Enabling Aging Workers to Excel in the Modern Job Market

Meera Vinod, Eileen Cahill, Shai Dhaliwal, Catherine Yu, Kedari Lahari Narra

Advised by **Professor Annalee Saxenian**
University of California, Berkeley School of Information
Master of Information Management & Systems 2023 Capstone Final Report

Table of Contents

Introduction

As the job market becomes increasingly digital, job websites have become a critical tool for job seekers of all ages. However, many of these websites are designed with a narrow focus on younger job seekers, ignoring the needs and preferences of older users. This lack of age inclusivity can create significant barriers for older job seekers, who may struggle to navigate complex interfaces or encounter ageist language and assumptions in job postings. Another area where job search based age discrimination is particularly evident is in job postings. Employers often use language that implies a preference for younger candidates or even excludes older workers entirely. This practice is not only unfair but also illegal, yet it persists on job sites across the internet.

In this project, we will explore two prototypes with which we try to promote age inclusivity on job websites:

1) improving website design to better meet the needs of older users
2) mitigating the impact of ageist language in job descriptions.

By prioritizing these approaches, job websites can become a more accessible and equitable platform for job seekers of all ages, while employers can benefit from a more diverse pool of candidates.

Problem Statement:

1) How can job websites be better designed to be more inclusive towards older job seekers
2) How can older job seekers apply to more jobs at age-inclusive companies and achieve better employment opportunities?

Research

We utilized the bulk of the Fall semester, and winter break/early Spring semester, to interview multiple stakeholders and reviewed literature to develop a better understanding and empathy for the aging worker lifestyle in the United States. Given that none of the team members are members of this demographic and with 3 out of 5 members being international students with little context into the aging worker needs in the United States, this was a really important process for us to go through.

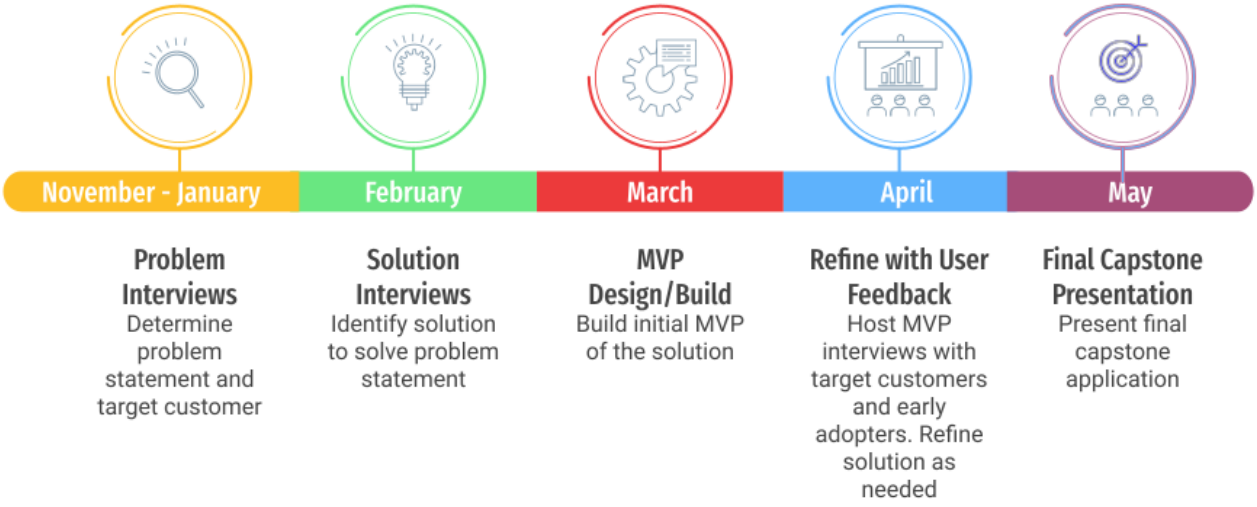


Figure 1: Capstone Timeline

We interviewed 24 people, ranging from senior citizens in different age groups, working professionals, directors of Berkeley vital aging groups, doctors, Berkeley senior citizen service representatives, an age-tech investor, and a fellow MIMS student who majored in Gerontology. In addition, a google survey was circulated on LinkedIn and Facebook which had 4 quality respondents. Team members also spoke with various local workers on the NextDoor application to gather information on the unique struggles they face navigating work while they age. We received very enthusiastic responses and willingness to be interviewed through this route.

These generative interviews and survey results have provided us with a clear path forward for our product.

Many seniors in the United States are postponing their retirement and deciding to continue staying in the workforce due to various reasons. The bad economy, rising inflation, inadequacy of their fixed income to sustain lifestyle goals are few of them. However, they enter an environment where age discrimination is rampant. Even in the absence of active discrimination, it is fair to say that no infrastructure exists to support older job seekers to find jobs considering their unique situation. We have summarized our findings, which fed directly into our prioritized set of product features in the chart below:

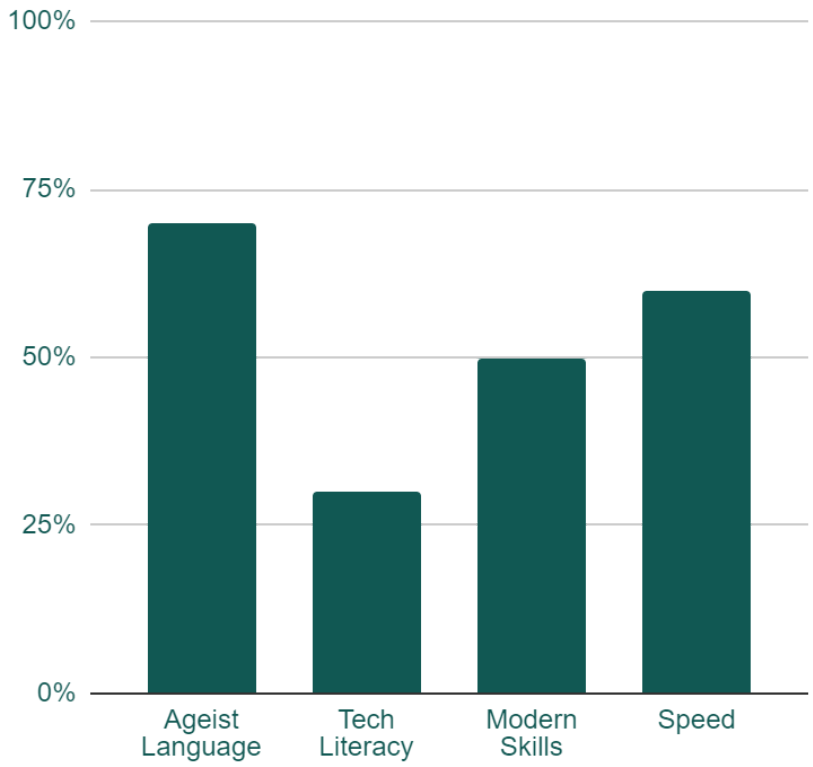


Figure 2. Top pain points for aging workers

Working professionals spanned across a wide variety of industries including healthcare, social services, technology and education which offered a depth of information and a specific product tailored to the needs of aging workers. This is a product we believe in.

Product Summary

Sage is a machine-learning/natural language processing powered browser plug-in that assists aging workers with managing the impacts of ageism in the job search process. The Natural Language Processing model easily sifts through job posts and detects age-unfriendly language. It is built into a larger front-end

browser plug-in that highlights the terms detected on a scale of least harmful to very harmful while also adding in a set of functionality that improves the user experience for older workers. We believe Sage will assist older workers navigating online sites and specifically enables aging workers looking to re-enter and navigate the job market with attention to the needs that the elderly face.

Our solution aims to address the following:

1. Difficulty re-entering the workforce after a gap in work
2. Trouble updating skill sets to match modern job descriptions
3. Struggle finding companies to apply where age discrimination does not exist
4. A learning curve in adjusting to conventions of tech-centric work culture

With Sage, users can assess whether a job post has harmful, age-unfriendly language that may negatively impact their potential employment. Sage also has a variety of helpful features that aid aging workers with navigating online sites to quickly and effectively complete online tasks. These include:

1. Help Buttons: users can ask for help when navigating through an online site. Pop-ups assist users on how to proceed with a given task.
2. Speech-to-Text
3. Font Size, Color Contrast Adjuster: adjust font/color contrast due to the symptoms of aging. This allows for a pleasant user experience when navigating online sites.
4. Declutter: reduce the amount of images and text on a site to focus only on the most critical information

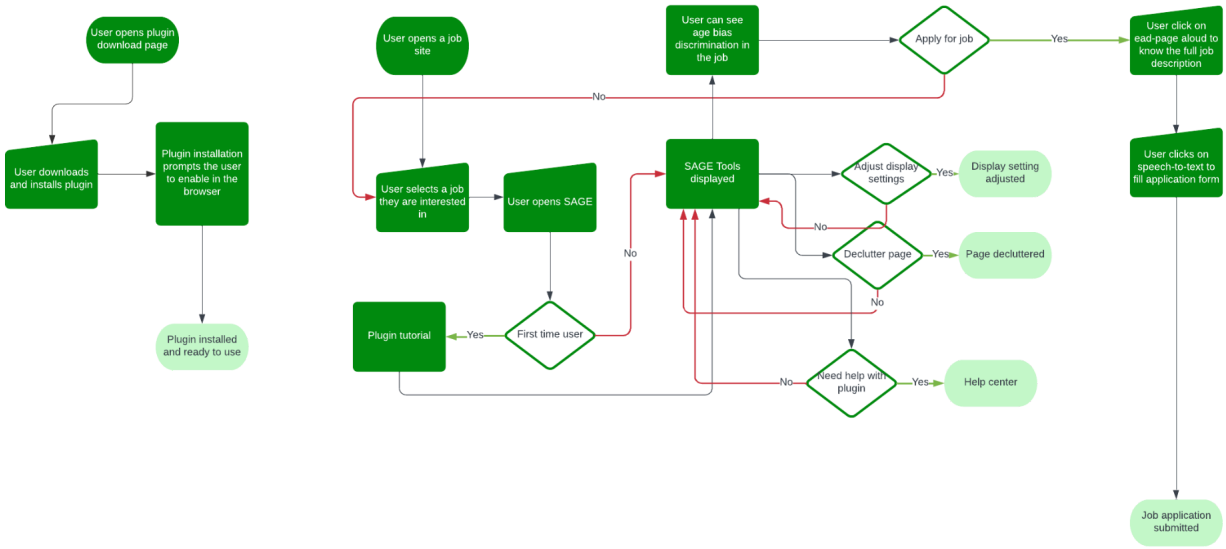
Sage aims to reduce the information asymmetry around age-bias in job descriptions; empowering applicants who deal with ageism while they apply for jobs. Sage not only detects & flags harmful age-unfriendly language but also provides tools that are designed to make the entire job application process easier, all within the same browser plugin.

Sage was researched, ideated, designed, and built by five masters students from the University of California, Berkeley’s School of Information as part of their final capstone project.

Methods

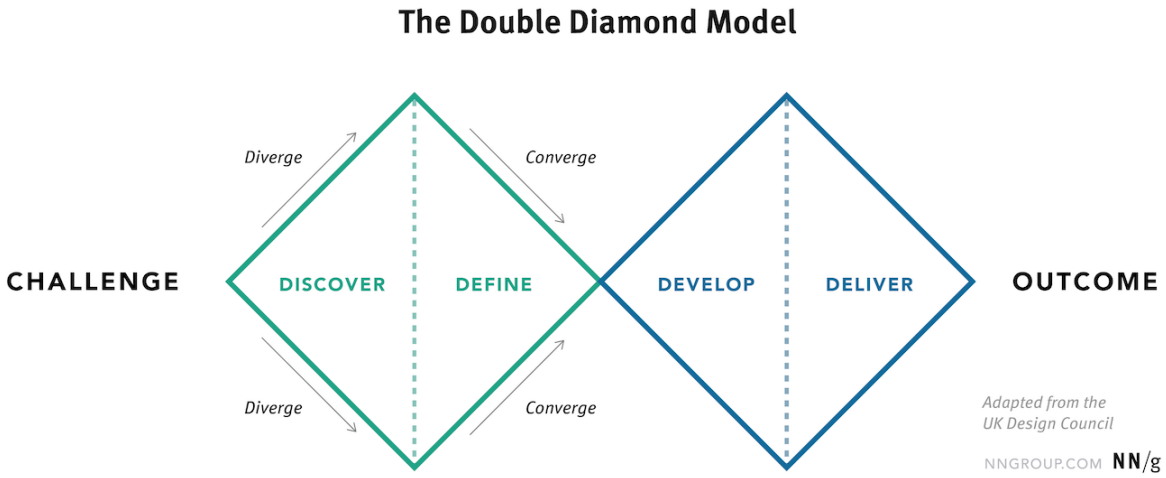
User Experience: Ideation & Design System

We started mapping out the user flow after finalizing on the core functionality of Sage. We created the flows to map out the steps that a user takes to complete a task and identify associated pain points and areas for improvement in our design.



[User flow diagram]

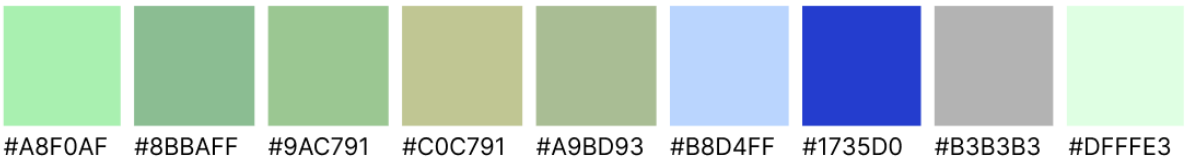
We then began sketching as it allows us to quickly and easily generate and iterate on ideas by focusing on user needs without being bogged down in technical details and visual aesthetics. This effort was part of the divergence phase in the second diamond of the design process. We sketched out multiple versions of the plugin, finalized the design for MVP and proceeded with low-fidelity(lo-fi) prototyping in Figma.



Lo-fi prototypes gave us the flexibility to focus on the functionality. As it was still the early stages of UI design, we were able to get honest feedback from other members of the team to make changes quickly instead of spending too much time on the visual design aspects.

As we were iterating on interaction design, we started working on the design language for the plugin. We researched extensively on ways to make designs accessible, especially for people in the 50+ age group[1] [2].

Colors



[1] https://healthfully.com/colors-elderly-8551183.html

[2]https://www.fonts.com/content/learning/fyti/situational-typography/designing-for-seniorseniors - Fonts.com | Fonts.com

Figure 3. Color palette, text sizing, and color contrast requirements

The palette is chosen to represent the energetic colors of green, play on the name of the plugin, maximize color contrast and highlight important information. The design system is built with accessibility in mind, keeping the font size minimum to 18 points and using a color palette passed color contrast and color blindness thresholds as per WCAG AAA standard.

We proceeded to create a design component library, a collection of pre-designed and reusable user interface components that can be used to create the high-fidelity prototype for user testing. Using a design component library ensures a consistent visual design across all pages. As all the UI components are designed with the same styles, colors, and typography it creates a cohesive and polished user experience.

Design component libraries also save time and effort by providing a pre-designed set of UI components that can be easily reused eliminating the need to design each UI component from scratch. Design component libraries make it easier to maintain a consistent design over time. This is because all the UI components are stored in one place and can be easily updated or modified as needed. This reduces the risk of creating outdated or inconsistent designs

User Experience: Design & Testing

High-fidelity designs

After establishing the design system and sketching out initial ideas of components and the user flow, creating the high-fidelity prototype was the next step. At this point, the plugin was defined as an age bias-detecting plugin with job application tools to foster the knowledge, speed, and productivity required to apply for jobs. Developing the high-fidelity prototype was done in Figma over the span of 7 weeks with weekly feedback from team members and a few colleagues. This was done to prevent progressing too far in the wrong direction. The features included in the plugin were decided based on existing tool offerings, previous user interviews we

had done, and desk research on what 50+ workers might want and need from a plugin. We landed on an age bias tracker, declutter page, speak to fill out forms, read page text out loud, adjust text display, and a help center. Although this plugin was designed with LinkedIn and Chrome in mind, the goal was once developed that this would work for any browser on any job application site from Indeed to USAJobs, and more.

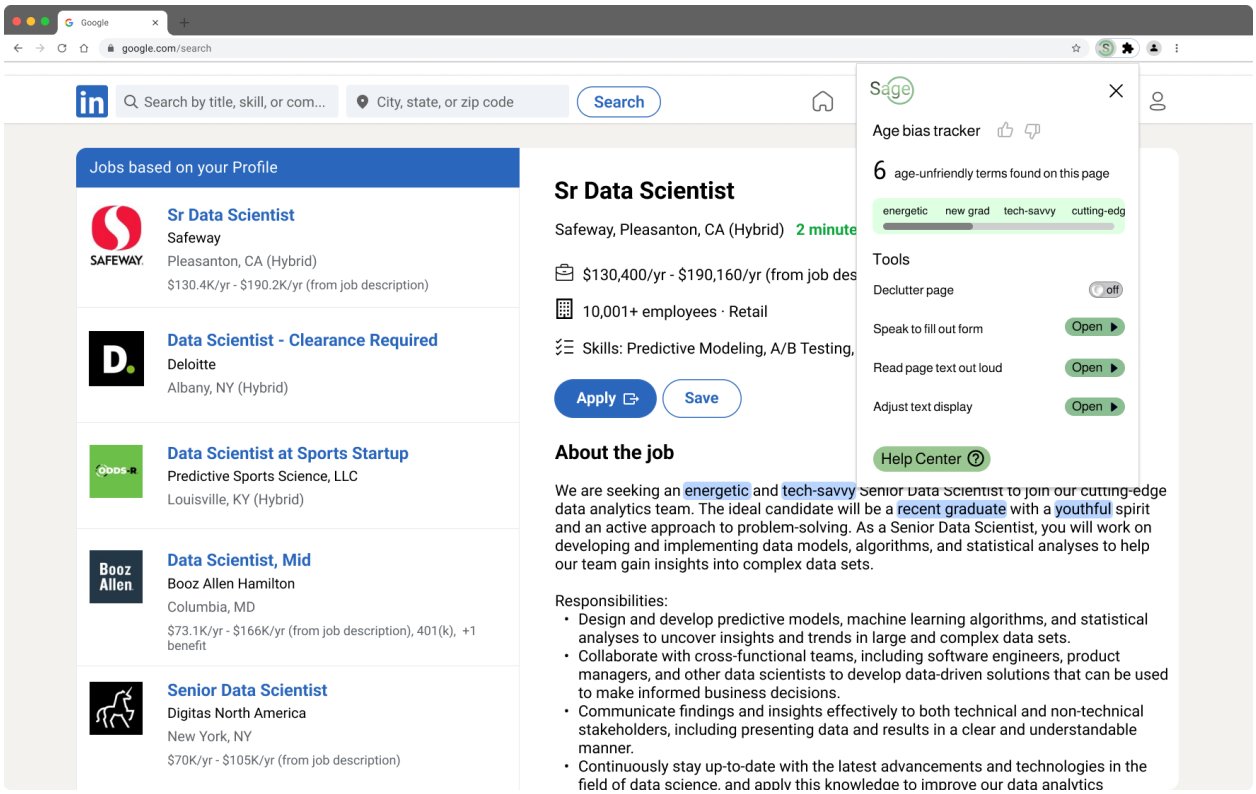


Figure 4. Landing page of Sage plugin

When a user searches for a job on LinkedIn, they open the Sage plugin in the Extensions section of Chrome (top right corner of screen indicated by puzzle piece icon). Once open the user sees the plugin pop up with the age bias tracker at the top and tools in the middle, with a Help Center button at the bottom.

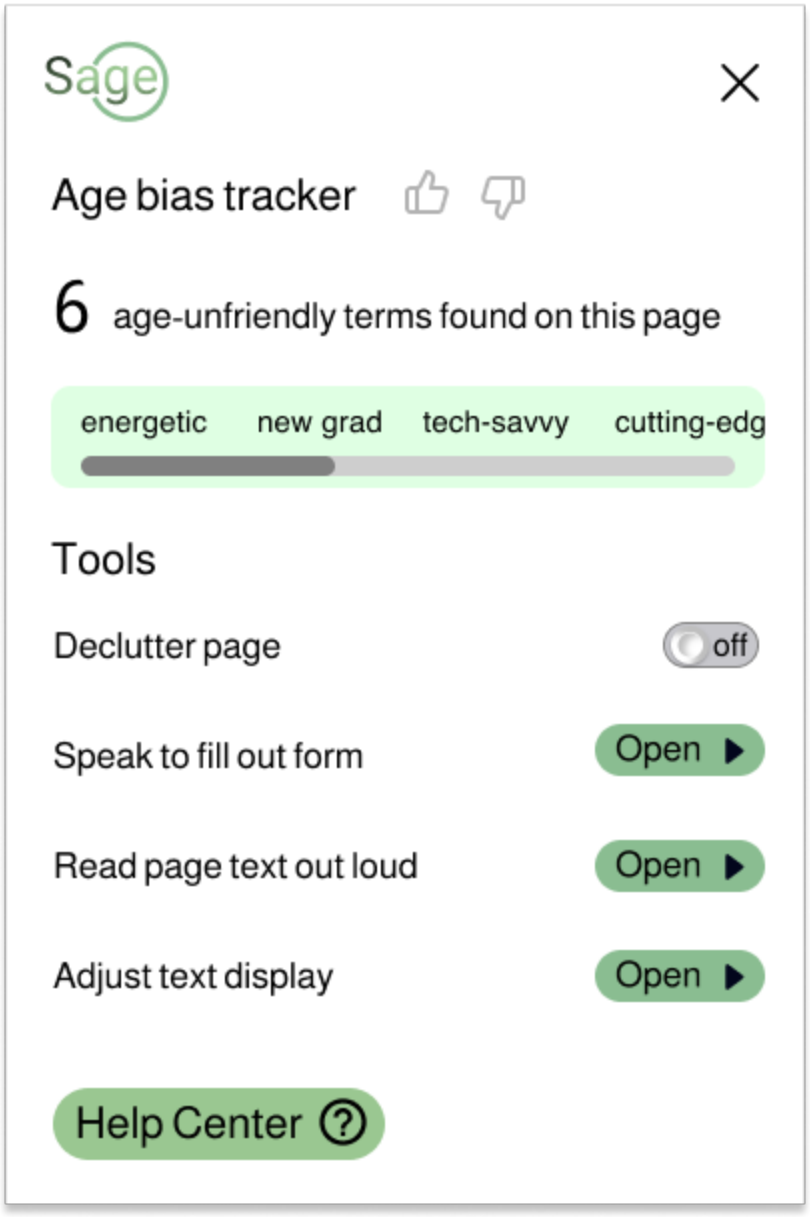


Figure 5. Age bias tracker utilizing NLP to inform users of potential age bias

The age bias tracker in the plugin shows the number of age-unfriendly terms found on the job post page. In this case, 6 terms are detected and all of the terms are listed underneath the total. A scroll bar is located in the section with the terms so that no matter how long the list of terms is, it can be kept within the same vertical distance. The terms are cut off toward the right to indicate that more information is present in that section. These terms are determined by the NLP model we developed, which detects age unfriendly language. This insight enables users to detect job posts and hirers which might discriminate against aging workers.

Watch a demo of the age bias tracker interactions here:
https://youtu.be/uzap0WQ9GYI

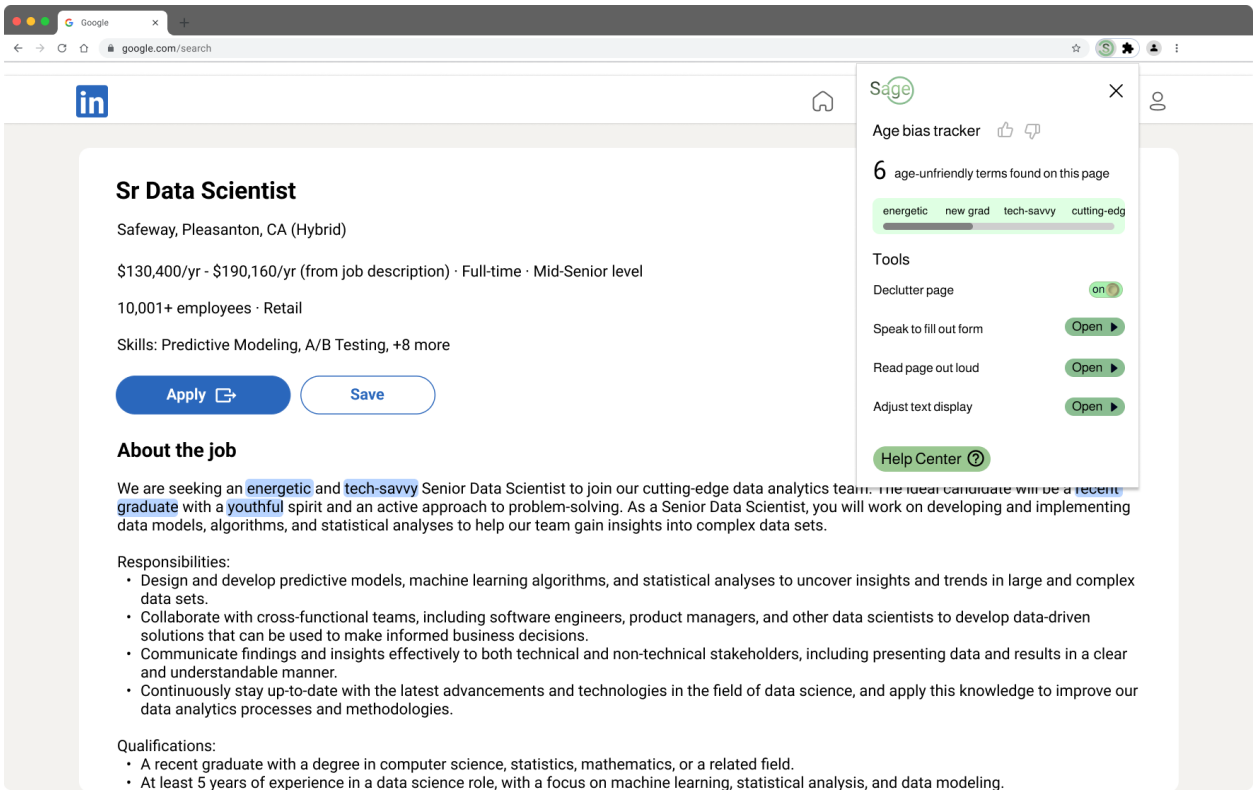


Figure 6. Declutter page tool simplifies LinkedIn job description

To offer a focused reading view for users, the declutter page tool removes all unnecessary elements except for the job title, features, apply and save buttons, and description. The user is able to reduce the amount of images and text on the LinkedIn page and allows the job description to take up the entire page to promote focus. The feature is included with the goal of reducing the cognitive load of aging users.

Watch a demo of the declutter interaction here:
https://www.youtube.com/watch?v=F4unAQR_GPk

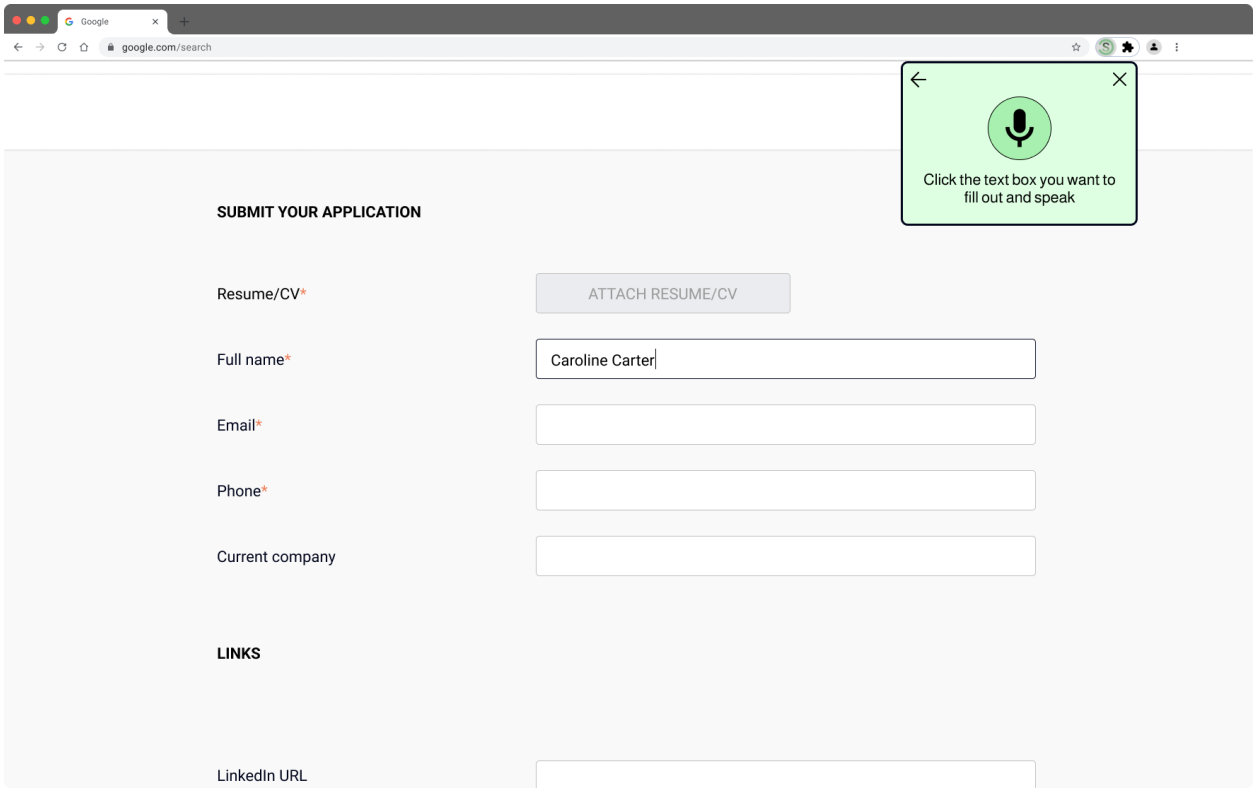


Figure 7. Speak to fill form tool

When filling out a job application form, the user can open the speak to fill form tool, which opens up a movable popup with a microphone button on it to turn on/off voice detection. Instructions on how to use the feature are included below the button, stating "Click the text box you want to fill out and speak". This feature helps those with hand issues, carpal tunnel syndrome, or who simply want to fill out forms using their voice.

Watch a demo of the speak to fill form interaction here:
https://youtu.be/OlkRjRsU7EM

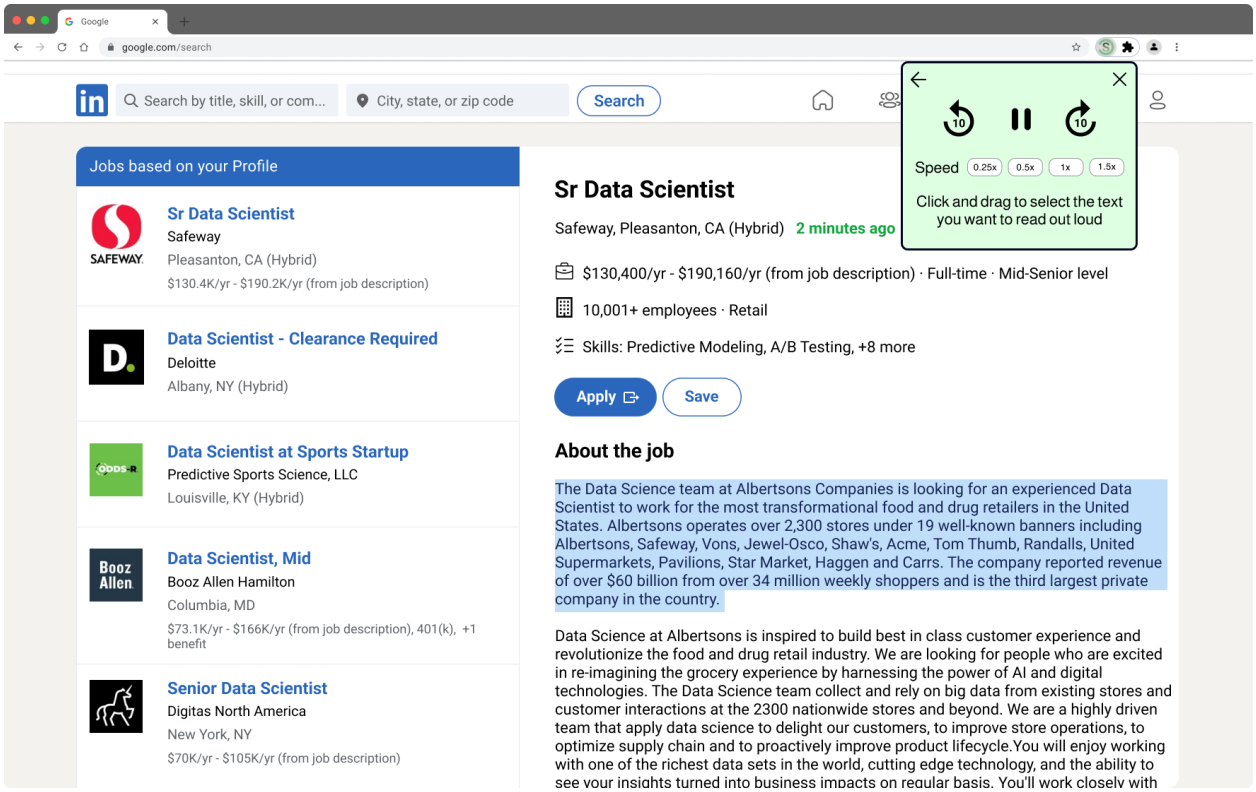


Figure 8. Read page text out loud tool

On a webpage with text, the user can open the read page text out loud tool, which opens up a movable popup with a play/pause button, a 10-second forward button and a 10-second back button, and varying speed buttons. Instructions on how to use the feature are included below the button, stating "Click and drag to select the text you want to read out loud". This feature aims to help those with vision problems or lower visual acuity to hear the text on a webpage, and not have to rely solely on their vision.

Watch a demo of the read page text out loud interaction here: https://youtu.be/qdd6tp80FWk

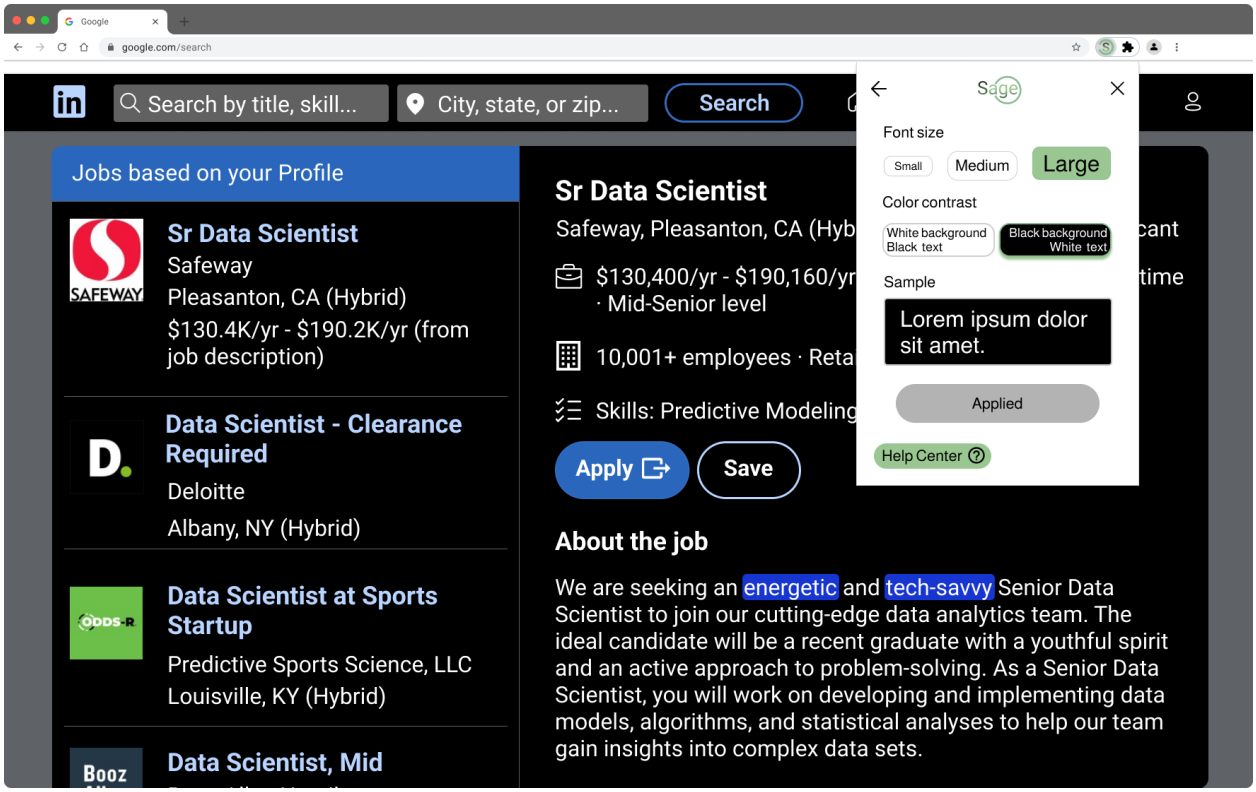


Figure 9. Adjust text display tool

On any job webpage, there is the option to increase or decrease the text size and to invert the color contrast. Color contrast allows for elements and text to be distinguished from each other while larger text size is typically more readable for anyone with eyesight degeneration. These are features that aim to improve the job search experience by taking into account symptoms of aging, such as reduced ability to distinguish colors and losing the ability to see things up close[3].

Watch a demo of the adjust text display interaction here: https://youtu.be/S8km8WctWZ0

[3] https://www.nia.nih.gov/health/aging-and-your-eyes

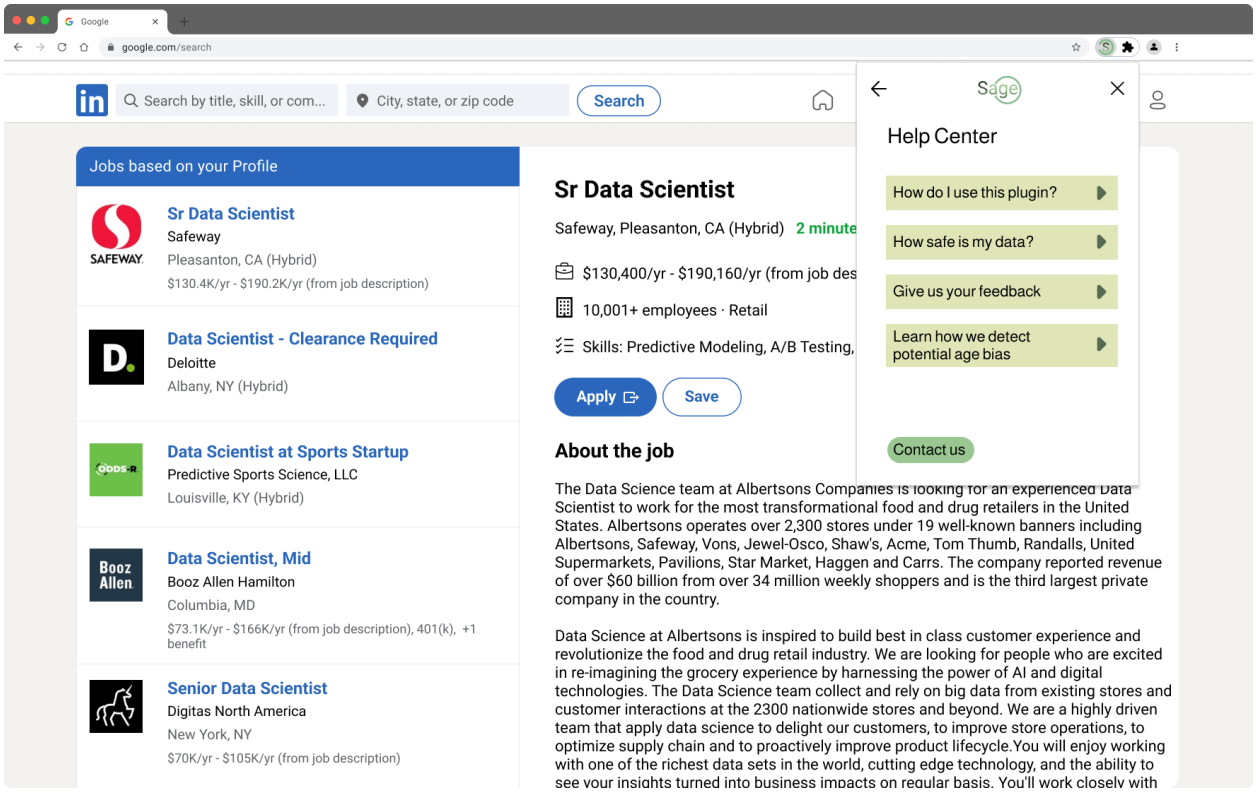


Figure 10. Help Center

The plugin allows aging workers to access help when navigating a site. The questions featured in the help center include “How do I use this plugin?”, “How safe is my data?”, “Give us your feedback”, and “Learn how we detect potential age bias”. These questions were determined based on which areas our team members thought it would be most useful to have information on and from what was heard in feedback throughout the design process.

Heuristic evaluation

Two team members performed a heuristic evaluation of the high-fidelity prototype before bringing it to usability testers (See Appendix Section B: Heuristic Evaluation). Nielsen’s heuristics were used to determine any issues with the prototype and rank their severity:

Nielsen’s heuristics:

#1: Visibility of system status

#2: Match between system and the real world

#3: User control and freedom

#4: Consistency and standards

#5: Error prevention

#6: Recognition rather than recall

#7: Flexibility and efficiency of use

#8: Aesthetic and minimalist design

#9: Help users recognize, diagnose, and recover from errors

#10: Help and documentation

The team members found that overall, the design had a couple of imperative issues to fix regarding help and documentation and system status visibility, which were all corrected before performing usability testing. Doing the heuristic evaluation allowed team members to view the design more critically and to troubleshoot existing issues with the flow, elements, and copy.

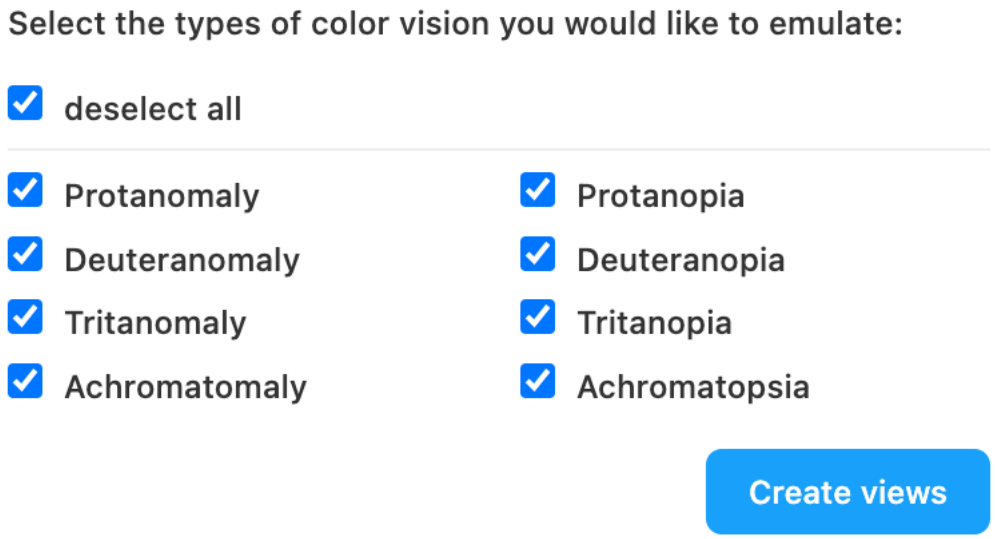


Figure 11: Assessing the high-fidelity prototype for color blindness accessibility

One key design driver throughout the entire design process was implementing accessibility for various visual disabilities. One of these disabilities is color blindness, which was checked that all text would be visible to users with various forms of color blindness. The prototype passed the color blindness check on all screens.

Usability testing

To get a sense of whether the plugin accomplished the goal of providing helpful insights to our users about potential age-biased language and offering a combination of productivity tools, we performed usability testing. Our designers tested the high fidelity prototype with 7 different people, 3 of whom were above age 50. The testing interviews were held remotely using Zoom, a link to the prototype was sent to the tester, and we asked the testers to perform a series of tasks and to provide their thoughts on the usefulness and usability of the plugin (see Appendix Section A for usability testing script).

All testers had confusion distinguishing between “Speech-to-text” and “Read page out loud” when asked to use the tool that would help them verbally write a job application. Speech-to-text is a term familiar to the field of accessibility, but not necessarily widely known, so the text was changed to “Speak to fill out form” to be more explicit. “Read page out loud” was then changed to “Read page text out loud” to further differentiate the two tools.

Among those who were above 50 years old, there was a general consensus that the tools were useful, but the ageism detection section needed a simple summary data point about the age bias detected to be useful. This finding greatly impacted our approach to the age bias tracker, as we transitioned from showcasing each age unfriendly word that was found and its severity to the number of age unfriendly terms found and simply listing which ones were found (Figures 12 and 13).

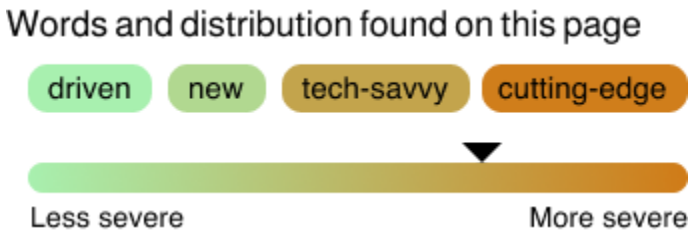


Figure 12. Previous version of age bias detection

Although this design initially was aesthetically pleasing to testers, two large problems were found with this way of showcasing age bias detection. Our NLP model did not have information about the severity of each age unfriendly word, so including this was not feasible. Secondly, it placed a strong cognitive load on the user to figure out not only how to interpret the information, but how to average the severity of all the words they saw into one data point.

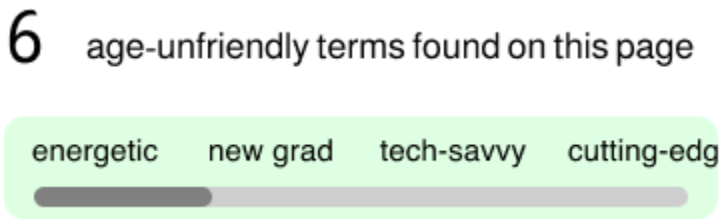


Figure 13. Final version of age bias detection

We landed on quantifying the number of age unfriendly terms on the page to ensure that the data point was apparent to the user. The severity legend in the previous version was removed so that more users will be able to grasp the information without requiring some knowledge of how to interpret a color gradient legend. To save space for any length of terms, the scroll feature was added so that the list of terms would take up the same amount of vertical space, no matter how many there were.

Among this age group, we also heard that an onboarding tutorial would be helpful to show the users how the plugin would be helpful to them. Rather than create an entire tutorial flow, our lead UI designer implemented an onboarding popup that would occur one time, when initially downloading the plugin, pointing the user to the Help Center for any confusion points (Figure 4).

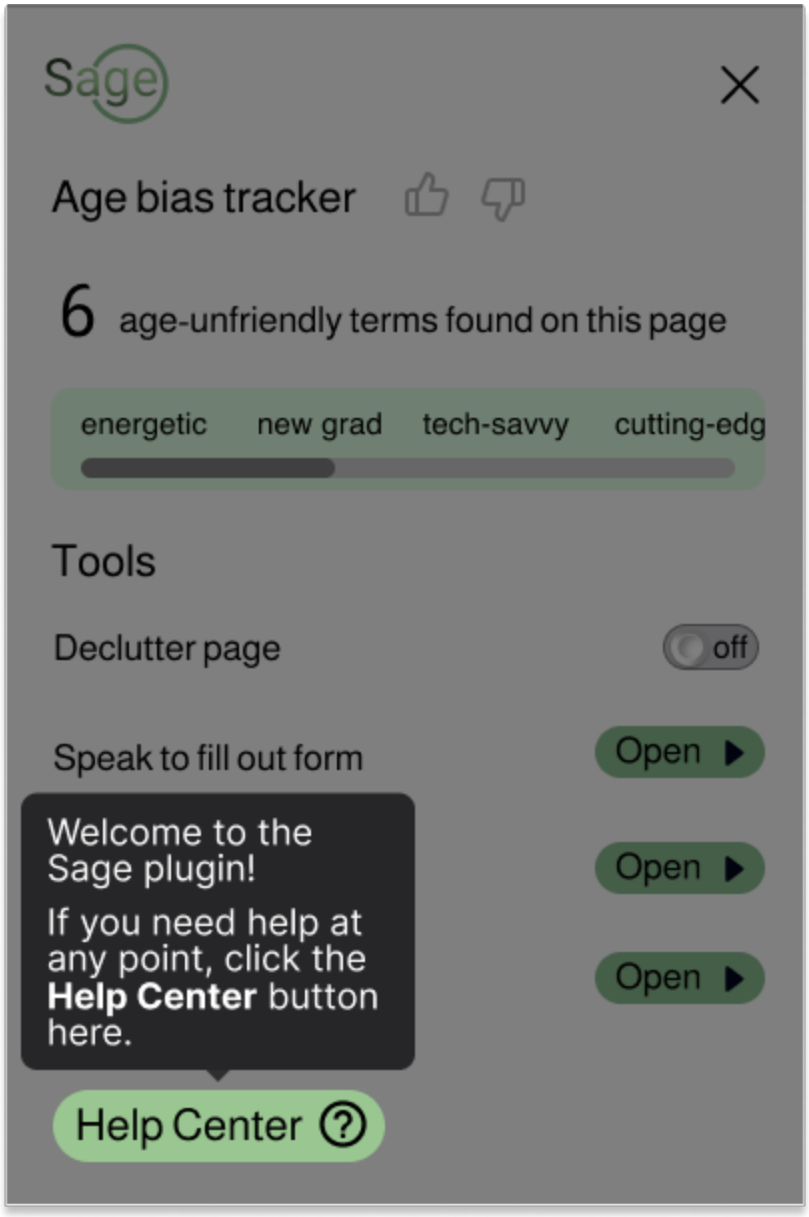


Figure 14. Sage plugin onboarding includes a one-time popup directing users to the Help Center

The usability tests were imperative to gathering insights about how users perceived our tool and what they expected from it. We were limited by the number of participants above 50+ that we had access to. More testing with the target age group would be needed to validate the aforementioned design changes and to continue to improve the plugin. There was some additional confusion about using a Figma prototype to do testing, where some of the features were not functional yet. Testing on an engineer-developed version of the plugin that could be integrated into the browser would yield useful insights.

Testers over 50 years old all thought that this kind of a plugin would be useful to them with some of the improvements addressed above. One participant said:

"This is good! Simplified for people like us who are not able to **keep up with the pace of tech advancements**. It is user friendly, and I believe **I can now apply easily**."

With this feedback, we believe that the plugin accomplished our goals of improving a user's speed and confidence in applying for jobs while simplifying the process and helping them to be more informed about age bias in job search.

NLP Functionality

A major component of this plugin is the machine learning algorithm that detects and flags age discrimination on job descriptions (JD) to help older candidates apply to companies that have the most age-inclusive policies. Studies on gendered wording in job ads (Gaucher et al., 2011[4]; Wille & Derous, 2018[5]) shows that minority candidates may feel discriminated against based on the phrasings of the job advert and that gendered wording commonly employed in job recruitment materials can maintain gender inequality in traditionally male-dominated occupations. Similarly, ageist wording in job recruitment materials could also maintain age discrimination in the workforce. Many jobs advertised on platforms like Linkedin, Glassdoor etc.. contain age discriminative language such as those looking for certain personality traits (young, agile), work experience (under certain years of experience), or physical traits (ex: able to sit on a chair for long periods of the day). While explicit discrimination in the form of expressing preference for younger candidates is illegal, linguistic choices in the JD serve to proxy that effect. Many older candidates are discouraged by JDs that consciously or unconsciously express a preference for a younger crowd, and end up not applying to jobs they are suitable for. We hope this section of the project will serve 2 purposes:

1) Help older candidates find jobs that are age-friendly.
2) Encourage companies to write more age-inclusive job descriptions in the long term

[4] Gaucher, D., Friesen, J. and Kay, A.C., 2011. Evidence that gendered wording in job advertisements exists and sustains gender inequality. Journal of personality and social psychology, 101(1), p.109.

[5] Wille, L., & Derous, E. 2018. When job ads turn you down: how requirements in job ads may stop instead of attract highly qualified women. Sex roles, 79, 464-475.

We implement the model in the form of an NLP (Natural language processing) algorithm that verifies the text in the JD, and highlights ageist words within the plugin interface. It also provides users with context about the intensity of discrimination of the words through a slider scale, thus helping them rank and choose more age-friendly jobs (Figure 4).

Below is a description of the working of the model in detail.

The model

The goal of this project is to build a model that helps mitigate ageism in job ads by detecting the ageist terms in job descriptions and requirements. The use of NLP in the recruiting process is not new but more from a resume screening side. In the study of Wings et al. (2021)[6], the researchers use both context-free and context-aware embeddings to build a model that classifies whether a given sentence contains soft skills, hard skills, or no skills. The results of their works show that using contextualized word embeddings, BERT in this research and part-of-speech tagging (POS) as the features, along with logistic regression provide the best results. Referring to prior studies, we use pre-trained BERT model to extract features from the sentences of each job posts. The job descriptions are first sentence-tokenized using SpaCy, with special characters removed. Given that ageism is highly context-dependent, we use the ”bert-base-uncased” pre-trained BERT model as a contextual word embedding method. The entire sentence will be loaded into BERT. POS tags are also added as a feature. Dummy variables will be used to represent the POS tags. After generating the POS tags for a sentence, the tags were counted and added as features. The features are then passed to a Logistic regression classifier to perform binary classification.

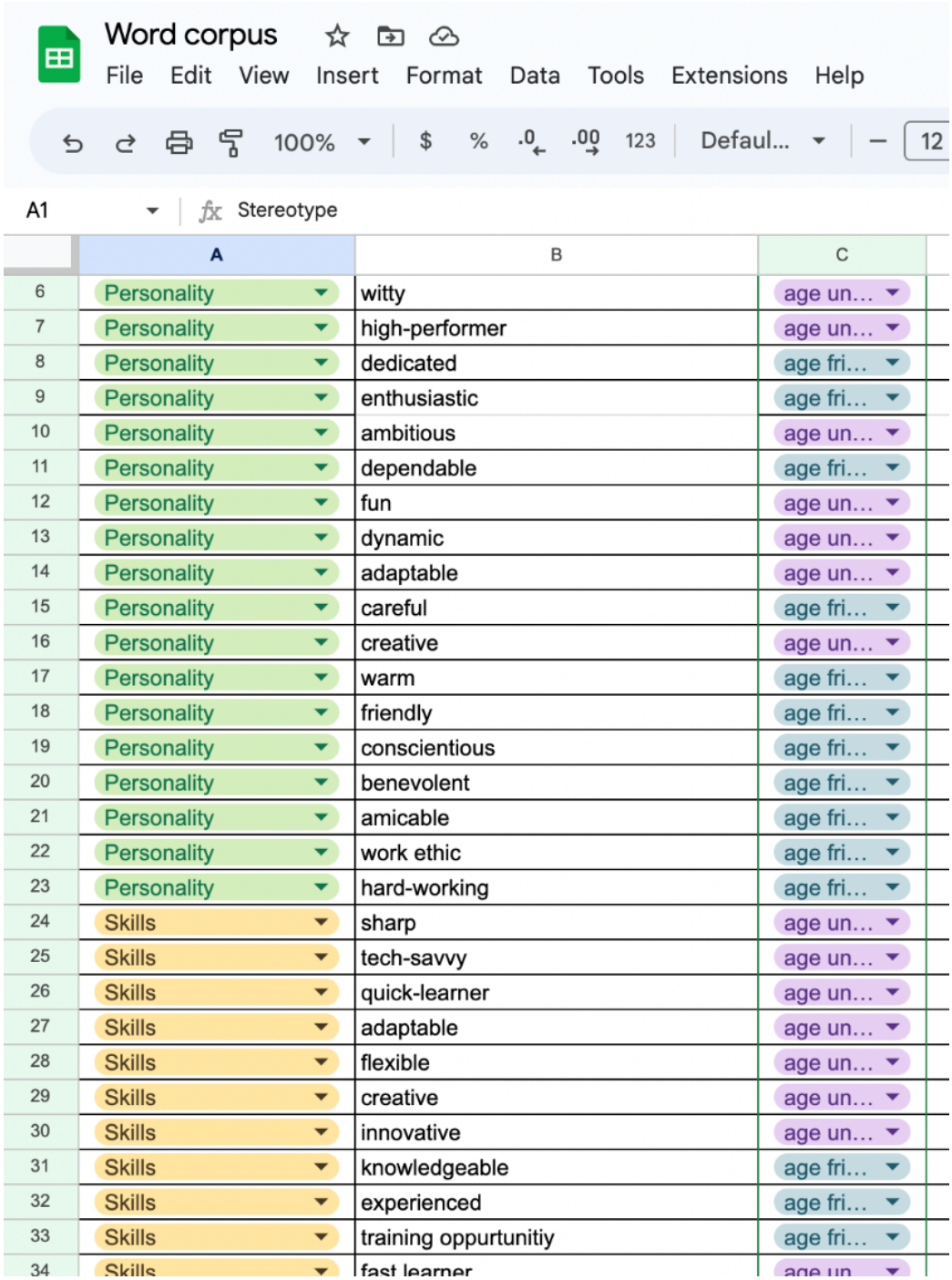
Annotation data

In order to train the model we sampled our data from 2 main datasets: the EMSCAD dataset which contains about 17,000 real world job posts, and a dataset from Data.world which contains 10,000data scientists job descriptions. Both these datasets contain job information from companies in the United States.

In order to identify ageist language, we first created a word corpus containing frequent ageist words used in JDs. We categorized them based on personality, skill sets, and age based discriminatory language. Extensive literature research and qualitative interviews were conducted to ensure an exhaustive list of words. We

[6] Wings, I., Nanda, R., & Adebayo, K. J. 2021. A context-aware approach for extracting hard and soft skills. Procedia Computer Science, 193, 163-172

then annotated over 800 lines of text from various job descriptions in these data sets in order to create a training dataset.



Word corpus

File Edit View Insert Format Data Tools Extensions Help

100% $ % .0 .00 123 Defaul... 12

A1 Stereotype

	A	B	C
6	Personality	witty	age un...
7	Personality	high-performer	age un...
8	Personality	dedicated	age fri...
9	Personality	enthusiastic	age fri...
10	Personality	ambitious	age un...
11	Personality	dependable	age fri...
12	Personality	fun	age un...
13	Personality	dynamic	age un...
14	Personality	adaptable	age un...
15	Personality	careful	age fri...
16	Personality	creative	age un...
17	Personality	warm	age fri...
18	Personality	friendly	age fri...
19	Personality	conscientious	age fri...
20	Personality	benevolent	age fri...
21	Personality	amicable	age fri...
22	Personality	work ethic	age fri...
23	Personality	hard-working	age fri...
24	Skills	sharp	age un...
25	Skills	tech-savvy	age un...
26	Skills	quick-learner	age un...
27	Skills	adaptable	age un...
28	Skills	flexible	age un...
29	Skills	creative	age un...
30	Skills	innovative	age un...
31	Skills	knowledgeable	age fri...
32	Skills	experienced	age fri...
33	Skills	training oppurtunitiy	age fri...
34	Skills	fast learner	age un...

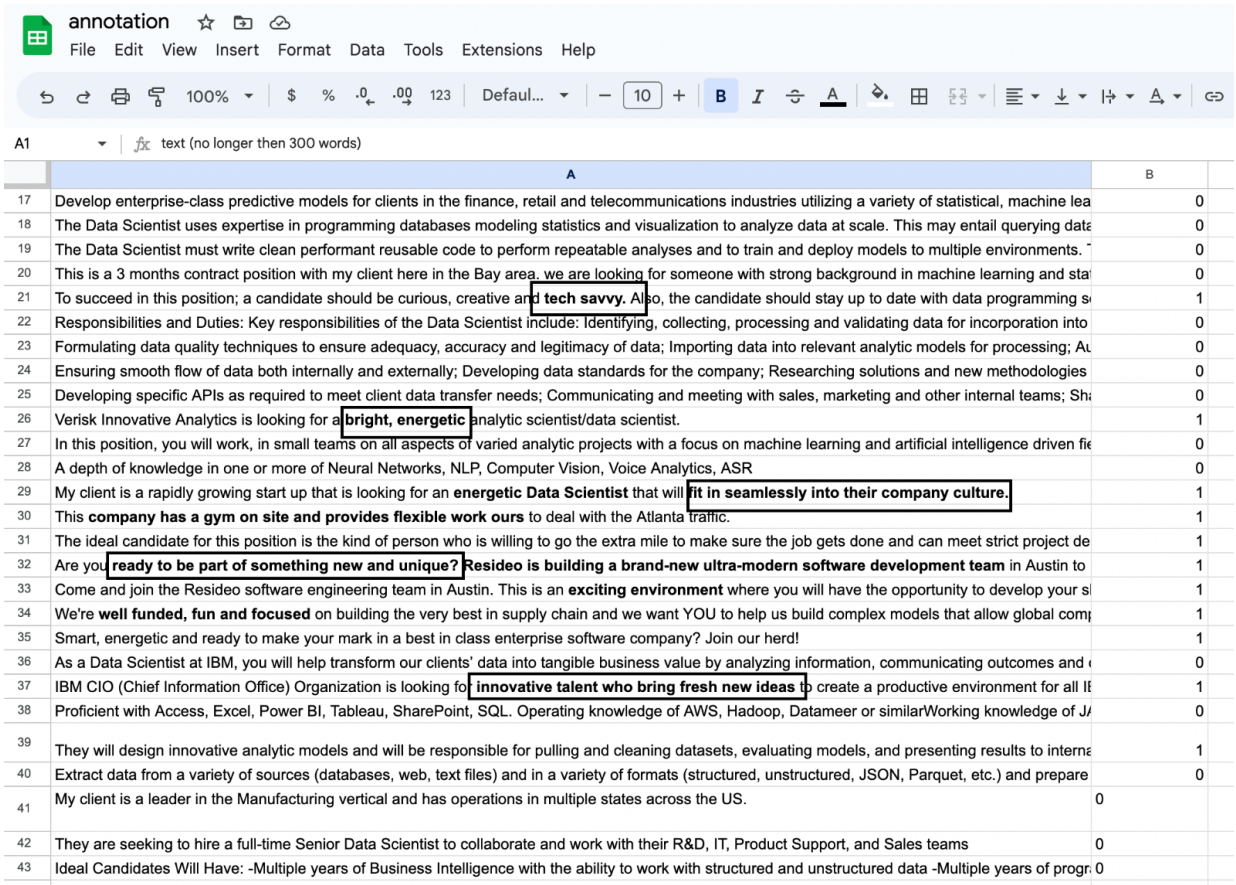
Figure 15: Age discriminatory terms

The training dataset depicts various types of age discrimination. For example: “We’re seeking someone who is fun, focused and can keep up with the demands of

a dynamic environment” is an example of ageist language that could be interpreted as discriminatory towards an older person’s lifestyle preferences (coded as personality type discrimination in our word corpus)

Type of discrimination	***Example usage***
Personality	You are a team player with an absolute winning mentality
Physical capability	Must be willing to sit continuously for 8+ hours a day
Lifestyle interests	You’re a digital native with a keen interest in pop culture
Skill sets	Must be tech savvy and a digital native
Age	Looking for an energetic, ambitious self starter who’s not afraid to question the status quo

Table 1: Types of discrimination and examples



	A	B
17	Develop enterprise-class predictive models for clients in the finance, retail and telecommunications industries utilizing a variety of statistical, machine lea	0
18	The Data Scientist uses expertise in programming databases modeling statistics and visualization to analyze data at scale. This may entail querying data	0
19	The Data Scientist must write clean performant reusable code to perform repeatable analyses and to train and deploy models to multiple environments. T	0
20	This is a 3 months contract position with my client here in the Bay area. we are looking for someone with strong background in machine learning and stat	0
21	To succeed in this position; a candidate should be curious, creative and **tech savvy.** Also, the candidate should stay up to date with data programming s	1
22	Responsibilities and Duties: Key responsibilities of the Data Scientist include: Identifying, collecting, processing and validating data for incorporation into	0
23	Formulating data quality techniques to ensure adequacy, accuracy and legitimacy of data; Importing data into relevant analytic models for processing; Au	0
24	Ensuring smooth flow of data both internally and externally; Developing data standards for the company; Researching solutions and new methodologies	0
25	Developing specific APIs as required to meet client data transfer needs; Communicating and meeting with sales, marketing and other internal teams; Sha	0
26	Verisk Innovative Analytics is looking for a **bright, energetic** analytic scientist/data scientist.	1
27	In this position, you will work, in small teams on all aspects of varied analytic projects with a focus on machine learning and artificial intelligence driven fie	0
28	A depth of knowledge in one or more of Neural Networks, NLP, Computer Vision, Voice Analytics, ASR	0
29	My client is a rapidly growing start up that is looking for an **energetic Data Scientist** that will **fit in seamlessly into their company culture.**	1
30	This **company has a gym on site and provides flexible work ours** to deal with the Atlanta traffic.	1
31	The ideal candidate for this position is the kind of person who is willing to go the extra mile to make sure the job gets done and can meet strict project de	1
32	Are you **ready to be part of something new and unique? Resideo is building a brand-new ultra-modern software development team** in Austin to	1
33	Come and join the Resideo software engineering team in Austin. This is an **exciting environment** where you will have the opportunity to develop your sl	1
34	We're **well funded, fun and focused** on building the very best in supply chain and we want YOU to help us build complex models that allow global comp	1
35	Smart, energetic and ready to make your mark in a best in class enterprise software company? Join our herd!	1
36	As a Data Scientist at IBM, you will help transform our clients' data into tangible business value by analyzing information, communicating outcomes and	0
37	IBM CIO (Chief Information Office) Organization is looking for **innovative talent who bring fresh new ideas t**o create a productive environment for all IB	1
38	Proficient with Access, Excel, Power BI, Tableau, SharePoint, SQL. Operating knowledge of AWS, Hadoop, Datameer or similarWorking knowledge of JA	0
39	They will design innovative analytic models and will be responsible for pulling and cleaning datasets, evaluating models, and presenting results to interna	1
40	Extract data from a variety of sources (databases, web, text files) and in a variety of formats (structured, unstructured, JSON, Parquet, etc.) and prepare	0
41	My client is a leader in the Manufacturing vertical and has operations in multiple states across the US.	0
42	They are seeking to hire a full-time Senior Data Scientist to collaborate and work with their R&D, IT, Product Support, and Sales teams	0
43	Ideal Candidates Will Have: -Multiple years of Business Intelligence with the ability to work with structured and unstructured data -Multiple years of progr	0

Figure 16: Annotation in implementation

Evaluation

Dataset

As part of our model training process, we reserve 20% of the dataset for evaluation purposes in order to fine-tune the model. Additionally, another 20% of the dataset is kept aside for testing the model's performance. Given that our data has highly imbalanced labels. We chose the precision, recall, and f1 score to evaluate the success of our models. Our baseline model was a Linear regression model that yielded a F1 of 0.47. Our best model yielded a F1 score of 0.7. Moving forward, we aim to incorporate more manually annotated job description data from publicly available datasets to enhance the quality of our model's training and evaluation. Furthermore, we intend to gather job postings in public online job boards, combining with user feedback on the model's predictions of these job postings to evaluate its performance in the future. This feedback will be used to further train and evaluate the model.

Motivation

We decided to use job descriptions from publicly available datasets to develop our model, as we wanted to ensure the data used was free from privacy protection concerns from the very outset of the model's design. As we move towards the future development of the model, our goal is to apply it in real-life scenarios to mitigate age discrimination in job postings. Therefore, utilizing data from public online job boards is a crucial step that must be taken to achieve this objective.

Preprocessing

In order to evaluate the data, each job description undergoes the following preprocessing steps: Firstly, the job description is tokenized into sentences, and any blank spaces, punctuations at the beginning or end of the sentence, and HTML tags (if present) are removed. Next, each cleaned sentence is transformed into word embeddings using BERT embeddings, and fed into the classifier.

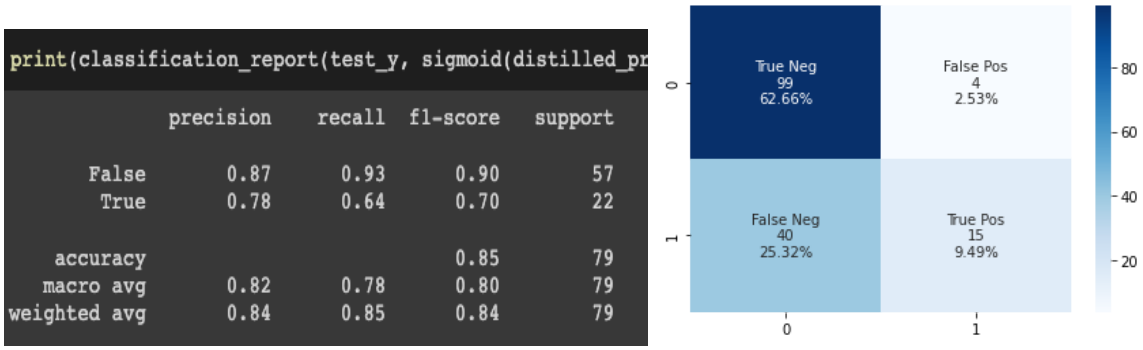


Figure 17: Model performance
Left: precision/recall on test data. Right: confusion matrix on test data

Future work

For a potential future iteration of the model, there are numerous improvements that can be made. We have enumerated some below:

Improving model performance

Currently, the model is trained on a static dataset annotated by students members of the project. Therefore, any possible difference in perception of discrimination by real-world users as compared to that learned by the model from the student researcher annotations would not be captured presently.

In order to capture any bias introduced by third party annotators and course correct the model to capture discrimination as perceived by the real world users, the next step would be to introduce a user feedback loop, so the training of the model can be dynamic based on real user provided feedback (such a like/dislike action) and less reliant on bulk annotation by external annotators thereby reducing the bias.

Improving user experience

A future iteration could provide users with a grammarly-style user experience where ageist language is highlighted directly on the job description instead of the plugin. This would mean that we can highlight even examples of age discrimination which is contextual instead of only based on words in our word corpus.

We could also collect more information about user behavior like website usage data, for example: hover data, click behavior, jobs applied information, so we can get a more nuanced understanding of the user's preferences of jobs.

Product iteration

Incorporating the above improvements, a next step in the process is - creating a stand alone experience that gives users a much easier way to apply to age-inclusive jobs. The plugin could have a resume feature that collects user's resumes and then matches them with age-inclusive job recommendations that are tailored to their job and industry preferences.

Model Card - Ageism in Job Ads

Model Details

- Person or organization developing model: Catherine Yu for Team Sage
- Model date: April 28, 2023
- Model version: 1.0
- We used the PyTorch-Pretrained-BERT library for the model building. We fine-tuned the pretrained ‘bert-base-uncased’ model on the labeled set, along with a linear layer and a Sigmoid activation to perform binary classification.
- BERT feature extraction: use pre-trained BERT model to extract features from the sentences of each job description

Factors

Relevant factors:

- Primary factors: personality based characteristics, physical health metrics, and skills, job sector type
- Secondary factors: Gender, age,

Evaluation factors:

- The current version has been trained on a dataset across all primary factors mentioned above.
- Although currently not implemented in the current prototype UI, we will measure real world performance against dynamic user feedback provided by users on individual job postings for the next iteration.

Intended Use

- Primary intended users: Job seekers (typically older) looking to avoid ageist jobs
- Out-of-scope use cases: Recruiters, laid-off job seekers

Training Data

- Currently, we allocate 60% of the dataset for training the model, while 20% is reserved for evaluation during the model training process. As with the evaluation data, we plan to introduce more manually annotated job description data from publicly available datasets in future iterations of the model's training. Additionally, we aim to collect job postings from public online job boards, and use user feedback on the model's predictions as the true labels to improve the model's training over time.

Evaluation Data

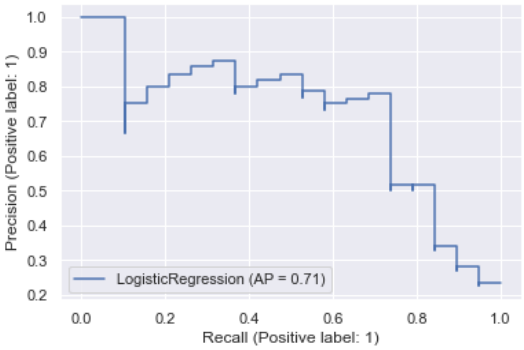
- Datasets: As part of our model training process, we reserve 20% of the dataset for evaluation purposes in order to fine-tune the model. Additionally, another 20% of the dataset is kept aside for testing the model's performance. Given that our data has highly imbalanced labels. We chose the precision, recall, and f1 score to evaluate the success of our models. Our baseline model was a Linear regression model that yielded a F1 of 0.47. Our

Metrics

- 158 sentences of job description are used as the testing data to measure the model performance.
- The performance of the model is evaluated using precision, recall, and F1-score.
- A decision threshold is used to convert the logistics output by the model to binary classifications. The threshold is chosen based on the precision recall curve (fig.1). We want to ensure a high precision rate while not sacrificing the recall rate too much, and thus a decision threshold of 0.5 is ended up being chosen. The 0.5 decision threshold gives us a precision rate of 0.73 and a recall rate of 0.42 on the true label of the testing data.

Quantitative Analysis

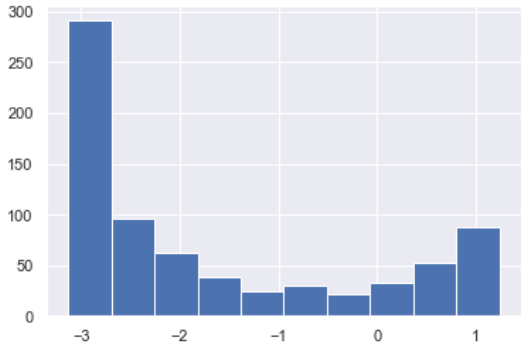
Precision-Recall Curve



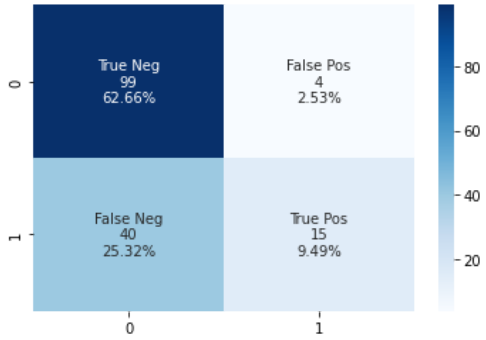
best model yielded a F1 score of 0.7.

- Motivation: We currently train the model using job descriptions from publicly available datasets to develop our model. As we move towards the future development of the model, our goal is to apply it in real-life scenarios to mitigate age discrimination in job postings. Therefore, we will consider utilizing data from public online job boards.
- Preprocessing: each job description undergoes the following preprocessing steps: Firstly, the job description is tokenized into sentences, and any blank spaces and punctuations are removed. Next, each cleaned sentence is transformed into word embeddings using BERT embeddings.

Distribution of training logits



Confusion Matrix on test data



Security & Privacy Needs

Our NLP model does collect partial user data such as their age, gender and feedback in order to determine user understanding of age discrimination over time. This allows us to fine-tune the model for greater accuracy in correctly detecting age discrimination in a site. As the model becomes more powerful, this creates valuable insights such as how perception of age discrimination varies across age groups or how women and men may perceive this differently. Our NLP model does not collect information on the job board itself or on the website that it is hosted on. Users may request copies of their data collected to ensure they are comfortable with this process.

Since our plug-in is not gathering any user data, the following personally identifiable information is considered out of scope for our system: Government issued identification numbers: Social Security Number (SSN), Driver's license number, California ID card numbers, Tax ID number, passport number, etc, Financial account numbers, credit or debit card numbers and financial account security codes, access codes, or passwords, Personal medical information, including protected health information (PHI) covered under HIPAA, Personal health insurance information, Biometric data used for authentication purposes, including facial recognition, A username or email address, in combination with a password or security question and answer that would permit access to an online account.

Users do not need to create an account when downloading and installing the plug-in. Therefore, we do not manage user credentials and we do not create a repository of identities. Users will also not be required to authenticate to use the plug-in, as all authentication requirements must be owned by the host site, for example, existing LinkedIn login credentials must be used in order to leverage the corresponding plug-in features. It is up to LinkedIn's current authorization policy to determine if the user is allowed to access the site after needed authentication controls have been met. Our team follows NIST 800-53 security controls for our plug-in and will also follow guidelines in GDPR and CCPA accordingly.

Acknowledgements

A special thanks to our advisor Annalee Saxenian, Professor at UC Berkeley's Information School who provided us with excellent feedback on a weekly basis and fostered our learning and growth during this project. Thank you to David Bamman, Professor at UC Berkeley's Information School who guided us through the development of the NLP model. Thank you to Filipe Santos for collaboration with Eileen Cahill and his work on ageism-related information visualizations as part of a separate project for a class INFO 247, Information Visualization and Presentation at UC Berkeley on age-related bias in the workforce.

Appendix

Section A: Usability Testing Script

April 2023

Interviewers
Eileen Cahill
Kedari Narra

Primary Objectives

- Uncover usability problems with existing interface when performing different tasks
- Compare different interface designs and discover intuitive interaction patterns
- Seek feedback on the current UI and prompt for feature requests or improvements

Script

Introduction
Hello, my name is [insert name here] and I'll be leading our interview today. I'd like to start by thanking you for making the time to speak with me. I'll mostly be talking about our plugin, Sage, which is designed to help aging workers navigate the modern job market.

Let me give you an outline of what's going to happen. I am going to ask you to perform a series of tasks on our plugin. It would be very helpful if you could talk through the tasks. We want to understand things from your perspective. It's important to highlight that this isn't a test.

This interview will last approximately 30 minutes, if at any point you want to take a break or stop the interview, please just let me know. With your permission, I'd like to record this interview. The recording will only be used to help us in our research, and it won't be shared with anyone except those with a need-to-know. Recording this interview also helps me, because I don't have to take as many notes! Is that okay? Great. Do you have any questions for me at this time?

Warm up Questions

- Do you use any job sites like LinkedIn, Indeed etc?
- Were you in the market looking for a new job in the last 2 years?

Tasks

- The font size of the page is too small. How would you change this?
- Can you fill out an application form using your voice instead of typing?

- The text on the page is too bright. How would you rectify the situation?
- You would like the plugin to read out the contents of the webpage. How would you do that?
- You want to know how the plugin is detecting age-bias in job postings, where do you find this information?
- You want the second paragraph to be read aloud again. How would you do that with the help of the plugin?
- You want to provide feedback to the team. How would you do it?
- You want to fast forward the screen reader. How would you make it happen?
    - Follow up: What do you expect will happen when you do this?
- You want to control the speed of the screen reader. Is it possible to do that on the plugin?
    - Follow up: What do you expect will happen when you do this?
- You are not comfortable with the layout of the page and want to simplify it. Which tool on the plugin would you use to do that?
    - Follow up: What do you expect will happen when you do this?
- You want to know if a job listing has age-bias. Where would you find this information?
    - Can you talk through what you see here?
    - What do you think the horizontal bar represents?
    - What do you infer from the words?
    - What do you think the different colors of words mean?
    - How would you compare different jobs based on this information?

Follow-up questions:

- What did you think of this prototype overall?
- Were there any confusion points?
- Would you use a tool like this?

<u>Wrap up</u>

Thank you so much for your valuable time, this conversation has helped greatly. Before we finish, is there anything that I am forgetting or would you like to add anything.

Section B: Heuristic Evaluation

Mon. Apr. 10

Participants

Eileen Cahill
Kedari Lahari Narra

Nielsen’s Heuristics

Neilson's heuristics:

#1: Visibility of system status

#2: Match between system and the real world

#3: User control and freedom

#4: Consistency and standards

#5: Error prevention

#6: Recognition rather than recall

#7: Flexibility and efficiency of use

#8: Aesthetic and minimalist design

#9: Help users recognize, diagnose, and recover from errors

#10: Help and documentation

Severity of violations:

0 - don’t agree that this is a usability problem

1 - cosmetic problem

2 - minor usability problem

3 - major usability problem; important to fix

4 - usability catastrophe; imperative to fix

Eileen notes:

Page	Heuristic	Severity	Fixes/comments
Homepage/landing	aesthetic/minimalist design	1	Perhaps not minimalist
	Recognition rather than recall	2	Some recall required - familiarity with what color contrast means, what speech to text means
Help page	aesthetic/minimalist design	1	could use more aesthetic design, but is minimalist which is great
	Help and documentation	N/A	need to make sure we add answers to each of these questions and link to contact page
Read page aloud	Consistency and standards	2	volume control is potentially redundant

	Recognition rather than recall	2	would older people recognize the 10 second markers?
	User control and freedom	3	how does the user know where to start and stop what they want read? What if they only want a paragraph read? Would they need instructions on how to get there?
	User control and freedom	2	what if the user wants to be able to change the type of voice?
	Visibility of system status	3	user doesn't know where the reading is happening on the page
	Aesthetic and minimalist design	2	the design is cluttered, not especially logical and it's unclear if the bottom blue

			rectangle is a button or not.
	Consistency and standards	2	button looks different here than in other screens
Minimizing plugin	Recognition rather than recall	2	going back to the plugin may be confusing with just symbols present
	Aesthetic and minimalist design	1	color may blend into background, might want a standout color
	Help users recognize, diagnose, and recover from errors	2	what if user exits out when they want to go back to plugin - do they know how to reopen the plugin?
Speech to text	Help and documentation	4	User could be very confused if never used speech to text before
	Visibility of system status	4	Unclear where on the page the speech to text is

			filling out
	Error prevention	3	No method for user to correct text that might be recognized incorrectly
	Consistency and standards	3	It doesn't look like the job is imported from the page - it looks like the job title just pops up out of nowhere
		3	How does user even get to this page - do they click a button on the page to fill out using speech to text? Does the plugin popup with the suggestion to use speech to text?
		4	Where would continue lead?
Font adjust	Aesthetic and minimalist design	1	Sample text sits awkwardly in the middle
	Matching the real world	2	Is the typeface selection necessary? It might give the user too many unnecessary choices
			On medium

			page, link to small page doesn't work, help button doesn't work, back button and X doesn't work on any of these pages
Color contrast	Aesthetic and minimalist design	2	A lot of white space - could be used for something
	Visibility of system status	2	Need to reflect on page the light mode/dark mode - showing an example of what this looks like
	Recognition rather than recall	2	Would user know what light mode and dark mode refer to?
Overall	Help and documentation	3	What if user wants help SPECIFIC to a page? Not just the general help page? They should be able to receive guidance for each particular page as well as general help.

Total score: 58

Kedari notes:

Page	Heuristic	Severity	Fixes/Comments
Homepage	Visibility of system status	1	‘On’ button text can be prominent
Speech to text	Match between system and the real world	2	Should this page open only when there is a form on the screen
	Aesthetic and minimalist design	0	Form fields can be taller
Font adjust	Consistency and standards	1	Help button at the button is not present in other screens
	Consistency and standards	0	Padding can be consistent
	Consistency and standards	1	“Apply” button at the bottom can make the experience consistent
Color Contrast	Visibility of system status	2	Sample text is always present and not indicating the mode selected
	Consistency and standards	1 1	“Apply” button at the bottom can make the experience consistent Toggle button instead of a normal button unlike the other pages
	Aesthetic and Minimalist design	1	Sample text seems like a button

	Match between system and the real world	1	Toggle button might not seem intuitive
Help center	Recognition rather than recall	1	The text doesn’t seem clickable
	Consistency and standards	1	Contact us button left justified, unlike other pages

Total score: 13