionais.

Em seu artigo *Halt the use of facial-recognition technology until it is regulated*, Crawford (2019) conta que a tecnologia de reconhecimento facial ainda não está pronta para expansão e implementação pública, assim como nem mesmo os governos estão prontos para evitar que essas AIs causem mais danos do que benefícios. Regulamentos mais fortes são urgentemente necessários, assim como um debate público mais amplo sobre o impacto que já está acontecendo. A legislação deveria garantir restrições ao seu uso, bem como transparência e outros direitos básicos. Até que ocorram essas implementações, é preciso uma prorrogação sobre o uso dessa tecnologia em espaços públicos.

A autora comenta que até mesmo o maior fornecedor mundial de câmeras corporais policiais, o Axon, situado no Arizona, anunciou em 2019 que não implantaria tecnologia de reconhecimento facial em nenhum de seus produtos por não acreditar que seria confiável para o trabalho policial e “poderia exacerbar as desigualdades existentes no policiamento, por exemplo, penalizando comunidades negras ou LGBTQ” (Crawford, 2019). No início de setembro de 2020, Portland aprovou a proibição da utilização de bancos de dados de reconhecimento facial por policiais, a nova lei proíbe o uso público de reconhecimento facial. Já a proibição do uso privado da tecnologia entra em vigor no dia 1º de janeiro de 2021²⁹.

É graças ao trabalho de pioneiros sobre o design ético da tecnologia, incluindo as professoras de ciência da informação e influentes da área Batya Friedman e Helen Nissenbaum, que junto de pesquisadores e de empresas, começaram a implementar estruturas como design sensível a valores ou 'inovação responsável', para ajudá-los a identificar prováveis interessados e seus valores (Cal & Crawford, 2016).

Um dos exemplos que resultam dessa trajetória é o de Tristan Harris, cientista da computação e fundador do Center Human Technology³⁰. O projeto consiste na interseção da natureza humana, da tecnologia e da transformação de sistemas, onde o objetivo é mudar a mentalidade a partir da qual os sistemas de tecnologia persuasiva são construídos, usando esse processo para apoiar mudanças cruciais em nossos sistemas econômicos e sociais, realinhando a tecnologia com a humanidade.

Técnicas de design são utilizadas para estabelecer as opiniões das pessoas sobre a privacidade pessoal, o meio ambiente etc. Os valores e opiniões das pessoas em potencial são incorporados ao design da tecnologia, seja um aplicativo ou um sistema de carro autônomo. É necessário que desenvolvedores de sistemas de AI abram os olhos e comecem a se basear mais nesses métodos importantes (Cal & Crawford, *idem*).

A tecnologia e o bem estar podem andar juntos, para isso acontecer é inevitável que haja regulamentação e que as grandes corporações façam parte dessa mudança. Apesar do capitalismo intrínseco, é possível criar um equilíbrio onde produtos e serviços sejam também inclusivos, onde

²⁹ Informação de acordo com a reportagem "Portland officials pass strict ban on facial recognition systems" publicado por Mariella Moon, 9 de setembro de 2020. Disponível em: <https://www.engadget.com/portland-facial-recognition-ban-035952590.html>

³⁰ Informação sobre Center Human Technology disponível em www.humanetech.com

a intenção de utilização de grandes sistemas de inteligência artificial seja para tornar um mundo melhor.

O reconhecimento de imagem, assim como várias outras ferramentas derivadas do AI, tem um grande potencial não só para a acessibilidade ou para servir de auxílio, mas para tornar a vida em sociedade mais englobante, mais possível e justa para todas e todos. A criação de alt-text inteligentes que possuem a capacidade de tornar imagens acessíveis para todos na web, é um dos passos importantes para isso.

O AI é uma fechadura, onde a porta é a acessibilidade e a inclusão é o destino final. Quem quer que consiga abrir essa porta, está dando o primeiro passo para tornar o mundo um local mais inclusivo.

Investigação ativa: Projeto Imagible.app

Um grande desafio natural para a comunidade do computer vision é projetar tecnologias para auxiliar pessoas deficientes visuais a superar seus verdadeiros desafios visuais diários. —Captioning Images Taken by People Who Are Blind Danna Gurari, Yinan Zhao, Meng Zhang, Nilavra Bhattacharya University of Texas at Austin, 2020.

1. Nota introdutória

Imagible - Accessible images for all, é um projeto que visa explorar e investigar como o reconhecimento de imagem pode auxiliar ferramentas de acessibilidade. O projeto nasceu sob a investigação dessa dissertação, na qual foram identificadas soluções para tornar a web mais acessível.

Atualmente ferramentas tecnológicas combinadas com inteligência artificial são elementos poderosos para a acessibilidade. Como resultado da pesquisa, decidiu-se criar uma plataforma com a habilidade de descrever imagens cotidianas, ambientes e cenários, para auxiliar designers e desenvolvedores na construção de legendas, descrição de imagens e textos alternativos adequados.

O projeto identificou, após pesquisas, testes, entrevistas e feedbacks, uma grande urgência por inovação em projetos focados em acessibilidade e tecnologias assistivas. Hoje em dia o texto alternativo necessita ser adicionado manualmente, a criação de alt-text automáticos com revisão é um benefício que surge a fim de otimizar o tempo. Conscientemente, o projeto não faz o processo completo de automação de descrições de imagens para *alt-text*, pois considera a revisão humana ponto essencial para a qualidade neste momento, visando impactar o modo que designers e desenvolvedores pensam os textos alternativos, assim como visa criar imagens mais acessíveis e, conseqüentemente, uma web mais inclusiva.

Os métodos e experimentações desse projeto resultam, por fim, na criação de um app web capaz de descrever cenários cotidianos por texto. A construção desta aplicação web funciona através do reconhecimento de imagem, proporcionando reflexões sobre o que é a "imagem sem

imagem". Esta aplicação, que ainda está em fase Beta e em testes, tem como foco auxiliar tecnologias assistivas de maneira que possa ser utilizado em benefício da inclusão. Os resultados alcançados são:

- Uma aplicação web open source, chamado Imagible (www.imagible.app), no qual pessoas podem utilizar da ferramenta reconhecimento de imagem para descrever imagens e gerar *alt-text* automaticamente, sob revisão, para seus projetos.

2. O projeto

Imagible, *accessible images for all*, é um projecto que nasceu no âmbito da investigação académica na Universidade de Lisboa e este projeto tem como ponto central explorar e investigar como o reconhecimento de imagem pode auxiliar a acessibilidade, onde as ferramentas tecnológicas são elementos fundamentais para a inclusão.

A relação entre os conceitos desta investigação resulta, finalmente, na criação de um sistema capaz de descrever objetos e cenários do dia a dia, a partir de quaisquer imagens enviadas pelo usuário.

Imagible é um projeto social para ajudar a tornar as imagens acessíveis para todos. O projeto usa a API do Microsoft Azure Computer Vision para ajudar os desenvolvedores e designers a criar um texto alternativo mais confiável e apropriado para seus sites e empresas. Ao tentar acompanhar o cenário da acessibilidade, o projeto procura auxiliar novas tecnologias assistivas, assim como dar suporte à acessibilidade diante dos avanços nas tecnologias front-end.

3. Metodologias

A criação do projeto utiliza de métodos e processos de Design Thinking, desde a parte de descoberta e investigação dos temas, onde se procura entender os *pain points* e conhecer o caminho a ser explorado, assim como se definem as jornadas e a partir da ideação, onde explora-se múltiplas alternativas de soluções na criação de valor e se transforma as possíveis soluções em ideias e projetos. Após essa etapa, parte-se para a criação do produto, passando pela

arquitetura de informações, definições de tecnologias, criação de wireframes e a interface UI, onde se parte do início de um desenvolvimento experimental de *computer vision* em paralelo junto a testes de *user testing e rapid testing* para criação de melhorias e validação do produto. Fases do projeto incluem:

- **Exploração:** Descobertas, Ideações, Grupos, Personas e Foco
- **Criação:** Wireframes, Branding, UI, Prototipação, Desenvolvimento
- **Validação:** Testes com usuários - user testing, rapid testing e roadmap

Esse projeto utiliza empatia a todo momento, pensando na perspectiva de suas *personas*, a fim de escutar as vozes de tais para criar soluções mais acessíveis e inclusivas. Em adição, esse projeto conta com muitas técnicas experimentais, procurando sistematicamente evidências e soluções para apoiar as ideias presentes no projeto, mantendo em mente a complexidade para testar e desenvolver o mesmo.

Esse projeto tem como objetivo auxiliar grupos de personas como desenvolvedores, criadores de conteúdo e designers para criar produtos digitais. Em consequência de criar imagens com descrições apropriadas a fim de criar uma equidade na web para pessoas com deficiência visual.

4. Fase: Exploração

a. Ideação e descobertas



Figura 7. *Board* de post-its com algumas das principais descobertas. Descrição de imagem: A imagem contém post-its com os dizeres: *Screen readers* não leem imagens ou não conseguem identificar sem tags apropriadas; Desenvolvedores não se sentem responsáveis por criar descrições de imagens; Designers não são conscientes de como criar descrições de imagens apropriadas; Criadores de conteúdo não são conscientes do que são textos alternativos; Acessibilidade e principalmente descrições de imagens são constantemente deixadas por último; Membros do time de produtos digitais constantemente empurram a tarefa uns aos outros; 2.2 bilhões de pessoas com algum tipo de deficiência visual no mundo; Descrições de imagens são constantemente negligenciadas por corporações; Relatórios e documentação de acessibilidade são longas e muitas vezes difíceis de entender; Times de UX nem sempre incluem pessoas com deficiência durante testes de produtos; Regras acessibilidade são constantemente ignoradas principalmente em times e produtos menores; Redes sociais meramente visuais são excludentes sem alt-text ou descrições de imagens.

Explorou-se diante dos seguintes cenários:

- Leitores de tela possuem muitas limitações para lerem imagens.

- Descrições de imagens são frequentemente negligenciadas dentro de corporações e times de produtos digitais.
- Desenvolvedores, designers e criadores de conteúdo, frequentemente não são conscientes de como criar descrições de imagens apropriadas.
- Designers e criadores de conteúdo não são usualmente conscientes do que são textos alternativos.
- Dentro de times de projeto, designers e desenvolvedores encontram dificuldades para criar descrições de imagens.
- Durante o desenvolvimento de produtos digitais, descrições de imagens são constantemente deixadas por último, membros do time constantemente empurram a tarefa uns aos outros.

Por fim encontrou-se às seguintes necessidades:

- Criação de uma documentação simples e de fácil entendimento sobre acessibilidade.
- Criação de projetos em redes sociais com conteúdo ilustrativo e informativo relacionado às práticas de acessibilidade, focados em desenvolvedores e designers.
- Conscientização, dentro de documentações para educação de times de produtos digitais, sobre os papéis e a importância da acessibilidade.
- Plataforma experimental de inteligência artificial, como ferramenta para auxiliar na criação de descrições de imagens.
- Criação de um plugin que funciona junto a leitores de tela, com a habilidade autônoma de descrever imagens.
- Um pacote NPM para projetos em javascript (como React.js e Vue.js) que descreve imagens do projeto.

b. Grupos, Personas & Foco

Os grupos focados nesta pesquisa foram: desenvolvedores e designers criadores de produtos digitais, com o objetivo de aprenderem sobre os benefícios de criarem-se imagens

descritivas apropriadas para pessoas deficientes visuais, desenvolvedores e designers portadores de deficiência visual e também profissionais especializados em acessibilidade na web.

Grupo	<i>Pain Points</i>
Designer	Sentem-se frequentemente excedidos pelas regras, alguns importam-se muito com acessibilidade maior parte do tempo, porém alguns apenas presumem que não faz parte da sua responsabilidade de criar <i>alt-text</i> apropriados ou descrições de suas imagens.
Desenvolvedores	Sentem-se frequentemente sobrecarregados em criar tags de acessibilidade, não são constantemente conscientes da importância de <i>alt-text</i> apropriados, apenas colam no código o que lhes foi enviado e também manifestam o sentimento que não deveriam ser criadores ou responsáveis por conteúdo.
Criadores de conteúdo	Sentem-se frequentemente surpresos por não saber e não conhecer tags de acessibilidade ou o que um <i>alt-text</i> significa.
Deficientes Visuais	Sentem-se frequentemente excluídos sobre as imagens visuais na web, e também são grandes entusiastas de novas tecnologias para navegar na web.

As personas desta pesquisa correspondem a usuários pertencentes aos segmentos acima, todas elas foram criadas com base em pesquisas contemporâneas, junto, principalmente, da base modificada de personas definidas pelo W3C Web Accessibility initiative, dentro do guia WAI Site UCD User Personas, de 2009, construídos pelos colaboradores Carol Smith, Tom Croucher, Charmane Corcoran, Jon Dodd, Matt May, and Shawn Lawton Henry e desenvolvido pelo WAI Site Task Force of EOWG.

I. Tanya, 32, Product Designer

"Acessibilidade faz parte da minha rotina diária de trabalho, mas mesmo tomando todos os cuidados de acessibilidade, contraste e hierarquia das páginas, as imagens muitas vezes ficam em segundo plano."

Tanya é designer de produto sênior, trabalha há muitos anos na área com interfaces, usabilidade e experiência do usuário. Ela sabe que faz parte do seu trabalho tornar seu produto acessível, porém precisa ficar pedindo e testando o produto para que os desenvolvedores sigam as regras de acessibilidade. Muitas vezes ela não sabe quem deve realizar a tarefa de descrever as imagens, achando que talvez os desenvolvedores devessem fazê-la. Sem o conhecimento adequado de como descrever as imagens apropriadamente e também sem saber se essa tarefa deve ser parte do seu trabalho, a descrição fica muitas vezes a cargo de desenvolvedores que a irão definir como um simples detalhe de acessibilidade.

Potencial: Tanya ama criar produtos para pessoas e sempre se preocupa em tornar seus designs acessíveis. Apesar de sempre apontar para os desenvolvedores como a acessibilidade irá funcionar no seu design, as imagens ficam de fora. No momento, ela acha que a acessibilidade é um grande passo para inclusão, mas quando seu design possui muitas imagens para descrever, essa responsabilidade fica em mãos de outras pessoas.

II. Marcus, 26, Designer

"Quero ter certeza de que meu design aparenta ser bom, mas também sei que preciso seguir esses padrões."

Marcus é formado em design gráfico e trabalha em uma pequena agência digital, recentemente começou a pesquisar como desenhar sites e sistemas. Atualmente ele está criando um novo produto digital, mesmo sabendo que deve torná-lo acessível, ele coloca a aparência do design em primeiro lugar, pois inclusive acha as documentações de acessibilidade muito longas. Sem o conhecimento adequado, Marcus não faz nem ideia de como tornar as imagens de seu

produto acessíveis e, segundo ele, são os desenvolvedores que devem definir como os detalhes de acessibilidade irão funcionar no seu produto, não o designer.

Potencial: Marcus ama o design que faz e ele poderia se acostumar a fazer todos os seus sites com as questões de acessibilidade em mente. Assim, com um pouco mais de conhecimento da documentação, poderia sempre apontar para os desenvolvedores como a acessibilidade iria funcionar no seu design. Mas fazer sites e produtos digitais não é sua principal função no dia a dia da agência que trabalha, então ele apenas tenta fazer o melhor trabalho que pode no tempo que lhe é dado. No momento, ele acha que a acessibilidade apenas o atrapalha.

III. Andrei, 39, Desenvolvedor

"Eu tenho um prazo e, de qualquer maneira, não temos nenhum usuário com deficiência. Eu olhei os logs do servidor, eu sei."

Andrei é um desenvolvedor de software sênior, há muitos anos executa servidores da web, constrói sites e bancos de dados, é também a pessoa mais técnica da empresa e assume a liderança nas decisões técnicas. Ele pensa que todas as regras de acessibilidade são apenas deixas do "politicamente correto" e comenta que "não é como se pessoas com deficiência realmente acessassem seu site".

Andrei não vê necessidade de acessibilidade e quando seu chefe soube disso, pediu para dar uma olhada no site da empresa. O objetivo do desenvolver é reunir algumas informações rápidas e concisas sobre acessibilidade para implementá-la no site. Seu principal interesse é aplacar seu chefe para que ele possa trabalhar em outras coisas.

Como Andrei é a única pessoa da empresa que sabe fazer sites, a decisão é dele. A menos que ele veja necessidade e o benefício da acessibilidade, ele não vai se dedicar a essa implementação.

IV. Martina Prado, 32, Front-end developer

"Eu não tenho certeza de quais diretrizes seguir. São muitas. Não existe apenas uma biblioteca que podemos instalar e aplicar os códigos?"

Martina Prado é desenvolvedora web de uma pequena empresa de design, na qual ela se orgulha muito do design visual que desenvolveu. Ela conseguiu manter sua empresa funcionando durante a recente recessão e quer posicionar sua empresa em local privilegiado mirando o crescimento futuro conforme a economia se recupera. Para fazer isso, ela sente que precisa ter muito conhecimento sobre todas as áreas de desenvolvimento da Web, incluindo a acessibilidade.

Martina recentemente teve um cliente perguntando sobre a conformidade com os padrões nacionais e, conversando com outros desenvolvedores da Web sobre o assunto, ela descobriu sobre WCAG. Ela está um tanto familiarizada com as idéias gerais de ambos, no entanto não tem certeza sobre as diferenças entre os dois, não sabendo qual funcionaria melhor para seus clientes. Além disso, para agregar acessibilidade aos serviços de sua empresa, ela gostaria de encontrar informações que pudesse usar para convencer seus clientes do valor agregado ao se ter um site acessível.

Martina lida com uma ampla variedade de responsabilidades, incluindo: reuniões individuais com clientes, programação, gerenciamento de projetos, design de sites da Web etc. Martina é muito criativa e aventureira, ela adora experimentar coisas novas e até faz seus próprios cartões comemorativos usando o computador.

As personas primárias abrangem os grupos de usuários em que alguns membros da equipe não sabem muito sobre, ou que alguém da força-tarefa se sentiu compelido a escrever.

c. Estudos de caso e entrevistas

Desenvolveu-se os estudos de campo com entrevistas para explorar a área de acessibilidade e reconhecimento de imagem, assim como recolha de feedback qualitativos para desenvolvimento de soluções.

Procurou-se realizar um mapeamento de leituras aprofundadas sobre estudos de caso realizados por Sarah Horton (2013) e também por Laura Kalbag (2017), assim como entrevistas e estudos junto à pessoas como Karthik Mahadevan, co-fundador e chefe de Design da Envision, e Anirudh Koul, Head of AI & Research em Aira e fundador do Seeing AI e com a especialista em audiodescrição Fernanda Brahemcha (Entrevistas disponíveis em anexo).

Todo objetivo é possível a partir desse momento, o mundo atual é tecnologicamente avançado o suficiente para fornecer as ferramentas e habilidades necessárias na criação de produtos mais acessíveis. Quando se trata de inclusão, todos podem ser um pouco mais atentos, mostrando-se dispostos a modificar costumes e aprimorar hábitos. Nas bem colocadas palavras da educadora Regine M. Gilbert (2019), "estamos no caminho de construir uma visão para o futuro, não apenas para nossos futuros usuários, mas também para nosso futuro". Segundo ideias apresentadas pela autora, criadores devem se lembrar que além de serem usuários, também são agentes responsáveis na criação de um futuro para todos.

A empatia é um fator essencial para que produtos acessíveis sejam criados, é necessário além de perceber os ambientes em que se convive, modificá-los na prática. É preciso que tanto criadores quanto usuários questionem tudo que está estabelecido, pois se nem todas as coisas e espaços forem acessíveis, significa que uma parcela de pessoas está em detrimento de outras, e sobre isso deve-se perguntar o porquê. É muito importante que explorem-se possibilidades que sejam cada vez mais inclusivas, mais horizontais, de abrangência maior. Familiarizar-se com outras perspectivas mundo a fora apenas engrandece o conhecimento, possibilitando maior diversidade na criação de ideias e resolução de problemas.

Pode-se apenas imaginar até onde a Inteligência Artificial chegará no futuro. Hoje a Microsoft é uma das maiores empresas de construção de tecnologias focadas em acessibilidade, possuindo projetos como o *Seeing AI*. O Google também está trabalhando no uso de AI relacionada à acessibilidade, com produtos como o *Lookout*. Ambos funcionam de forma semelhante, com a capacidade de reconhecer objetos e o ambiente em tempo real (Gilbert, 2019).

Existem hoje alguns recursos que visam tornar o mundo mais acessível, tanto no mundo digital quanto físico. Um dos exemplos além do *Seeing AI* é *Orcam My Eyes*, um dispositivo criado para aumentar a autonomia das pessoas com deficiência, o aplicativo pode ler com precisão documentos impressos e está disponível em todas as 54 bibliotecas municipais da cidade de São Paulo (Torniero, 2019).

Lúcia Miyake, especialista em pesquisa e tecnologia assistiva, em entrevista com Gustavo Torniero (2019) argumenta que mesmo com todo o avanço na área, é visível que suas limitações se dão em respeito à falta de pesquisadores e recursos na área de pesquisa em TA e acessibilidade. Torniero (2019), também aponta que além disso a criação de produtos com inteligência artificial podem ter um custo alto caso sejam desenvolvidos exclusivamente para o

uso de pessoas com deficiência pelo fator da baixa demanda.

Porém a especialista apresenta uma visão otimista em relação ao futuro onde "A tecnologia assistiva está sendo cada vez mais conhecida e é uma área de boa aplicabilidade da inteligência artificial, além do design universal", diz. (Torniero, 2019)

I. Estudo de caso: Seeing AI, Microsoft

No final do dia, você precisa de alguém que seja apaixonado pelo projeto como você, pela missão e quem quiser trabalhar nesse campo deve procurar tecnologias emergentes e também uma verdadeira paixão para construí-lo. -Anirudh Koul

Atualmente o aplicativo *Seeing AI*, produzido pela Microsoft, permite com que pessoas com deficiência visual realizem tarefas do dia a dia através de seu uso. Ele é um dos maiores exemplos do benefício trazido por meios de exploração de técnicas de Inteligência Artificial e *Computer Vision* para a acessibilidade .

Seeing AI é um aplicativo gratuito que narra o mundo ao seu redor, foi projetado pelo expert na área Anirudh Koul para a comunidade deficiente visual e de baixa visão. O programa possui a capacidade de reconhecimento de imagem e de descrição de pessoas, textos e objetos.

Em entrevista autoral com Anirudh Koul em 2020, ele explica que o projeto foi iniciado durante um Hackathon, evento no qual programadores juntam-se na criação de um projeto em 24h.

Para Koul, existem três passos essenciais na área de tecnologia para criação de projetos de sucesso: primeiro, é necessário conhecer seu público para agregar valor ao projeto, ele deve possuir importância pessoal para o usuário; segundo, o projeto deve pensar no futuro, o que é algo que quase sempre envolve Machine Learning e Inteligência Artificial; terceiro, o projeto precisa ser excitante, único e algo que o criador sinta orgulho de falar sobre.

Para alcançar esses três objetivos é necessário ter uma equipe multi diversificada e qualificada, afirma Koul, projetos como o *Seeing AI* precisaram de pessoas que puderam ajudar no desenvolvimento da capacidade um do outro. O experiente engenheiro conta que é essencial

ter um designer à frente do projeto, pois sem *design thinking*³¹ e um design focado na experiência do usuário, é muito provável que o projeto não alcance seus objetivos. Ainda assim, para ele a qualificação necessária não é ser um técnico ou saber programar, vai muito além disso.

Koul iniciou o que viria a se tornar o *Seeing AI*, depois de ter construído um software de leitura de livros para seu avô. Esse projeto foi desenvolvido inicialmente como um *e-reader*³² e evoluiu para ser capaz de identificar códigos de barras em livros, tornando possível o usuário identificar qual livro tinha em mãos. Para criar um leitor de código de barras foi necessário criar um banco de dados de imagens de livros, para que com o auxílio do *Machine Learning* o computador pudesse processar e entender onde o código estaria localizado. Após alguns experimentos iniciais, sua equipe percebeu que a inteligência estava reconhecendo objetos como tapetes e estantes que ficavam no fundo das fotos, a partir desse momento o time passou a focar em fazer reconhecimento de objetos no geral.

Esses passos que pareceriam serem tomados ao acaso, foram na verdade resultado de muito esforço e tempo dedicados ao projeto, a cada momento a equipe descobria algo novo e conseguia evoluir no seu desenvolvimento. Na primeira vez que a inteligência para reconhecimento de objetos estava rodando em um celular, seu processamento levava quase 7 minutos. Koul compartilha que foram após muitos finais de semana trabalhando no projeto, que finalmente a equipe conseguiu evoluir o processamento para 90 segundos, então para 40 segundos e por fim, hoje o processamento gira em torno de 2 segundos.

³¹ O design thinking é um processo não-linear e interativo que equipes utilizam para entender usuários, desafiar suposições, redefinir problemas e criar soluções inovadoras para prototipar e testar. Retirado de Interaction Design Foundation, em Outubro de 2020, de <https://www.interaction-design.org/literature/topics/design-thinking>

³² Um e-reader, também chamado de leitor de e-book, é um dispositivo eletrônico móvel projetado principalmente para a leitura de e-books e livros digitais.

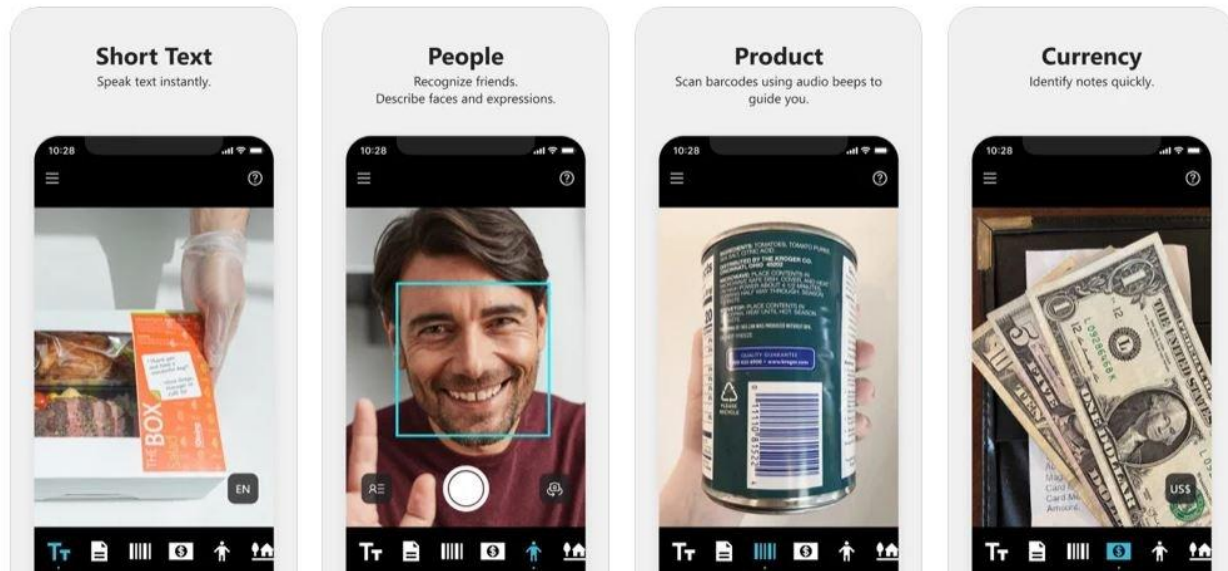


Figura 8. Imagens de Microsoft Seeing AI app (2020). Descrição de imagem: A imagem contém quatro elementos mostrando as principais features do app. No primeiro elemento mostra-se o recurso de texto, na imagem escrito “Short Text, speak text instantly” e mostra a imagem de uma embalagem dentro da aplicação. No segundo elemento mostra-se o recurso de reconhecimento facial, escrito “People, recognize friends, describe faces and expressions” e mostra a imagem de um rosto de uma pessoa de cabelos escuros, branco, com barba, sorrindo e com um quadrado azul de identificação do rosto. Terceiro elemento mostra-se o recurso de reconhecimento de códigos de barras, escrito “Product, scan barcodes using audio beeps to guide you”, e mostra imagem de uma lata de produto verde com um código de barras. Por último, mostra-se o recurso de identificar notas de dinheiro, escrito “Currency, identify notes quickly” e abaixo mostra a imagem de três notas com um, cinco e dez dólares.

A Inteligência Artificial pode causar grande impacto em todos os campos, como tem-se feito evoluções na detecção de câncer e na capacidade de reconhecer objetos e pessoas melhor do que a maioria dos humanos conseguiria. Anirudh Koul acredita que a Inteligência Artificial necessária para causar um grande impacto na sociedade pode já existir, no sentido de hardware e software, porém é preciso torná-las mais inteligentes e compreensíveis sobre o ambiente ao seu redor, para assim ser possível ensinar a máquina como entender o mundo que a cerca dentro da ética.

Ao questionar o engenheiro se *Computer Vision* e *Machine Learning* podem ser chaves

essenciais para o futuro da tecnologia assistiva, principalmente para pessoas com deficiência visual, Koul fala que sim e concorda que são ferramentas essenciais para o futuro, porém defende que para quem quer causar um grande impacto, definitivamente deve procurar uma combinação de AI com um design inteligente de usuário, algo que afirma faltar em projetos atuais.

No início, o *Seeing AI* não possuía um designer na equipe e com isso a interface não era muito compreensível, Anirudh Koul reconhece que foi com a chegada de uma designer para liderar a experiência que o projeto deu um salto. Em 2019 as técnicas de UX foram implementadas e ele conta que a sua equipe aprendeu que algumas das idéias que tiveram não eram válidas para o usuário final, e eles conseguiram entender isso graças aos milhares de e-mails que receberam de feedback do aplicativo (Koul, 2020).

Segundo Koul (2020), o *Seeing AI* foi apenas um pequeno passo que a Microsoft deu para as questões de acessibilidade e projetos voltados para comunidade. "Quando a Microsoft viu o sucesso dele, você sabe... Eles valorizam profundamente o AI para o bem social! Então eles se comprometeram com 165 milhões de dólares para dar às pessoas que participarem do projeto *AI for Good*." (Koul, idem). Na visão do criador, a Microsoft quer ter certeza de que os outros projetos tenham sucesso também, pois acreditam no bem maior para a sociedade.

II. Seeing AI User Story

História / Depoimento de Andre Louis, músico de Londres, usuário do *Seeing AI*:

Minha opinião sobre a vida é que a vida definitivamente está lá para ser tomada. O que você coloca é o que você sai. Duas coisas importantes para mim na vida são família e música. Ambos desempenham um papel enorme na minha vida... Eu tento ser o mais independente possível e a maneira como faço isso é principalmente com low-tech, que é uma bengala branca e hi-tech, que é um telefone.

Seeing AI é um das melhores tecnologia no momento, para usuários com deficiência visual. É um aplicativo que você pode ler praticamente qualquer

coisa em qualquer lugar. Nome de uma estrada, nome de um espaço, rótulo de maionese. Tem uma infinidade de funções e elas trabalham juntas para produzir algo que você pode usar sempre que precisar....

Na minha opinião, uma das melhores partes do *Seeing AI* é a capacidade de ler imagens das mídias sociais. Muitas vezes alguém me envia uma foto de algo que eu não consigo entender então vou e envio essa foto para o aplicativo do Seeing AI. Então eu vou ter uma descrição do que é: "Provavelmente um grupo de pessoas sentadas em um banco em um parque." "Pessoa." "Menino de oito anos de idade, com cabelos pretos, parecendo feliz." "Menina de sete anos parecendo enojada." Mudou a forma como posso ver as mídias sociais. Parece que posso participar. Gostar. Respostar. Muito mais do que eu podia antes...

Acho que ver a AI é realmente um grande passo tecnológico no caminho e espero que possamos continuar por um caminho desses por muitos anos.

Andre Louis,
Seeing AI Testimonial

III. Projetos de Referência

Até o momento, existem projetos e softwares que descrevem imagens e ajudam em melhorias de marketing e SEO. Existem também alguns projetos como Envision e Seeing AI, focados em descrever imagens e em acessibilidade, feitos para a utilização do usuário final, a pessoa com deficiência visual. Também existem alguns projetos com um foco global em auxiliar empresas com a automatização da acessibilidade. Este projeto tem como objetivo principal, portanto, auxiliar os responsáveis de projetos digitais a criarem imagens mais acessíveis, a partir da conscientização e da criação de ferramentas com foco em acessibilidade para imagens.

5. Fase: Criação

a. Branding

O conceito de branding deste projeto foi escolhido com base em cores que atingissem um alto padrão de contraste level AA, para que fosse possível a utilização de fontes tipográficas acessíveis como Georgia e Arial.

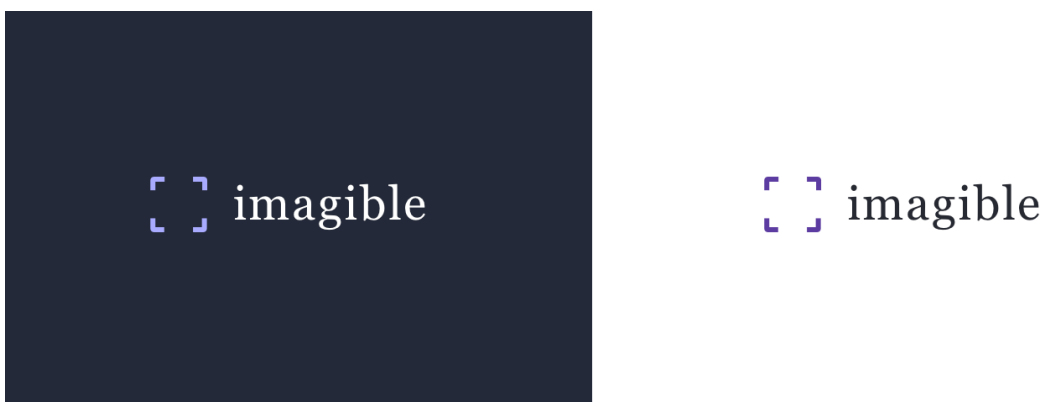


Figura 9. Logo e brand do Imagible.app. Descrição de imagem: A imagem contém um logo na versão noite e em um fundo escuro, assim como contém uma versão dia com a cor escura e fundo branco. A ideia do projeto é conter versão noite e dia da interface para uma melhor acessibilidade de contrastes de acordo com o horário do dia.

b. Regras de criação e desenvolvimento pré-estabelecidas

A aplicação, evidentemente, também foi criada com base em regras acessíveis na sua implementação, algumas delas foram:

- Contraste entre texto e fundo deve ser mínimo AA.
- Elementos nativos devem ser utilizados para implementação sempre que possível.
- Campos de texto e botões devem conter navegação por teclado e foco nos elementos.
- Imagens e elementos visuais devem conter descrições apropriadas em *alt-text*.
- Vídeos devem sempre conter closed captions.

- Imagens traduzidas por inteligência artificial devem conter campo de feedback e avaliação.
- Código fonte deve estabelecer clara hierarquia de conteúdos como H1, H2, H3 etc.
- A aplicação deve conter ARIA labels.

c. Arquitetura de informação

Arquitetura do projeto, que consta na página inicial com as funcionalidades de descrição de imagens e subpáginas. Diagrama de fluxo:

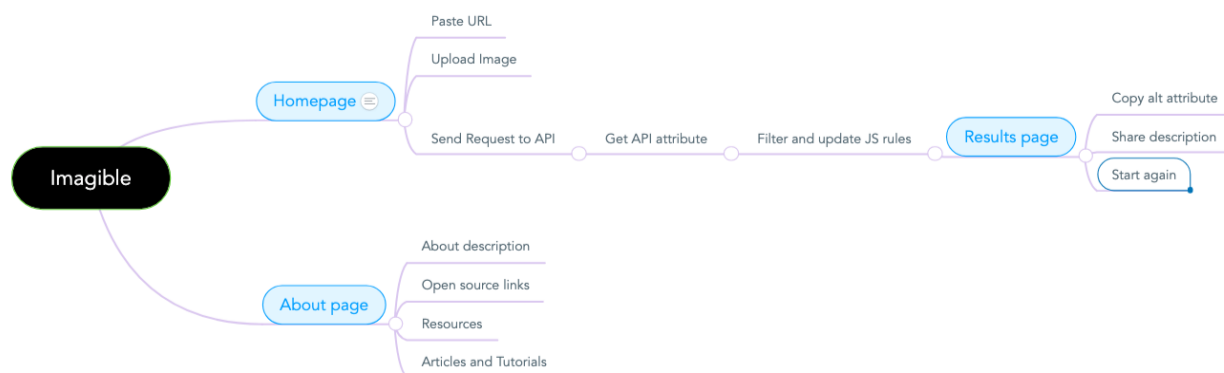


Figura 10. Diagrama de fluxo da aplicação Imagible.app. Descrição de imagem: A imagem contém um diagrama de fluxo inicial com Imagible.app segue para Homepage, que inclui: *Paste URL input*, *Upload Image input* e fluxo para descrição da imagem “*Send Request to API*, para *GET API attribute*, *Filter and update JS rules*” para por fim chegar no *Results page*, com funções de *Copy Alt attribute*, *Share description* and *Start again button*. Assim como mostra subpáginas como *About Page*, com *Description*, *Open source links*, *Resources pages* e *Articles and tutorials*.

d. Wireframes

Aqui algumas estruturas wireframes iniciais³³ para web e mobile:

³³ Para acesso completo de todo wireframes e saber mais sobre o projeto, acesse o arquivo: <https://www.notion.so/973f824047cf4834ad5a2bf235ebad53?v=b0833b20be6845a0b5b4e7de4b638ea9>

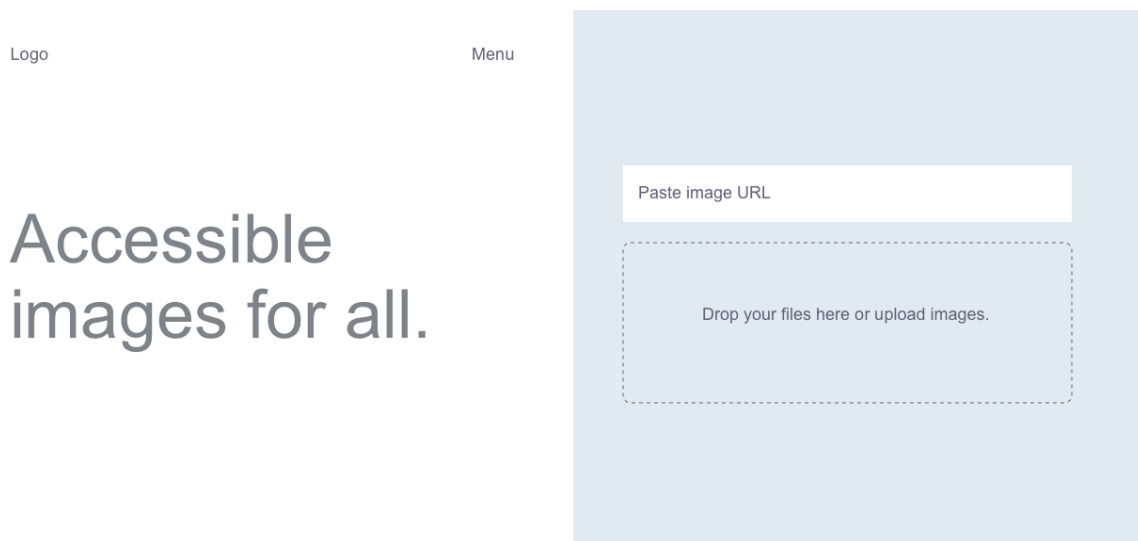


Figura 11. Wireframe da página inicial de Imagible.app. Descrição de imagem: A imagem contém ao lado esquerdo o logo, menu e um título escrito *Accessible images for all*, ao lado direito uma área de upload com input escrito “Paste image URL” ou “Drop your files here or upload images”.

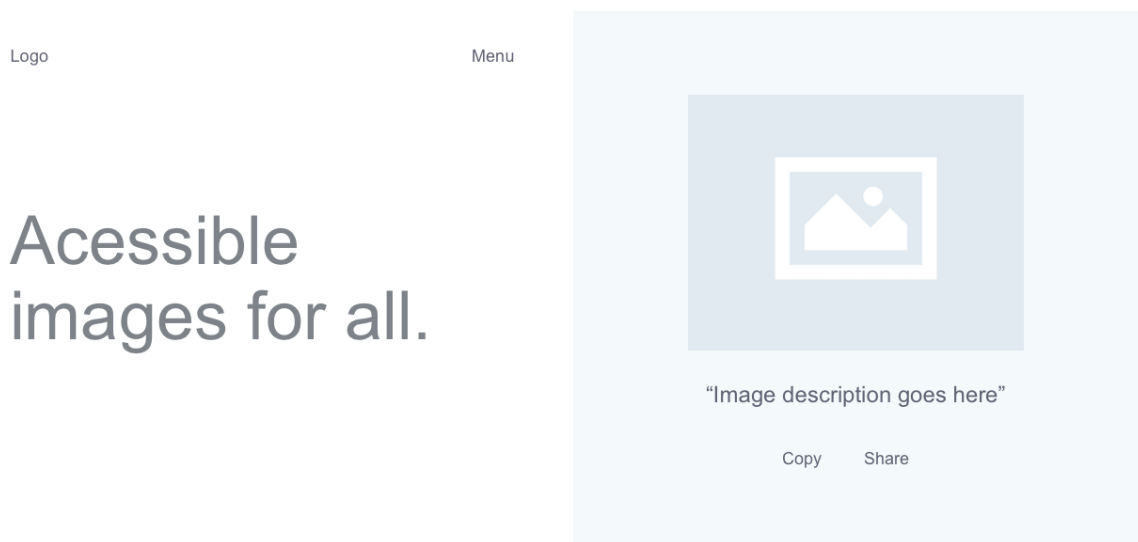


Figura 12. Wireframe da página de resultados Imagible.app. Descrição de imagem: A imagem contém ao lado esquerdo o logo, menu e um título escrito *Accessible images for all*, ao lado direito uma área de resultado do upload, contendo uma imagem vazia de exemplo e um título “Image description goes here” e dois links de ação abaixo para copiar ou compartilhar.

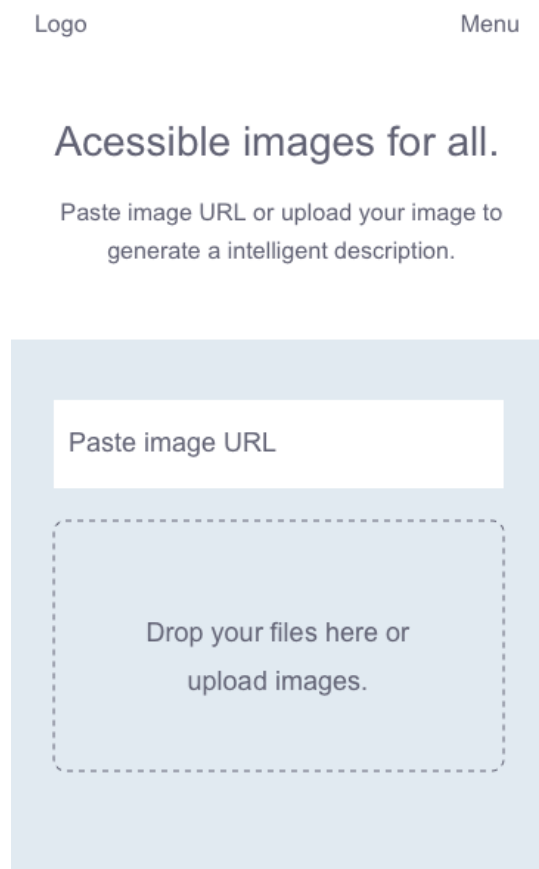


Figura 13. Wireframe versão mobile da página inicial de Imagible.app. Descrição de imagem: A imagem contém ao topo o logo, menu e um título escrito *Accessible images for all*, abaixo uma área de upload com input escrito “Paste image URL” ou “Drop your files here or upload images”.

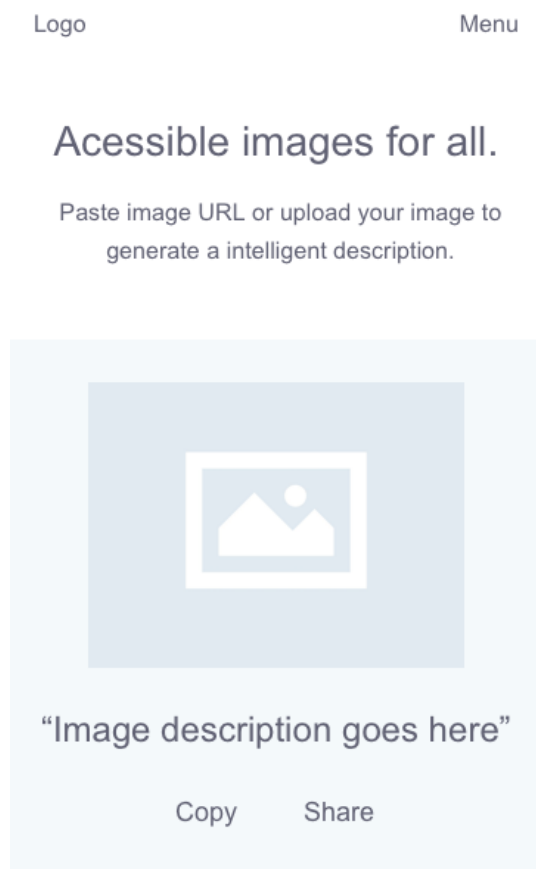


Figura 14. Wireframe versão mobile da página de resultados de Imagible.app. Descrição de imagem: A imagem contém ao topo o logo, menu e um título escrito *Accessible images for all* e abaixo uma área de resultado do upload, contendo uma imagem vazia de exemplo e um título “Image description goes here” e dois links de ação abaixo para copiar ou compartilhar

e. Protótipos

No protótipo do projeto consta o User Interface (UI) de como a aplicação web foi desenvolvida junto à evolução e alterações da interface que foram necessárias devido aos testes com usuários, aqui abaixo algumas telas³⁴ de exemplo utilizadas:

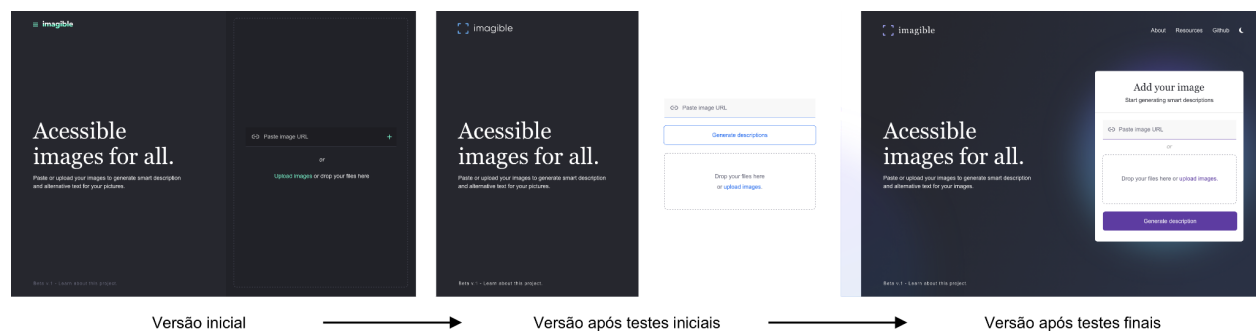


Figura 15. Evolução da interface antes, durante e após testes da aplicação Imagible.app.

Descrição de imagem: A imagem contém uma evolução em três etapas da interface, ao lado esquerdo a interface inicial antes dos testes em tons escuros, ao centro a interface durante os testes dividido em tons escuros e claros, ao lado direito a versão da interface após testes com um fundo escuro e contendo uma caixa delimitada em branco para área de upload.

³⁴ Para acesso completo de todas telas UI e saber mais sobre o projeto, acesse o arquivo: <https://www.notion.so/973f824047cf4834ad5a2bf235ebad53?v=b0833b20be6845a0b5b4e7de4b638ea9>

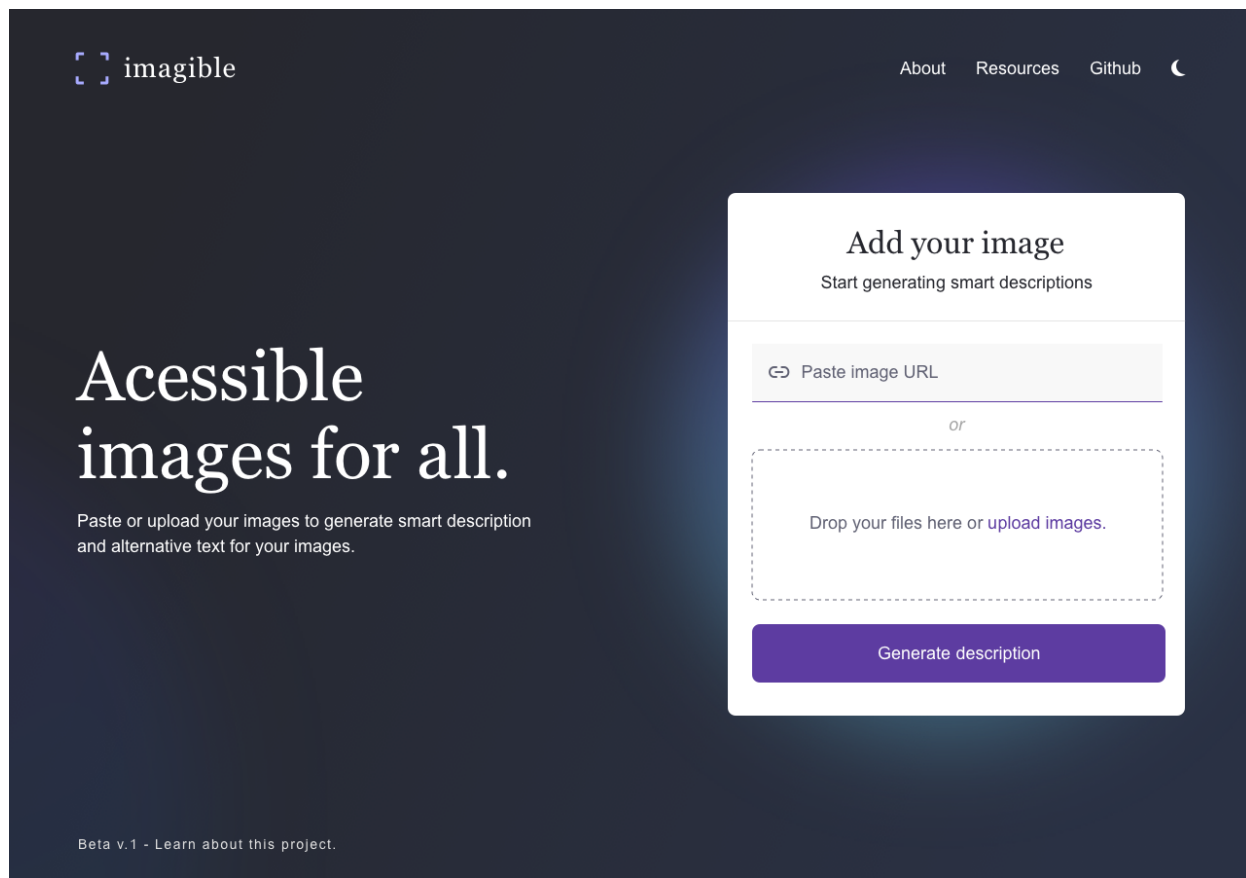


Figura 16. Página inicial Imagible.app. Descrição de imagem: A imagem contém ao lado esquerdo o logo, menu e um título escrito *Accessible images for all*, ao lado direito uma caixa delimitadora da área de upload, contendo um texto “Add your image, start generating smart descriptions” abaixo um input para colar a a URL da imagem e um campo de drag and drop para fazer upload de imagens, abaixo um botão “Generate description”.

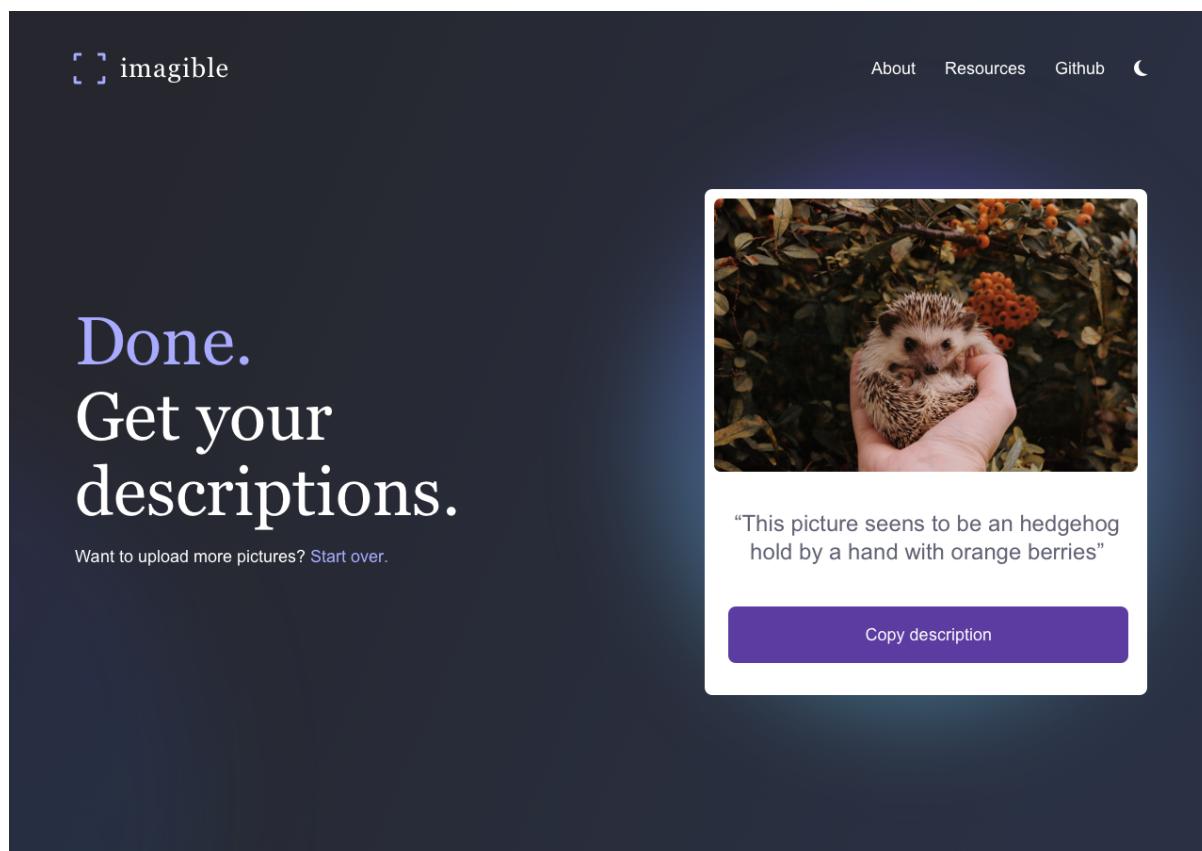


Figura 17. Página de resultados Imagible.app. Descrição de imagem: A imagem contém ao lado esquerdo o logo, menu e um título escrito *Done. Get your descriptions*, ao lado direito uma área de resultado do upload, contendo uma imagem exemplo uma descrição “This picture seems to be an hedgehog hold by a hand with orange berries” e dois ícones de ação abaixo para copiar ou compartilhar, com um botão abaixo para processar nova imagem.

f. Desenvolvimento e Implementação

O projeto utiliza React.js como linguagem principal e APIs com ferramentas da Microsoft Azure Computer Vision para criar descrições de imagens com maior precisão. Hoje em dia o texto alternativo necessita ser adicionado manualmente, a criação de alt-text automáticos com revisão é um benefício que surge a fim de otimizar o tempo.

O projeto não irá registrar ou guardar em nuvem nenhum dado de seus usuários. O potencial usuário irá colar a URL ou fazer upload de uma imagem na plataforma, então essa imagem será processada pela API com intenção de obter um string com a descrição da imagem,

onde o usuário poderá finalmente copiar, revisar e colá-la em seu projeto.

Para fins de desenvolvimento e com base em termos de inclusão, este projeto irá utilizar a API de Computer Vision da Microsoft. Diante de análises, o projeto classificou o produto da Microsoft como o mais neutro dentre outros, por mostrar que pessoas não devem ser sumariamente classificadas ao levar em consideração a diversidade multicultural no mundo e a complexidade do assunto, ao mesmo tempo que também existem iniciativas da mesma empresa que são alinhados ao tema central deste projeto, a inclusão digital. Aqui uma tabela de comparação entre as principais APIs analisadas em 2020 e a conclusão para escolha da tecnologia, baseado na tecnologia mais completa e com menor taxa de *bias*:

	Microsoft Azure Computer Vision API	Google Vision API	Amazon Rekognition	IBM Watson Visual Recognition
Face Detection	yes	yes	yes	yes
Face Recognition	yes	x	yes	x
Facial Landmarks	yes	yes	yes	x
Feature Detection	yes	yes	yes	x
Similar Faces	yes	x	yes	x
Emotion	yes	yes	yes	x
Label Detection	yes	yes	yes	yes
Landmarks	yes	yes	x	yes
OCR	yes	yes	yes	yes
Image Analysis	yes	yes	yes	yes
Video Analysis	yes	yes	yes	x
Custom Model	yes	x	x	yes
Natural Language	yes	x	x	yes
Gender Bias	low	low	high	high
Racial Bias	low	low	medium	medium

O reconhecimento de imagem realizado pela API escolhida é semelhante à classificação de imagens, mas neste caso a interface retorna as coordenadas da caixa delimitadora (em pixels) para cada objeto encontrado. Por exemplo, se uma imagem contém um cachorro, um gato e uma pessoa, a operação de detecção listará esses objetos em conjunto com as coordenadas na imagem. Essa funcionalidade permite processar as relações entre os objetos em uma imagem, não apenas identificar suas classificações, pois ela também determina se há várias instâncias com a mesma classificação em uma única imagem. Além disso a API também retorna a pontuação de confiança ou o nível de precisão em que cada classificação se encaixa, na qual a classificação pode variar entre 0 e 1, onde 0 é nenhuma (ou nível inexistente) e 1 é totalmente confiante do resultado. Nessa análise, qualquer resultado acima de 0.6 expressa um bom nível de precisão.

Dentro da mesma API da Microsoft também é possível executar Natural Language sobre o conjunto de itens processado em uma detecção de objetos, na qual se cria uma relação entre cada classificação e leva-se em consideração uma pontuação, ou de precisão ou de confiança, de cada descrição em linguagem natural, para assim criar descrições de imagens com uma linguagem legível por humanos. Em resumo, este serviço pode analisar uma imagem e gerar uma sentença que descreve seu conteúdo para o usuário, recurso que não é oferecido por outras plataformas.

6. Conclusão e validação

Nesta investigação foram realçados pontos sobre a imagem e a inteligência artificial como elementos centrais para a criação de ferramentas que auxiliem a acessibilidade. Com foco em explorar tecnologias como o computer vision, foi possível revelar uma poderosa ferramenta na criação de imagens acessíveis.

Os principais levantamentos sobre a imagem vieram das ideias de Aumont (1992) e Sartre (1996) que exploram o conceito de imagem não-visual, do imaginário, da imagem sensorial. Com essas ideias foi possível iniciar uma investigação sobre as imagens invisíveis na web e, sobretudo, iniciar uma relação entre o reconhecimento de imagem para a acessibilidade e a inclusão.

A relação da investigação que parte para explorar a acessibilidade como a porta para inclusão apoia-se nas ideias de Kalbag (2017) que traz a relação da acessibilidade e do design, apontando assim a importância da inclusão digital e de como projetar produtos para o mais amplo espectro de pessoas possível, através da empatia, é essencial no processo de entender e quebrar barreiras na busca por soluções.

Assim como também sustenta-se as ideias de Sarah Horton e Whitney Quesenbery (2013) que apontam a necessidade de projetar e criar produtos com acessibilidade em mente, pois as decisões de design tomadas em nome da acessibilidade e da inclusão podem beneficiar a todos.

Foi então possível explorar recursos de descrições de imagem, que abriram caminho para apresentar-se a relação entre o reconhecimento de imagem e a acessibilidade para pessoas com deficiências visuais, possibilitando a transformação de textos, descrições e legendas através do seu uso.

Dentro dessa relação exploraram-se iniciativas e tecnologias com o desejo de criar uma web mais inclusiva com o uso da inteligência artificial. Entrevistas e estudos sobre projetos como Center Human Technology, Envision AI e o Seeing AI da Microsoft, mostraram e revelaram que é possível buscar na tecnologia soluções para a acessibilidade e a inclusão.

Com base em projetos como esses, também explorou-se tecnologias relacionadas ao reconhecimento de imagem, o Computer Vision e a Inteligência Artificial, com o propósito de criar um projeto experimental que pudesse descrever objetos e imagens.

a. Resultados de testes

O projeto passou por fase de testes com participantes representativos das personas criadas para fornecer uma experiência de usuário que os beneficie. As metodologias de teste em prática foram *user testing* e *rapid testing*. Sendo o primeiro uma técnica de pesquisa utilizada para avaliar um produto ou serviço, onde os testes de usabilidade são realizados com usuários representativos do público-alvo e cada participante tenta realizar tarefas típicas enquanto o analista observa, ouve e anota. (Usability.org) Enquanto o segundo teste é uma técnica para coletar feedbacks rápidos e de forma remota, através do uso de ferramentas online para testes de usabilidade.

I. Participantes do *user testing*

Participante	Função
Participante 1	Criadora de Conteúdo
Participante 2	Desenvolvedor
Participante 3	Desenvolvedora
Participante 4	Designer de Produto
Participante 5	Designer

Nota: Tabela da função de cada participante do teste com usuário.

II. Requerimentos aplicados para o teste

- Número de testes: 5.
- Local das sessões: via Zoom.
- Software para registrar clicks e interações dos testes: Useberry.
- Acesso ao ambiente de test: www.imagible.app.
- Moderador: Ana Batista.
- Análises de testes: Ana Batista.
- Tempo estimado: 30-60 minutos.

III. Objetivo do produto a ser testado

- Criar uma web mais acessível.
- Auxiliar desenvolvedores a criar descrições de imagens e *alt-text*.
- Auxiliar designers a criar descrições de imagens.
- Auxiliar criadores de conteúdo a criar descrições de imagens.

IV. Barreiras

- Falta de consciência sobre acessibilidade.
- Falta de consciência sobre como descrever imagens.

- Falta de tempo em projetos com milhares de imagens para serem descritas.

V. Soluções

- Auxiliar a descrever e criar melhores descrições de imagens.
- Processar e automatizar descrições de imagens, com revisão humana.

VI. Motivações

- Para criar uma web mais inclusiva.
- Para criar melhores descrições de imagens na web.

VII. Premissas e crenças iniciais

- O projeto ajudará o passo de criar descrições e alt-text de maneira rápida e eficiente

VIII. O que não está em teste?

- Múltiplo upload de imagens, que será implementado em segunda fase.

IX. Parâmetro de sucesso para o teste

- O participante conseguir alcançar o objetivo final de fazer upload de uma imagem e coletar a descrição final.

X. Jornada de entrevista

- Introdução e explicações
- Perguntas introdutórias
- Performance de tarefas
- Perguntas específicas e feedbacks
- Finalização

XI. Tarefas para performance

Tarefas	Objetivos
---------	-----------

Antes de iniciar as tarefas, gostaria de saber como você descreve imagens nos seus projetos hoje e como a acessibilidade é tratada dentro do seu time.	Expectativa em feedback aberto para compreender como o participante lida com acessibilidade em projetos.
O que você entende por descrição de imagem e texto alternativo?	Expectativa em feedback aberto para compreender como o participante lida com descrições de imagens.
Vamos para primeira tarefa. Gostaria que você explorasse a página e expressasse sua primeira impressão.	Expectativa em feedback aberto para compreender primeiras impressões.
Continuando, você pode tentar descobrir como você pode fazer upload de uma imagem na aplicação?	Conseguir abrir uma imagem, achar os campos de upload.
Agora, gostaria que você processasse essa imagem e narrasse todos os passos do que você está fazendo.	Conseguir gerar uma descrição de imagem.
Por último, gostaria que você tentasse processar uma imagem a partir de uma imagem que está na web, sem fazer upload dela.	Conseguir gerar descrição de imagem a partir de uma URL.
Feedbacks finais	Coletar impressão geral, o que poderia melhorar e como.

Nota: Tabela das tarefas performadas durante testes com usuários.

XII. Principais resultados primários

Problema	Comentário	Ação sobre o projeto
----------	------------	----------------------

Um participante não conseguiu achar o campo para colar URL do projeto.	A área de upload pode não estar totalmente clara, um título ou flechas indicativas podem ajudar.	Precisa novo design e implementação dos ajustes. Status: novos designs foram implementados à interface.
Um participante não conseguiu entender o propósito da aplicação. Antes de explicações introdutórias.	Fica mais evidente que criadores de conteúdo necessitam de mais engajamento e conteúdos para entender o que são descrições de imagens. Ações como artigos e mais explicações dentro da plataforma podem ajudar.	É necessário criar uma parte educativa dentro da aplicação. Necessita design, conteúdo e implementação de novas ideias.
Participante questionou se poderia compartilhar o resultado diretamente nas redes sociais	Como implementação, seria interessante um botão de compartilhamento e uma nova integração com redes sociais para publicação.	Ideia necessita refinamento e avaliação de APIs para compartilhamento.
Algumas imagens, de cunho abstrato, foram descritas de maneira errônea pela plataforma durante o teste.	Como imagens abstratas são uma barreira ainda hoje para o computer vision, uma solução breve seria colocar um tag de level de confiança sobre o quão precisa a descrição foi gerada.	Ideia deve ser implementada.
Participante não relaciona a	Fica mais evidente que é	Ação: Adicionar ao website

ideia inicial de como funciona a aplicação após processar imagem.	necessário mais conteúdo e tópicos de aprendizado para entender o que são descrições de imagens.	parte de como funciona passo a passo. Ideia deve ser implementada.
---	--	---

Nota: Tabela dos principais resultados de testes com usuários.

XIII. Implementações futuras sobre testes

Item	Status
Adicionar botão de compartilhamento para redes sociais	Ideia está em validação
Criação de conteúdos educativos sobre acessibilidade dentro da plataforma	Ideia está em ideação
Tag com level de confiança na descrição	Ideia está em implementação
Desenvolver um plugin que conecta a inteligência da plataforma com screen readers	Ideia necessita ideação e refinamento para implementação.
Pacote NPM e biblioteca para instalação em React.js e Vue.js	Ideia como consideração futura. Feedbacks e testes mostram que descrições por computer vision necessitam de revisão humana ainda hoje, pois pode conter erros.

Nota: Tabela resultados de testes sobre possíveis implementações futuras.

XIV. Roadmap

Período	Eventos
2021 Q1	<ul style="list-style-type: none"> ● Fim dos testes de usabilidade ● Alterações e melhorias de interface

	<ul style="list-style-type: none"> ● Início do desenvolvimento de resultado de testes ● Lançamento do projeto
2021 Q2	<ul style="list-style-type: none"> ● Nova fase de testes por acompanhamento de dados ● Início de documentação, blog & social media do projeto ● Pesquisa e início para desenvolvimento de plugin para leitores de tela ● Validação de outros resultados e ideias para futura implementação

Nota: Tabela de Roadmap com implementações futuras.

b. Notas finais sobre resultados

Diante da análise do projeto apresentado, tornou-se evidente que alguns grupos de personas, como criadores de conteúdo, não estão familiarizados com a necessidade de se criar descrições de imagens, assim como designers e desenvolvedores assumem que esta não é uma responsabilidade ativa dentro do time e que por fim, quando acontece, foi porque alguém sentiu-se compelido a escrever.

Sendo as descrições de imagens frequentemente negligenciadas dentro de corporações e times de produtos digitais, tornou-se ainda mais clara a necessidade de ferramentas para tornar a web mais acessível, assim como fica evidente uma grande indispensabilidade de educar e instruir membros de equipes para uma total compreensão da importância da acessibilidade.

Portanto, é perceptível que a web não é um espaço acessível por diversos motivos, o que não significa que deve-se continuar dessa maneira. Com o apoio de ferramentas de inteligência artificial e novas tecnologias é possível tornar milhares de websites e imagens acessíveis.

O projeto Imagible nasceu por esse motivo, um experimento web capaz de descrever imagens com o intuito de auxiliar desenvolvedores, criadores de conteúdo na web e designers, na construção de textos alternativos mais eficazes. A partir daqui, o futuro o desenvolvimento do projeto será acompanhado de uma nova fase de testes por acompanhamento de dados, pelo início de documentação técnica, criação de um blog e redes sociais do projeto, assim como pretende-se

iniciar uma pesquisa e início para desenvolvimento de plugin para leitores de tela além da finalização e validação de outros resultados e ideias para futura implementação.

Por fim, este trabalho explora seus resultados como aprendizados na criação de projetos inclusivos e acessíveis. Principalmente a conclusão de que todos os projetos devem ser orientados aos usuários, e para isso é necessário incluir um amplo espectro de personas para ser desenvolvido juntamente aos usuários finais, reunindo constantes feedbacks que fazem com que suas vozes se tornem absolutamente essenciais para o sucesso do projeto.

Assim como a acessibilidade necessita estar em mente em todos os passos do design de um projeto, é preciso projetar intencionalmente com a acessibilidade e inclusão em mente, caso contrário se estará excluindo intencionalmente -- seguindo o pensamento de Stephen Frost, líder do Frost Included, empresa de consultoria para inclusão dentro de organizações, “A menos que você inclua conscientemente, você inconscientemente excluirá”³⁵. E mais do que isso, é necessário que os projetos sigam os padrões de acessibilidade WCAG e a legislação vigente, pois a acessibilidade é fundamentalmente um direito civil, e sendo assim é essencial que na busca de solucionar tais falhas, abranja-se os aspectos legais da acessibilidade na web.

Com esse experimento, finalmente, podemos ressaltar que a imagem invisível deve ser extinta para que a inclusão e a acessibilidade possam prevalecer no mundo digital. Para que isso aconteça são necessários projetos, regulamentações e estudos cada vez mais aprofundados, são necessárias intenções reais de mudança que proporcionem ações diretas na busca de tornar a web um local verdadeiramente inclusivo.

A relação entre imagem, acessibilidade e inteligência artificial pode ser a resposta não apenas para tornar imagens mais acessíveis, mas também para transformar a maneira com que desenvolvedores e usuários entendem as possibilidades de comunicação na web. Para muito além de apenas uma ferramenta de auxílio à inclusão, o entendimento da acessibilidade globalizada poderá proporcionar inúmeros desdobramentos no futuro das relações humanas. Se todas as pessoas puderem se comunicar diretamente, sem empecilhos, não se tornará mais fácil encontrar denominadores e objetivos em comum entre todos os participantes da sociedade, sem a exclusão de "pequenos" grupos em favorecimento de outros? Quando nenhum ser humano for impossibilitado de participar da web, talvez, através de um olhar otimista e esperançoso, a inclusão e a acessibilidade não sejam apenas temáticas com pouca importância social, mas se

³⁵ Frase retirada de entrevista com Stephen Frost em 15 de março de 2021: <https://www.learnerbly.com/articles/unless-you-consciously-include-you-will-unconsciously-exclude>

transformem em pilares fundamentais em qualquer espaço, virtual ou não, assim como já deveria ser há muito tempo. É necessário incluir acessibilidade como ponto essencial de projetos digitais, para que assim possamos abrir portas para a inclusão e tornar a vida em sociedade mais justa e inclusiva para todos.

Bibliografia

- Aumont, J. (2001). *A Imagem*. Campinas: Papirus.
- Aumont, J. (2004). *O olho interminável: cinema e pintura*. São Paulo: Cosac & Naify.
- Barthes, R. (1989). *A câmara clara*. Lisboa: Edições 70.
- Bergson, H. (1999). *Matéria e memória: ensaio sobre a relação do corpo com o espírito*. São Paulo: Martins Fontes.
- Bittencourt, G. (2001). *Inteligência artificial: ferramentas e teorias*. Florianópolis: UFSC.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Carmona, D. (2019). *The AI Organization*. O'Reilly Media, Inc. Ebook Kindle.
- Coppin, B. (2010). *Inteligência artificial*. Rio de Janeiro: LTC.
- Dondis, D., & Camargo, J. (2007). *Sintaxe da linguagem visual*. Martins Fontes.
- Holmes, K. (2018). *Mismatch: How Inclusion Shapes Design (Simplicity: Design, Technology, Business, Life)*. The MIT Press. Ebook Kindle.
- Horton, S., & Quesenbery, W. (2013). *A web for everyone*. Rosenfeld Media. Ebook Kindle.
- Kalbag, L. (2017). *Accessibility for everyone. A Book Apart*. Ebook Kindle.
- Nielsen, J., & Loranger, H. (2007). *Usabilidade na Web*. Elsevier. Ebook Kindle.
- Goldsmith, S. (2000). *Universal design*. Taylor & Francis Ltd.
- Gilbert, R. (2019). *Inclusive design for a digital world*. Ebook Kindle.
- Zuboff, S. (2019) *The age of surveillance capitalism*. PublicAffairs. Ebook Kindle.
- Alves, I. S. (2015). *Consciência imaginante e analogon: o lugar do objeto estético na obra de Sartre*. Recuperado em 15 de fevereiro de 2020, de <http://periodicos.pucminas.br/index.php/SapereAude/article/view/10992>
- Anirudh, Koul. *Seeing AI, Microsoft*. (fev. 2020). Entrevistador: Ana Carla Batista Corrêa. Lisboa, 2020. A entrevista encontra-se disponível em

<https://www.imagible.app/resources/interviews>

Bacelar, Dr. S. (2010). *Terminologia Médica - Imagenologia, imaginologia, imagiologia*. Recuperado em 10 de setembro, 2020, de <http://www.imaginologia.com.br/imagenologia/pdf/Imaginologia-ou-Imagiologia-ou-Imagenologia.pdf>

Bird, S., Barocas, S., Crawford, K., Diaz, F., Wallach, H. (2016). *'Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI'*, *Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT-ML)*. New York University. Recuperado em 20 de setembro de 2020, de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2846909

Campolo, A., Crawford, K. (2020). *'Enchanted Determinism: Power without Responsibility in Artificial Intelligence'*. *Engaging Science, Technology, and Society*, 6, 1-19. Recuperado em 30 de setembro de 2020, de <https://estsjournal.org/index.php/ests/article/view/277>

Crawford, K., Schultz, J. (2019). *'AI Systems As State Actors'*. *Columbia Law Review*, 119(7), 1941-1972. Recuperado em 22 de setembro de 2020, de <https://columbialawreview.org/content/ai-systems-as-state-actors/>

Crawford, K., Paglen, T. (2019). *'Excavating AI: The Politics of Images in Machine Learning Training Sets'*. Recuperado em 10 de junho de 2020, de www.excavating.ai

Crawford, K. (2013) *'The Hidden Biases of Big Data'*. *Harvard Business Review*, April 1. Recuperado em 28 de setembro de 2020, de http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html

Crawford, K., Calo, R. (2016). *'There is a blind spot in AI research'*. *Nature*, October 13. Recuperado em 03 de outubro de 2020, de <http://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805>

Garcia, C. (2008) *Sociologia da Acessibilidade*. Curitiba: IESDE Brasil S.A. Recuperado em 10 de Julho de 2020, de <http://www2.videolivrraria.com.br/pdfs/23923.pdf>

Hill, K. (2020). *The Secretive Company That Might End Privacy as We Know It*. New York Times, Published Jan. 18, 2020. Recuperado em 10 de outubro de 2020, de <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>

He, K. Zhang, X., Ren, S., Sun, J. (2015). *Deep Residual Learning for Image Recognition*. Recuperado em 30 de maio de 2020, de <https://arxiv.org/abs/1512.03385>

Krizhevsky, A., Sutskever, I., Hinton, G. (2017). *ImageNet classification with deep convolutional neural networks*. Recuperado em 24 de maio de 2020, de <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

Mahadevan, Karthik. Envision CEO. (fev. 2020). Entrevistador: Ana Carla Batista Corrêa. Lisboa, 2020. A entrevista encontra-se disponível em <https://www.imagible.app/resources/interviews>

Metcalf, J., Crawford, K. (2016). 'Where are the Human Subjects in Big Data Research? The Emerging Ethics Divide,' *Big Data & Society, special issue on Critical Data Studies*. Recuperado em 10 de outubro de 2020, de <http://bds.sagepub.com/content/3/1/2053951716650211.full.pdf+htm>

Microsoft. (2019, 02 de dezembro). *Seeing AI: Making the visual world more accessible*. Youtube. <https://www.youtube.com/watch?v=DybczED-GKE=>

Motta, Lívia. Ver com Palavras. (dez. 2020). Entrevistador: Ana Carla Batista Corrêa. Florianópolis, 2020. A entrevista encontra-se disponível em <https://www.imagible.app/resources/interviews>

Okumura, M., Rudek, M., Canciglieri, J., (2013). *Application of Assistive Technology in a Concurrent Engineering environment for the special products development: a case study*. In. *Concurrent Engineering Approaches for Sustainable Product Development in a Multi-Disciplinary Environment*. Springer London. Recuperado em 10 de outubro, 2020, de https://www.researchgate.net/publication/236168853_Application_of_Assistive_Technology_in_a_Concurrent_Engineering_Environment_for_the_Special_Products_Development_A_Case_Study

Pupo, D., Melo, A., Pérez, Sofia., (2015). *Acessibilidade: discurso e prática no cotidiano das bibliotecas*. Recuperado em 27 de junho, de http://eurydice.nied.unicamp.br/portais/todosnos/nied/todosnos/artigos-cientificos/livro_acessibilidade_bibliotecas.pdf.1.pdf

Richardson, R., Schultz, J., Crawford, K. (2019). 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice', *N.Y.U. Law Review Online*, 192. Recuperado em 24 de setembro de 2020, de <https://ssrn.com/abstract=3333423>

- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, J., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A., Fei-Fei, L. (2015). *ImageNet Large Scale Visual Recognition Challenge*. Recuperado em 16 de dezembro de 2020, de <https://arxiv.org/pdf/1409.0575.pdf>
- Santoro, A. (2002). *Que animais enxergam por meio de sons*. Recuperado em 5 de maio de 2020, de <https://bit.ly/3ikC2xm>
- Sartre, Jean-Paul. (1996) *O Imaginário - Psicologia Fenomenológica da Imaginação*. São Paulo: Editora Ática. Recuperado em 10 de agosto de 2019 de <https://www.scribd.com/doc/122055237/SARTRE-Jean-Paul-O-Imaginario>.
- Torniero, G. (2019). *Aplicativos para pessoas com deficiência: como a inteligência artificial pode melhorar a vida de milhões de brasileiros*. Recuperado em 10 de novembro de 2020, de <https://www.bbc.com/portuguese/geral-46685874>
- UsableNet. (2019) *ADA Web Accessibility Report*. Recuperado em 10 de outubro de 2020, de <https://cdn2.hubspot.net/hubfs/3280432/2019EndOfYearRecap.pdf>
- W3C WAI. (2005) *WCAG 2.0, Web Content Accessibility Guidelines 2.0*, Recuperado em 20 de Dezembro de 2019, de <https://www.w3.org/WAI/standards-guidelines/wcag/>
- Yang, K., Qinami, K., Fei-Fei, L., Deng, J. Russakovsky, O. (2019). *Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy*. Recuperado em 30 de maio de 2020, de <https://arxiv.org/abs/1912.07726>