THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ELECTRICAL ENGINNERING AND COMPUTER SCIENCE

Implementing a Web-Based Application to Support Researchers in Improving the Diversity of their Citations

VENKATA SAI RENUSREE BANDARU
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Computer Science
with honors in Computer Science

Reviewed and approved* by the following:

Dr. Christopher McComb
Adjunct Professor of Engineering Design
Thesis Supervisor

Dr. Jesse Louis Barlow
Professor of Computer Science and Engineering
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Academic research papers often reference various publications to support the contents of the document. Since the number of citations is often used as a metric of academic success for researchers, it is important to avoid bias in the citations used. However, studies have identified that there tends to be a bias towards the choice of references used by individuals, owing to the history of academia that has been white male dominated. This thesis details the implementation of a web-based application that allows individuals a way to assess the diversity in the references they cite in their papers. More specifically, this thesis discusses the use of various packages and their algorithms to analyze citations and provide feedback to the user on the distribution of authors that they cite in terms of gender, race, and ethnicity. Based on the titles of the citations provided in the input, the web-application uses an API to retrieve papers of similar context and provide recommendations to the user.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to everyone who has supported me in any capacity throughout my academic journey. I extend my utmost appreciation to my thesis supervisor, Dr. Christopher McComb, for his patience, moral support, kindness, motivation, and guidance through each stage of my thesis research and writing process as well as my college career at Penn State University Park. I would also like to thank my honors advisor, Dr. Jesse Barlow for guiding me and constantly supporting my time at Penn State. I really appreciate my friends for their constant support and engagement throughout my whole journey as without them, I would have been completely lost.

Last but not the least, I want to acknowledge and express my gratitude to my parents and sister for always being there for me, through my thick and thins, for all the untold sacrifices to help me have a better life, and for inspiring me to constantly strive for excellence.

Thank you for everything!

Chapter 1

Introduction

Diversity and Inclusion (D&I) is an area that is being brought to light more often within businesses, schools, multinational corporations and other industries (Porterfield, 2021). It is important to ensure that, as individuals in a society, we recognize and acknowledge everyone thoughts and values. However, some fields have been long dominated by certain genders or ethnicities which self-perpetuates the bias existing within their domain. Academic publishing is one of the sectors that doesn't give high importance to the range of diversity in terms of citations.

Academic research papers reference various publications to support the contents of the document. Since the number of citations is often used as a metric of academic success for researchers, it is important to avoid bias in the citations used (Carpenter et al., 2014). In this context, diversity in citations refers to the incorporation of citations published by authors belonging to diverse backgrounds in terms of ethnicity and inclusiveness in gender identity. Considering how the choice of references chosen could imply multiple types of citation bias, increasing the diverseness of publications used is a method to minimize the partiality when selecting a research publication that supports one's research.

However, studies have identified that there still tends to be a bias existing towards the choice of references used by individuals. This means that individuals tend to cite people that identify similar to themselves and they identify with (King et al., 2017). Owing to the history of academia that has been white male dominated (Homaeipour, S., 2018), this implies that more

people will often cite white men because they are historically the most cited publications belonging to a specific ethnic group. As a result, this citation bias can potentially mean highly impactful work by minority authors is being under-cited. Additionally, citations are considered a metric of success. This means that other individuals who do not belong to this white dominated group are put at a major disadvantage for not being able to recognize their academic promotions and other awards.

The purpose of this thesis is to design and prototype a web-based application that would allow one to verify and evaluate the diversity in the references that they use in their academic research papers. The application feeds on bibliographic references that the user inputs. The ethnicity and gender of authors are inferred with the help of using APIs and various pre-existing packages. These distributions of results are displayed in an infographic format on the web-based application to give a visual representation of the diversity that the user has with their citations. In addition, the application recommends the user with various other academic papers with other references which are relevant to their research area. As a result users have the ability to include a potentially diverse range of citations within their papers and reduce the citation bias that previously existed.

Chapter 2

Background

In this modern era, there is a shift in understanding and recognizing the importance of having diversity and inclusion within any domain, be it industry workspace, classroom settings or multi-national organizations. Though they are generally interchanged, diversity and inclusion represent two different concepts. While diversity talks about the representation in terms of various backgrounds and cultures within the entity, inclusion refers to the significance given to the presence of various perspectives of individuals from different backgrounds within the environment (Nair & Vohra, 2015). The combination of these concepts, leads to a balance of thoughts and perspectives within a community allowing every person to feel valued, supported and involved within a society.

This complements the fact that involvement of individuals from various backgrounds makes it possible to reap better outcomes, be it productivity in a workforce or involvement within a community (Leone, 2020). Social interactions impact the way individuals perform their duties. Having the ability to be exposed to multiple perspectives from humans with different experiences allows one to expand their outlook and creativity (Reiter-Palmon & Illies, 2004). As a result, one gains many skills such as becoming more open-minded, more aware of bias, better able to collaborate and work well within a team, more committed to diversity and inclusion, and more culturally intelligent(Bourke & Titus, 2019). These characteristics are represented of an inclusive leader that knows how to get the best out of a diverse workforce by improving performance, collaboration and decision making of the group of individuals.

Many other sectors have recognized the importance of having a diverse and inclusive environment within their industry (Nair & Vohra, 2015). Academia is one of the domains that is actively engaged in bringing an inclusive and diverse experience for scholars of all ages – from kindergarten to university graduates. One study discusses how two different approaches of cultural diversity within schools impact the academic and socio-emotional development of an adolescent's moral reasoning as well as social and ethnic identity (Schachner, 2019). It highlights how they observed the possibility of having positive impacts on diverse students promoted due to both the approaches of equality and inclusion as well as cultural pluralism within multi-ethnic schools. Educating students on the benefits of diversity and inclusion prepares the future generation to collaborate well and be better leaders for their own success and the advancement of the society.

However, it is still important to understand that academic publishing is also a realm of area that is focused on educating the society by making known of new discoveries and/or inventions that technological advancements, historical events, or new creations within every domain of knowledge. As a result, a diverse perspective that is inclusive of various viewpoints within academic publishing also is significant to understand which can be better understood in the following section.

Importance of Diversity and Inclusion in Academic Publishing

Academic publishing is one of the factors affecting the measure of success for researchers and academic professionals. Publication metrics have been a part of evaluation metrics in many contexts to assess academic productivity and impact of research of an individual in his/her field of study. These metrics have been a basis for academic professional's success such as for evaluating

tenure and promotion, recruiting opportunities, grant applications and renewals, and administrative purposes for departmental or university performance reports (Carpenter et al., 2014).

Factors like the number of publications a researcher publishes and the number of citations that those publications receive can be a signal of how visible and credible the research and the researchers are (*What Is the Significance of Academic Journals?*, 2019). According to Lawani, there is more significance to the higher citations a paper obtains as it is directly related to higher quality of work and thus the number of citations increase with the quality of the paper (Lawani, 1986). The number of publications not only demonstrates the researcher's efficiency but also could be an indicator on the researcher's career growth. For example, having more publications in the early stages of an academic professional's career denotes the sign of having more publications in the future allowing them to have a higher recognition within grant funder applications as well as have better potential within their field. A study in radiology residency candidates shows how residents having a greater number of publications in their early career growth was a strong pointer to higher potential within the area of radiology (Rezek et al., 2012).

Another metric of academic career success includes author status: whether one is sole author, first author, or last author. Having multiple authors within a single publication is a sign of collaborative activity which demonstrates cooperation and productivity, while being first or last author shows that the individual contributed to majority of work (Nichani, 2013).

Peer-review is the process of "subjecting an author's scholarly work, research or ideas to the scrutiny of others who are experts in the same field" (Kelly et al., 2014).Whether the paper is published in a peer-reviewed journal or not, demonstrates the quality of the research as peer reviewers provide feedback on how the authors can improve their contents and identify errors that need to be modified. Additionally, peer-reviewing ensures that only high-quality research is

published based on analyzing factors such as validity of the content, originality as well as the significance of the research (Kelly et al., 2014). As a result, the greater number of peer-reviewed journals in which an academic professional published his/her work also impact the way they obtain funding in their future projects. Journals within varied areas of specialties highlight the diversity and depth of publication and the research itself (Nichani, 2013).

A crucial step in any research process is finding authentic resources and literature reviews. This stage is essential in gaining deeper understanding of the subject being explored. A good place to start would be by organizing information into defined categories. This involves being aware of known facts, current and existing studies and analysis followed by investigating further developments happening during that moment in time.

In academic settings, professionals are inclined towards looking at published articles and published research papers for references as they indicate recent or current trends. Additionally, published documentation is acknowledged to be more credible as publications are peer-reviewed. This leads to the question of how these resources are chosen and whether there are any biases involved when selecting these sources and citations. By nature, humans subconsciously possess bias that influences the way they think or act. This characteristic is inherent and cannot be detached. However, it is important to understand how this bias could affect the way an individual chooses his/her sources for academic references.

Bias in Citations

Bias can have both minor and/or major effects on how an individual interprets a situation and this is very subjective (Cain & Detsky, 2008). Therefore, it is crucial to acknowledge that a

person's observations and judgments might have unintentional influence of bias which it should be protected from by being objective when executing a task.

Undeniably, this bias is evident in academic referencing and publishing as well. Bias has played a role in many aspects of an academic research and is still actively present in the domain. Peer reviewing is a scrutinous process that could also be subject to bias which has impacted various the outcomes of numerous research studies (Lee et al., 2013). This is because, in the end, peer reviewers are also humans who are susceptible to bias. This is also known as reviewer bias which is influenced by many factors (Tvina et al., 2019). This prejudice from reviewers could lead to proposed qualitative and/or quantitative research to be rejected eventually leading to rejections from research funding bodies. Aspects like gender bias, ethnicity bias, regional bias, institutional affiliations also known as prestige bias are some of the biases that exert influence on reviewers when they are examining a scholarly study (Drieschová, 2020).

Recognizing and understanding that biases exist in research and citations is prominent as it refers to the credibility of the study itself. As mentioned, every induvial has a subconscious personal bias with them when taking any decision and that is part of every person. It is important to ensure that this bias doesn't affect the choices one makes in terms of the academic references they use for their study. This owes to the fact that when a scholar's manuscripts and publications don't employ any satisfactory mechanisms to minimize bias on their work, it is indicative of supporting bias. As a result, that authors publications will lose credibility and will likely be considered as an unreliable source of information (Galdas, 2017). As mentioned earlier, this impacts the number of citations that an author receives decreasing their validity in the scholastic domain they are pursuing.

Similar to the reviewer bias, scholars conceivably have identical biases. Personal bias associated with an individual could amplify with gender, ethnicity, regional, and institutional bias. In situation where these biases surface when researchers are accumulating their resources for their study, it could lead to unfavorable consequences such as having partiality with their findings/results. Since this is not credible, eventually their entire research can be disregarded. This then leads to evident implications on the professional's integrity in their respective career field. The impacts of these biases within the academic realm are discussed in further detail in the following sub-sections.

Prestige Bias in Citations

Prestige bias refers to the partiality of an individual when they predominantly prefer to choose to acquire knowledge from a scholar who is known to be prestigious within their domain because of their success which lead the scholar to gain attention respect and admiration from their peers (Brand et al., 2021). In the domain of academic publications, this is relevant to when authors who are inclined to subconsciously treat famous publishers or their affiliations work within their educational domain preferentially because of their prestigious status. Authors who tend to constantly cite references from prestigious people from their area would often refuse to acknowledge the work of other intellectuals from their domain that could have been momentous to their own study of research. This type of prejudice is also evident in peer reviewing where reviewers unconsciously tend to treat submissions of recognized individuals and their affiliations with an unfair preference (Frachtenberg & McConville, 2022). Studies have shown that this type of bias acts as a barrier to providing and supporting fair and impartial fields of academics (De

Cruz, 2018). As a result, it can lead to an exacerbation of an unjust structural representation of individuals leaving out the work of many well-educated scholars.

Regional and Ethnic Bias in Citations

Regional bias, more popular as Nationality bias denotes the prejudice of a person who tends to favor authors or publishers who are situated in the same country as the journal or paper (Lee et al., 2013). Studies have shown that this is evident in publication reviewers, some highlighting the inclination for US based authors and publications more favorably (Link, 1998). However, it has also been studied that this bias could be because the writing style and language of non-native speakers and not the nationality itself (Herrera, 1999). Another study that analyzed the comparison of citation practices by health professionals found that U.K authors publishing on the peer reviewed journal Lancet and U.S authors publishing in New England Journal of Medicine are more inclined to cite material composed in their own countries. Material from other countries tend to be published in a reduced amount due to the perception of the foreign information being inferior (Campbell, 1990). Generally speaking, these studies implicitly hint, those academic publications from ethnicities of various non-US and non-UK countries, are being disregarded by many publishers because of the patriotism towards their own country.

Gender Bias in Citations

Gender Bias is the act of treating an individual differently based on the person's real or perceived gender identity (Rothchild, 2007). Often these are stereotypical beliefs are related to the difference in behaviors towards females and males. Gender gap is observable in many domains

and the field of academics also experiences the same, mainly highlighted within the STEM fields. Studies show that gender bias in academic publications refer to the discrimination in recognizing and providing credit by citing female publishers and often tend to cite male authors for their research (Homaeipour, 2018) . According to the analysis done by Homaeipour, findings show that manuscripts written by female publishers, either first author, last author, or dominant gender of the authors, have less citations comparatively to male authors for similar topics while controlling factors year of publication and journal where it is published. Another study examined how this bias has implications on the academic career of individuals and deduced that because of the significant difference in peer-reviewed publication rates between males and females, males had a higher probability of benefits such as holding a PhD degree and being tenured or on track to get tenured, and held a higher position in their career compared to females (Kaufman & Chevan, 2011).

Given that the STEM field is still predominantly male populated, gender bias is highlighted in this area many academic publications. Another study revealed that women engineers work receives lower recognition in terms of number of citations compared to the male colleagues within the scientific engineering community even when they publish their papers with higher impact factors (Ghiasi et al., 2015). According to the same investigation, it was concluded that, in the domain of engineering, scholars, regardless of their gender, incline towards collaborating predominantly with males. Consequently, this contributes to the increase in male-dominated structures because of the continual forming and repetition of these collaborations. This demonstrates how, regardless of the gender, the is still prejudice towards the publications of female authors in this academic field.

Nonetheless, there are methods being implemented to reduce this type of bias which has shown positive results. Multiple studies have indicated that both male and female reviewers tend to give higher scores to identical work done by males than females (Budden et al., 2008) .Double-blind peer review is one of these methods being incorporated where the identity of author and reviewer is not revealed to each other throughout the review process (Tomkins et al., 2017) . Introduced first in the journal Behavioral Ecology, this technique has seen significant improvement in the number of female first-author publications when compared to single-blind review process for similar publications (Budden et al., 2008).

Another study examined what role the gender of first author has on the publishing frequency, self-citing tendency, and the number of citations that a paper receives. Some statistics disclosed this publication named "Quantitative Evaluation of Gender Bias in Astronomical Publications from Citation Counts" in 2016 show an increase in the number of papers which have female first authors by about 20%. However, it was still evident that male first authors still receive more citations per papers compared to female first authors (Caplar et al., 2017) .

As shown above, citation bias is noticeable in the academic realm, and it is important to minimize the effect of these biases for providing an impartial publication. In order to provide a web application to decrease the effects of bias in references cited, it is necessary to understand previous research methods that have been investigated.

Previous Research/Related Work

The approach for this project is to analyze the first name of the authors in a publication to infer gender and the last name of authors to infer the ethnicity. Previous research that explored

identifying gender and ethnicity based on the first and last name of individuals are described in the following subsections.

Gender Classification of Names with Character Based Machine Learning Models

This publication analyzes the first names provided by registered users for internet accounts to predict the gender of the individual. For this, the study tries to identify the gender of the person based on their first name and improves the prediction with the help of the individual's last name. It compares different algorithms, such as Long Short-Term Memory (LSTM), Character Convolutional Neural Network (Char-CNN) and Character-BERT (Char-BERT) to execute this investigation. Results of the study show that using dual LSTMS are effective in addressing the issue of having different gender connotations within different cultures and give accurate predictions. In the case of when the names are unisex, utilizing a content-based model was more efficient to obtain good results (Hu et al., 2021).

Predicting the Gender of Indonesia

In this publication, the authors investigate a method to predict the gender on Indonesian names with the help of using Character-Level Long Short-Term Memory (char-LSTM). Given that Indonesian names traditionally do not include surnames like known English names, the analysis focused on examining both first names and full names. To further understand which model is better for the classification, they compare their char-LSTM model with other conventional algorithms, namely, Naïve Bayes, Logistic regression, and XGBoost with n-grams as features. From their investigation, they deduced that using char-LSTMs improves the accuracy in predictions of

genders compared to the conventional algorithms tested. When tested on full names, the accuracy improved to about 92.25 % using char-LSTM while testing the model only on first names resulted in about 90.65% accuracy (Septiandri, 2017).

Name-Ethnicity Classification from Open Source

In this manuscript, the authors develop an ethnicity classifier that learns from an open-sourced dataset, i.e., public and non-confidential sources. Given that, their classifier implements a combination of Hidden Markov Models and Decision Trees to classify the data into 13 different ethnic groups. It has been identified that this type of research investigation is more evident within the realm of biomedical research yet encounters difficulty with regards to distinguishing linguistically distinct cultural/ethnic groups. According to the article, the classifier resulted in reasonably well predictions for the ethnicity prediction, except for some ambiguity when the training data was limited (Ambekar et al., 2009).

Development and Validation of a Computerized South Asian Names and Group Recognition Algorithm (SANGRA) for use in British Health-Related Studies

This paper discusses the establishment of an algorithm, known as SANGRA, short for South Asian Names and Group Recognition Algorithm, that classifies names of South Asian ethnicities. It highlights how it is important to classify the ethnicity of people from South Asian countries to identify differences in health-related exposures and disease risks in Britain. To distinguish South Asian names, they created a dataset by accumulating first names and surnames from countries like India, Pakistan, Bangladesh, and/or Sri Lanka together with the religious and

linguistic origin on which the algorithm was trained on. The algorithm was able to predict the ethnicity of individuals from the South Asian origin with an accuracy ranging from 89 – 98% (Nanchahal et al., 2001).

Chapter 3

Methodology

Overview and Approach

Cite Diversely Application

The objective of this thesis is to provide an approach that can be utilized to minimize implicitly bias an author might be incorporating in their publications. Aiming to reach as many users as possible and providing ease of use, the web application was proposed. This application intends to move a step closer to helping academics understand their own implicit bias in academic publications by evaluating the diverseness of the authors in the publications by feeding on citations as input. In order to be effective, the web platform would need to be able to take in numerous amounts of data in the form of a file or in the text area and provide a result that can be comprehended by the user. Using various open-source packages and APIs, the bibliographical data are tokenized and parsed to be passed into the model that infers the most likely gender based on their first name and most likely race and ethnicity based on the author's last name based on which the user can gain a perspective of their choices in references.

The application gives returns information on the authors in all the citations and the distributions of gender and ethnicity predict results are displayed in an infographic format on the

web-based application to give a visual representation of the diversity that the individual has with their citations. In addition, the application recommends various other academic papers with more diverseness in terms of gender and races/ethnicity similar to the topics that have been inputted. This way users can include a diverse range of citations within their papers, attempting to reduce any citation bias that previously existed.

To develop a web application, there are multiple steps are required to be taken. Below the general approach is first described followed by the packages and libraries used for constructing the interface are discussed to gain better understanding of the implementation.

Approach

The flowchart in *Figure 1* provides an overview of how the system works. The first step of the process is to take in the input of bibliographic citations in the form of APA, BibTeX, MLA or Chicago citation style into the web application. Based on the citation option chosen, this input is either passed directly into the BibTeX parser or written to a file to execute a function that preprocesses the text, discussed in detail in the following sections. Then, the processed text is passed into the gender inference and ethnicity inference functions depending on the citation option to get the predictions based on the names in citations. Using these predictions, the information is formatted and loaded into the donut plot to display the visual infographic. Simultaneously, the processed text is duplicated and refined to extract the titles of the publications from the citations passed. Once the clean titles are gathered from preprocessing, they are passed title by title into the recommendations function that retrieves related academic publications. With both pieces of information, the web application displays the infographics showing the probabilities of genders

and ethnicities in citations and suggests related academic publications for increasing the variety of the genders/ethnicities of the authors for referencing.

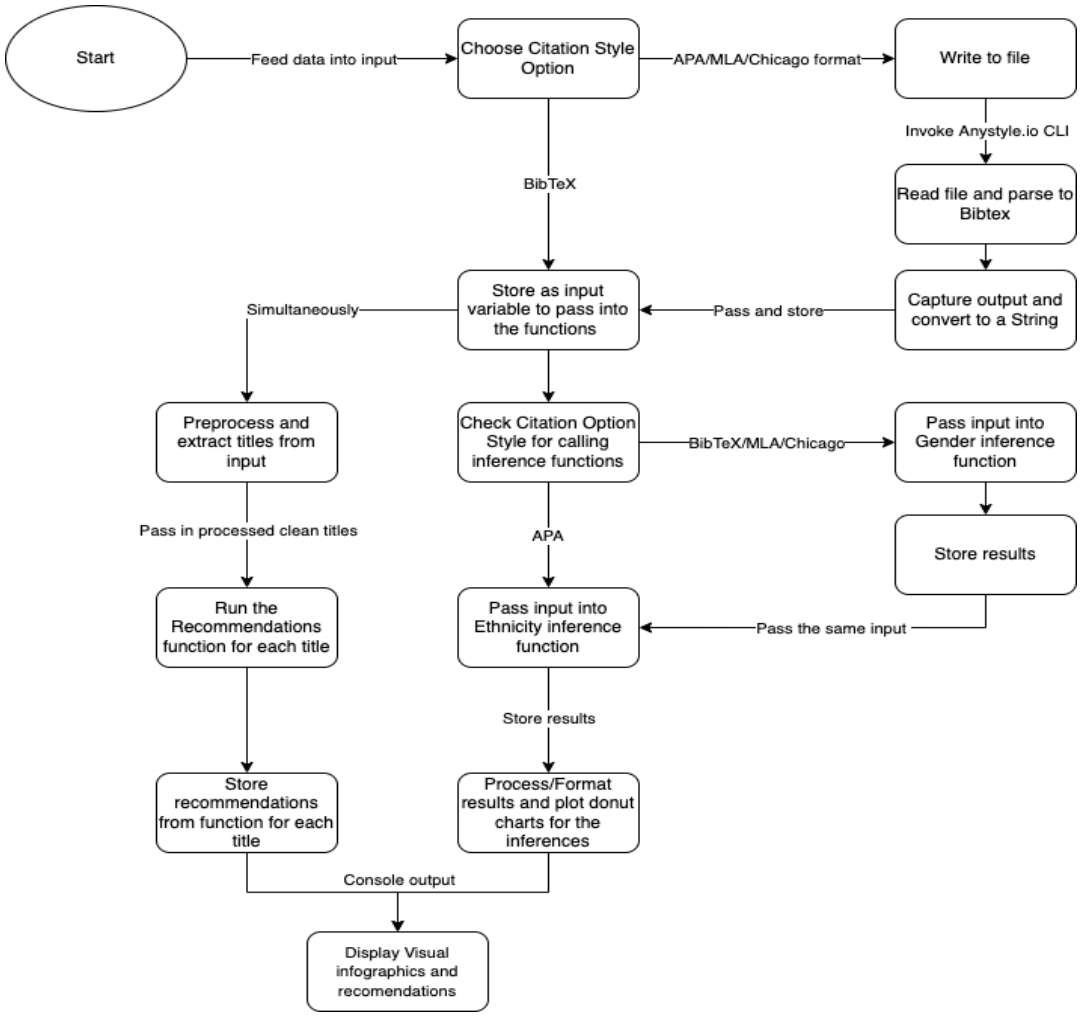

Figure 1: Application Process Flowchart

For the design and implementation of the web application, various existing libraries and packages have been utilized. The majority of these packages were in the Python language while one belongs to Ruby and the other was a framework for Python which is discussed in detail in the following subsections.

Preprocessing

Anystyle.io – Ruby Command Line Interface

In order to parse different citation styles inputted into the BibTeX format, the AnyStyle.io Command-Line Interface (CLI) has been implemented in this project. AnyStyle.io is an open-source software that parses academic references into various structured, bibliographic data. Regardless of the citation style, the web application takes in a single reference or a list of references and translates them into either a BibTeX format, CSL-JSON format, or XML format as required by the user. This software uses machine learning heuristics to segment data into specific attributes swiftly and includes the ability train another model with data relevant to one's parsing needs (Keil, 2011/2022). It is mainly available as a web interface but is always available as a ruby command-line interface that can be applied in other projects based on the users. To utilize it in this application for getting the right format for parsing, the command line interface has been incorporated. Considering how the software is only available in Ruby language, the python code executes it via a subprocess which is discussed in further detail in the following section.

Parsing the input

The web application can process a variety of different citation styles. The BibTeX style is chosen as the optimal citation type since it clearly segments various aspects of the references and denotes them a title/key which can be identified to be extracted. To infer the gender and/or race

and ethnicity from this input, it needs to be formatted into the BibTeX style for passing it into a package that extracts first names and last names.

Two approaches were considered for this– using regex or incorporating existing packages that parses the citations. Both techniques have their own positive and negative aspects. Regex was the more universal solution that would allow the ability to extract the contents of a citation. Considering how regex is known to be useful in the search and replace operations, implementing this in python would speed up the preprocessing and would reduce the dependencies of the code itself. On the other hand, python does not have any modules or packages that take in a citation and give out a BibTeX citation to be fed into the algorithms.

Due to this barrier, other languages were investigated for any such API that can be used to parse academic references which resulted in choosing AnyStyle.io. Despite finding a package that performs as needed, the issue with AnyStyle.io was that it was in Ruby which would need to be executed using an system call that invokes a process in the subshell, hence significantly slowing the process. In addition, the command line interface didn't have the ability to output to a variable as it was only called through the terminal and read from a file of references.

The approach to resolve this was to invoke a subprocess that called the AnyStyle.io command line interface from a file and parsed it into the BibTeX format. With the help of using the subprocess library embedded in python, the parsed information from the output was captured and converted to a string which was accessed in the python code to feed into reference inference code.

The reference inference code proceeds to take in the formatted data and picks out the details of the authors from the publications. Using another python package namely Nameparser, these references method segments human names into individual components of first and last names.

These names are accumulated and stored to be passed into the gender inference and ethnicity inference functions respectively.

Initially, all text from the input was passed directly into the Anystyle.io Command Line Interface (CLI) which converts the citation style to BibTeX and saves it. However, there was a difference in the gender and ethnicity results when BibTeX was passed into the Anystyle.io CLI, leading to discrepancies compared to when the BibTeX was directly passed in to the BibTeXparser. To solve this, radio button options were added, as shown in *Figure 2*, to differentiate each citation style for the data inserted. Selecting different options determined the next steps that would take place after preprocessing the data. The code was altered to invoke the Anystyle.io CLI subprocess only when the option of "BibTeX" was not selected, removing any further inconsistencies in the processing.

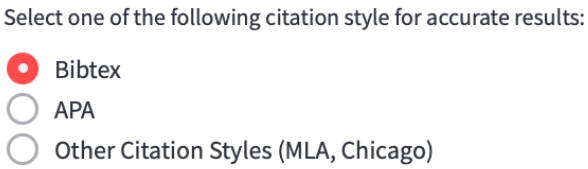

Figure 2: Radio options for Citation Styles

The gender prediction of the names faced some revision when incorporating the ability to recognize various types of citation formats. Initially, there was no separate choice between MLA, Chicago, and APA – only BibTeX and Other styles. The obstacle to this was that APA citation compresses the first name to an initial. Because the gender inference function bases the prediction on the first names of the individuals, this would result to nil results in the donut plot giving a blank graph for the gender distributions while the race and ethnicity distributions was displaying. This limits the use of the web application system exclusively to the ethnicity aspect without providing

its full potential. As a result, the option of the APA radio button was integrated allowing to differentiate the input. When this option is selected, the gender inference is disabled, and a warning prompt is shown to the user as displayed in *Figure 3*. When other options are selected, both gender and race/ethnicity are inferred providing both plots and references.

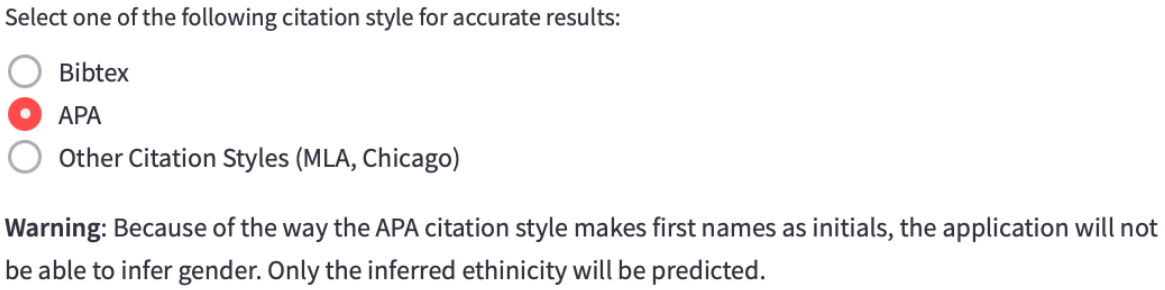

Figure 3: APA Option Warning Dialog Box

Table 3 below, the provides an overview of the various types of inputs that can be fed into the web application, examples of these citation styles and the types of inference that can deduced from the different kinds of styles fed into the system.

Table 1: Type of Inputs and Inferences expected

Type of Citation Style Input	***Example of Input***	***Gender Inference***	***Ethnicity Inference***
BibTeX	@article{caplar2017quantitative, title={Quantitative evaluation of gender bias in astronomical publications from citation counts}, author={Caplar, Neven and Tacchella, Sandro and Birrer, Simon}, journal={Nature Astronomy}, volume={1}, number={6}, pages={1--5}, year={2017}, publisher={Nature Publishing Group} }	Yes	Yes

APA	Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, *1*(6), 1-5.	No	Yes
MLA	Caplar, Neven, Sandro Tacchella, and Simon Birrer. "Quantitative evaluation of gender bias in astronomical publications from citation counts." *Nature Astronomy* 1.6 (2017): 1-5.	Yes	Yes
Chicago	Caplar, Neven, Sandro Tacchella, and Simon Birrer. "Quantitative evaluation of gender bias in astronomical publications from citation counts." *Nature Astronomy* 1, no. 6 (2017): 1-5.	Yes	Yes

Algorithm Implementations

Gender Inference

To infer the gender from the given input, the first names of the authors are extracted from the parsed BibTeX data and passed into a gender inference function. The algorithm that determines the prediction of the gender is based on a package called gender-guesser. Gender-guesser is a python package that predicts the gender given a first name. It utilizes the program code of "Gender-verification by forename" written by Jörg Michael (*Gender-Verification by Forename (Cmd-Line-Tool & Db) - Utilities*, n.d.). This is an algorithm that is trained on a manually made dataset that consists of more than 40,000 first names, genders associated with them, and frequency of each name. The dataset encompasses names from countries in Europe, and partly from China, India, Japan, and USA (*Gender-Verification by Forename (Cmd-Line-Tool & Db) - Utilities*, n.d.). In addition, the dataset had 600 pairs of equivalent names that were used to distinguish the ambiguity of names from different regions/countries. Based on the logarithmic frequency scale that ranges from 1 (rare) to 13 (extremely common), the different gender types of mostly male,

likely make, mostly female, likely female, androgynous, and unknown are inferred. Using this package, the genders of the publication authors are determined based on the following labels: Mostly Male, Likely Male, Mostly, Female, Likely Female, Androgynous, and Name not found. Consequently, the results are categorized and interpreted in the donut plots as shown in *Table 1*.

Table 2: Gender Types

Generalization in the visualization	Labels by Gender Guesser
Likely Male	Mostly Male, Likely Male
Likely Female	Mostly Female, Likely Female
Hard to Tell	Androgynous
Unable to Tell	(Name not found)

Ethnicity Inference

Similar to gender inference, the ethnicity inference algorithm feeds on the names of the authors extracted from the data inputted. However, it employs the last names of the individuals and deduces the race/ethnicity based on the highest probabilities. The algorithm performs a database lookup on the US Census data from 2010 for deducing the ethnicity. The dataset consists of last names and the likelihood of that last name belonging to the various ethnicity types. The ethnicity type labels are Likely White, Likely Black, Likely Asian or Pacific Islander, Likely American Indian or Alaska Native, Likely Two Races, Likely Hispanic, and Hard to Tell. Based on the last name, the algorithm looks up the dataset, identifies a corresponding last name and

obtains the ethnicity type corresponding to the maximum value. If the name is not listed within the dataset, it is classified as an unknown race. In this manner, the last names of authors are classified to a specific race/ethnicity type supported by the dataset.

Recommender System

The recommendation system is a tool provided to aid in getting research papers related to the citations inputted in order to diversify the ratio proportions of the author's gender and ethnicity. To obtain this, the arXiv API has been implemented as described following section.

arXiv API

The arXiv API allows one to programmatically access the electronic publications on arXiv.org. This API is intended to provide similar interface as the arXiv human web interface that has the ability to access hundreds of thousands of academic prints. Similar to other APIs, one can make a request with the parameters encoded in the URL. The search query allows to search based on a string, via an id of the publication, or through publication author name. Multiple other parameters options such sorting order and maximum results to get are also available to be passed into the arXiv API.

In order to minimize errors that can occur dealing with requesting queries and getting right results, a python wrapper for the arXiv API, known as arxiv.py, was implemented for this part of the code. This wrapper provided similar functionalities as the API but had specific functions and variables designated to the parameters that could be used to fetch the results (Schwab, 2015/2022).

Preprocessing data

To accomplish the task of retrieving the reference recommendations, the data passed into the recommendation function with the wrapper implementation has to be preprocessed. When the BibTeX information is formatted as required, the data is used to extract the titles along with the author's names. These titles go through tokenization and removal of stop words and stored in a list that is later accessed by a function that searches for the recommendations using the arXiv API.

Because this part of the project required the usage of arXiv API in order to retrieve other related academic papers based on the titles passed in as an input, the python wrapper was used. However, there were complications on how the titles were being passed to get queried. Since the titles were extracted and passed into the search call, the query couldn't identify papers for the entire title and rather only searched for the first word and supplied the recommendations which were unrelated. In the effort to resolve this issue, first the titles passed in were URL encoded to understand that the whole string passed in was the title to be searched for. Since the python wrapper couldn't retrieve a long string that was specific to the title, the query search title fed in was shortened to three words which allowed to get the API response of recommendations.

Frontend and Visualization

Streamlit.io – Framework

To skip the learning curve of learning a new language such as JavaScript or HTML as well as speed up the effort to deploy the application, Python was chosen as the primary language for the code. As a means to reduce the complexity of the implementation of the website, using a

framework that would let to work on frontend development in Python as well was more appropriate. Based on suggestions from other users, Streamlit.io was chosen as the framework. It was the most appropriate choice as it is an open-source python framework that can be utilized for building and deploying web apps for Machine Learning and Data Science (*Streamlit • The Fastest Way to Build and Share Data Apps*, n.d.). Given it acts just like an additional package in python, it provides the ability to write and execute an app like python code.

Visualization

Infographics are a visual aid that helps understand information better. For the web application, first a pie plot was considered to display the likelihood predictions of gender and race/ethnicity. Individual categories represented as slices in pie plots are an effective way in visualizing the overall proportion of the statistics. However, it has been identified that donut plots are recommended to be a better alternative as the hollow in the center diverts the attention of the readers towards the circumference of the circle rather than the center (Robertson, 2017). Given it could be represented as a stacked bar graph, using a donut plot would provide more clarity in the data represented. Therefore, the donut plot was chosen as the plot for the infographics. In the code, however, a donut chart is plotted with the pie chart component.

Plotly Dash

In the context of this application, it was a debate between which choosing Dash by Plotly or Matplotlib to obtain the plots for the gender and race/ethnicity inferences. Both packages have their strengths and weaknesses. Dash is more sophisticated in its approaches and the package itself

can deploy web dashboards as specified in the code. It differs in the way it renders the data and would need more hands-on experience to exercise its maximum potential. In addition, Dash is not completely open source and has a subscription to use its full capabilities. However, Plotly's Dash provides visualization that is more interactive and aesthetic to users. It is faster compared to Matplotlib and requires less code to implement single components compared to Matplotlib.

On the other hand, Matplotlib has an ease of use and is simpler to interpret as it uses a sequential logic of implementation. It is a comprehensive library that has a variety of different plots that are most commonly used in for visualization in python as is a numerical extension for Numpy. Despite that, matplotlib is comparatively slower than Dash and the features within the components can lead to not giving exactly what is needed. As mentioned in the earlier section, before a pie plot was considered as it would be an easier way to perceive the results of the predictions. Because of this reason, first matplotlib was used to display the pie chart on the webapp. However, since donut plots present a better visualization and attract the attention of the viewers to the results rather than the center, they have been employed later in the process. To understand whether dash was a better option or matplotlib, both the libraries were used to implement the donut plot to provide an analysis. In the final analysis, it was observed that matplotlib donut plot included more work in order to get the data to be plotted and was much slower in comparison to dash donut plot and was also giving issues regarding the formatting of the labels with the plot itself. This allowed to conclude to use Plotly's Dash components for the donut plot for better visualization

Visuals of Web Application

The following section provides visuals of the web application and the functions each part of the website executes.

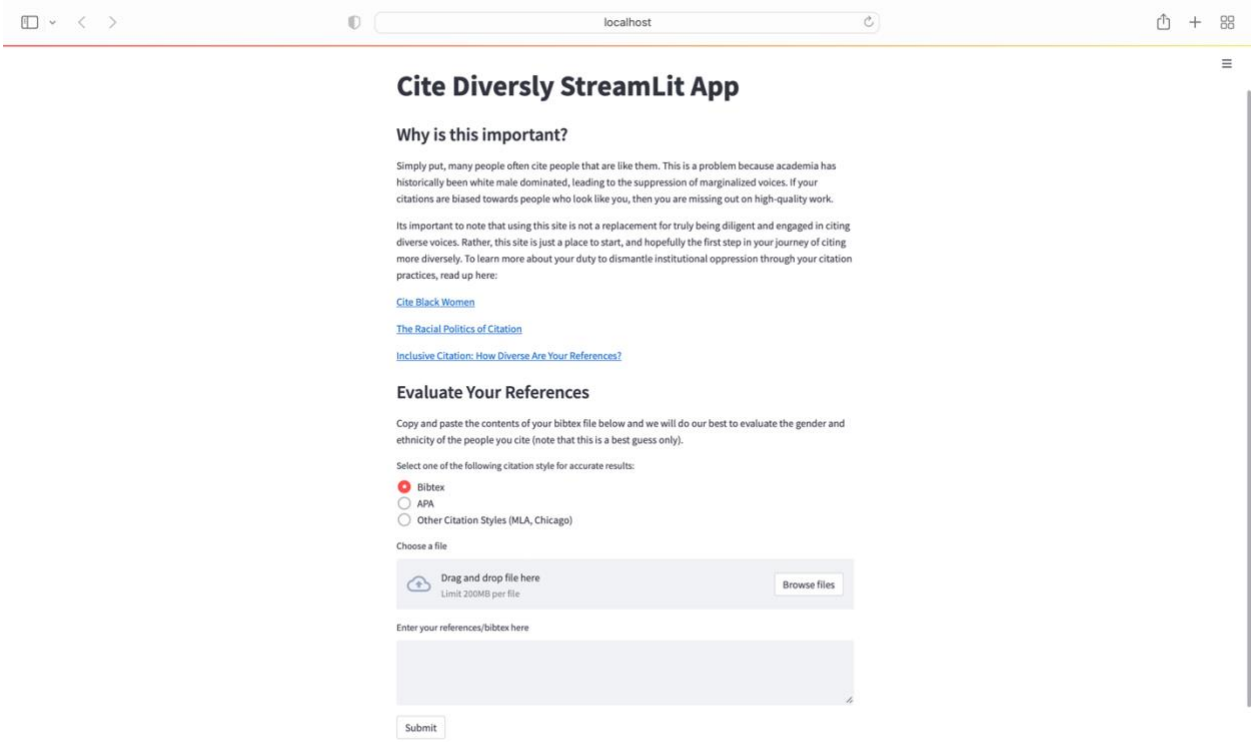

Figure 4: Cite Diversly Web Application

Figure 4 displays the overall application, constructed using Streamlit and hosted locally. When the website is pulled up, this is the initial screen presented to the users. As depicted, the user has the option to evaluate their references by selection one of the citation styles and either uploading a file or entering the references in the text area. By default, the BibTeX citation style is selected for the input.

Figure 5 below shows an example of inferring gender and race/ethnicity by uploading a file. When the file named test.txt is uploaded with the citation option style, the interface automatically starts running without needing to click on the submit button. When the file has been

read, the likelihoods of the genders and ethnicities of the authors are plotted, and the references are recommended.

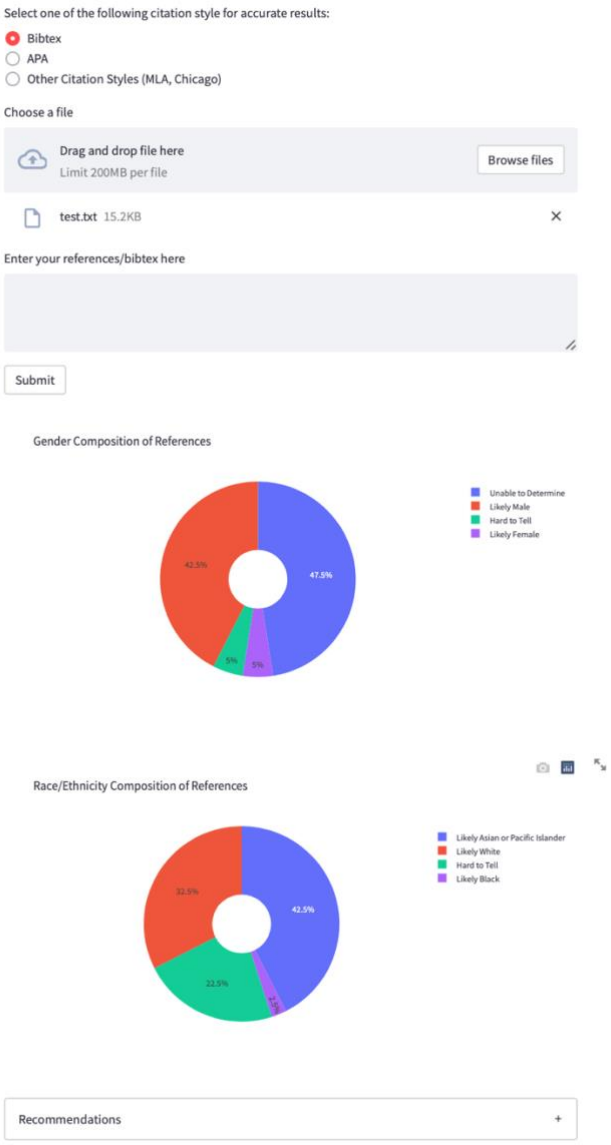

Figure 5: Example of File Input

Based on the titles extracted from the processed text file input, the recommendations from the arXiv API are retrieved which are then hyperlinked into the website as shown in *Figure 6*.

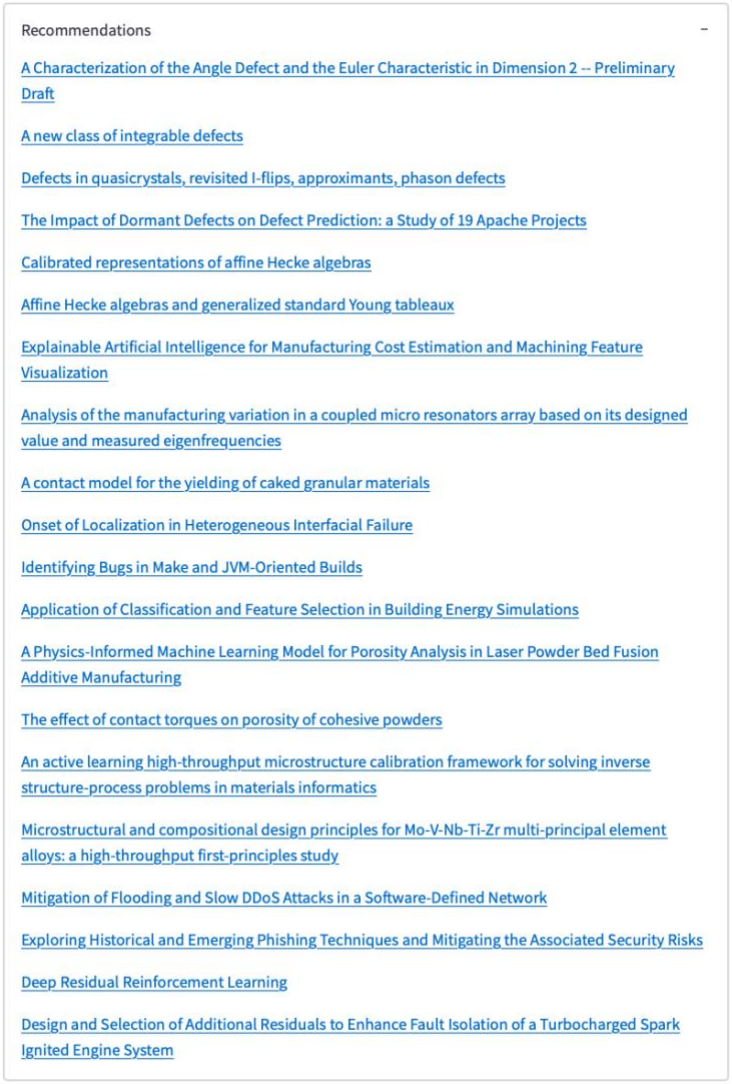

Figure 6: Recommendation of References

Figure 7 and Figure 8, exhibit the outcome of inputting APA citation style into the text area. As required, it is necessary to select the right option style when entering the reference into the text area Clicking on the submit button, only the race/ethnicity composition for the references is displayed as explained in the previous subsections. When the mouse hovers over the donut plot arc, it shows the type of race/ethnicity and how many citations within the input belonged to this category.

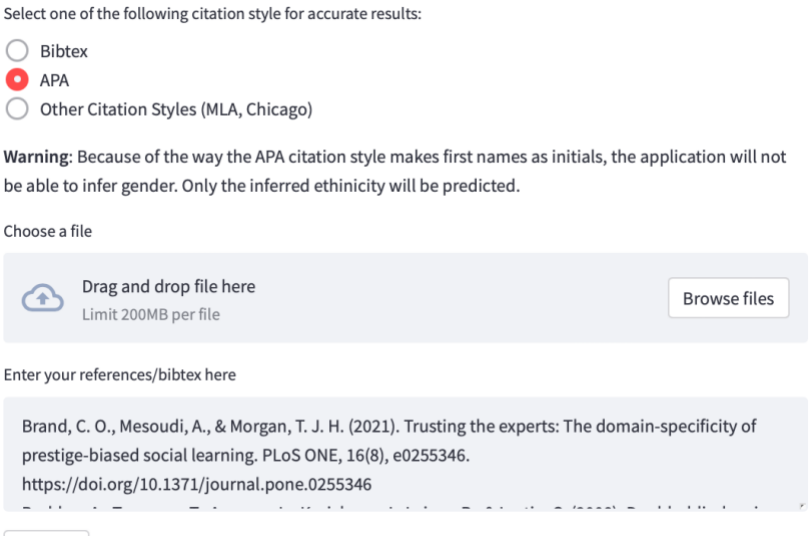

Figure 7: Inputting APA for Inference

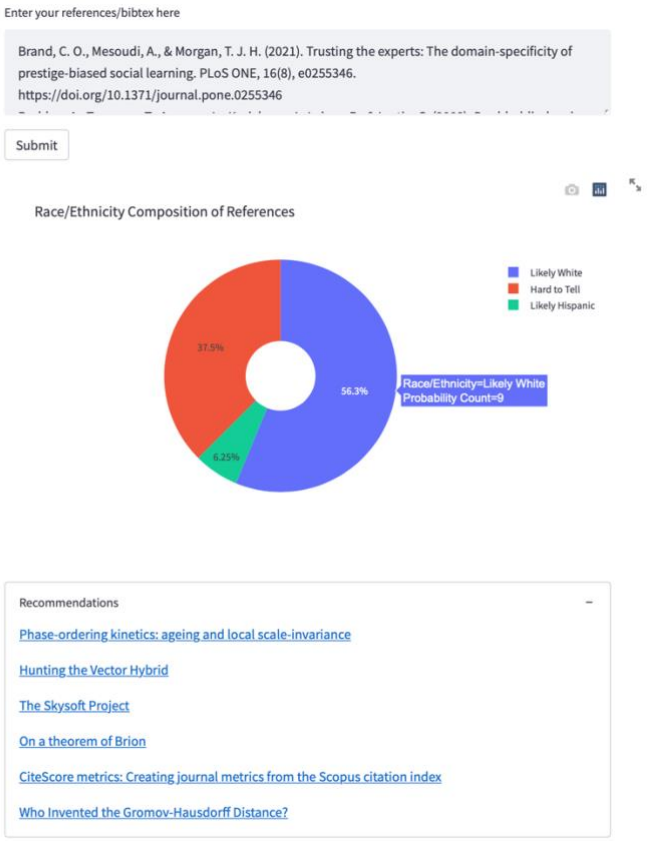

Figure 8: Ethnicity Inference for APA citation style

Chapter 4

System Testing

The following section reports the results obtained when a variety of different papers are tested on the system as well the performance accuracy of the algorithms.

Accuracy of Evaluations

Getting a good performance in the predictions is necessary for the effectiveness of the web application. To measure the correctness of the evaluations, authors of 15 publications have been extracted from the citations and have been classified based on most likely gender and most likely ethnicity manually. This manually labelled dataset was compared to the dataset of labels that have been predicted by the web application algorithm. Table 3 illustrates the names of the authors and their corresponding manual classification of gender and ethnicity labels as well as the inference classification of the labels.

Table 3: Gender and Ethnicity Label Comparisons - Manual Labeling vs Inference Labeling

First Name	Last Name	Manual Labeling		Inferences Labeling	
		Likely Gender	Likely Ethnicity	Likely Gender	Likely Ethnicity

Sean	Rismiller	Likely Male	Likely White	Likely Male	Likely White
Jonathan	Cagan	Likely Male	Likely White	Likely Male	Likely White
Joshua	Gyory	Likely Male	Likely White	Likely Male	Likely White
Christopher	McComb	Likely Male	Likely White	Likely Male	Likely White
Kenneth	Kotovsky	Likely Male	Likely White	Likely Male	Unable to Determine
Jonathan	Cagan	Likely Male	Likely White	Likely Male	Likely White
Christopher	McComb	Likely Male	Likely White	Likely Male	Likely White
George	Box	Likely Male	Likely White	Likely Male	Likely White
Saurabh	Basu	Likely Male	Likely Asian	Unable to Determine	Likely Asian or Pacific Islander
Norman	Draper	Likely Male	Likely White	Likely Male	Likely White
Joaquin	Quiñonero-Candela	Likely Male	Likely Hispanic	Unable to Determine	Unable to Determine
Carl	Rasmussen	Likely Male	Likely White	Likely Male	Likely White
Zhi-Lei	Wang	Likely Male	Likely Asian	Hard to Tell	Likely Asian or Pacific Islander
Toshio	Ogawa	Likely Male	Likely Asian	Likely Male	Likely Asian or Pacific Islander
Yoshitaka	Adachi	Likely Male	Likely Asian	Likely Male	Likely Asian or Pacific Islander
Chun-Teh	Chen	Likely Male	Likely Asian	Unable to Determine	Likely Asian or Pacific Islander
Grace	Gu	Likely Female	Likely Asian	Likely Female	Likely Asian or Pacific Islander
Bernard	Choi	Likely Male	Likely Asian	Likely Male	Likely Asian or Pacific Islander

Anita	Pak	Likely Female	Likely Asian	Likely Female	Likely Asian or Pacific Islander
Dan	Braha	Likely Male	Likely White	Likely Male	Likely White
Yoram	Reich	Likely Male	Likely White	Likely Male	Likely White
Benoit	Weil	Likely Male	Likely White	Unable to Determine	Likely White
Armand	Hatchuel	Likely Male	Likely White	Likely Male	Unable to Determine
Guido	Stompff	Likely Male	Likely White	Likely Male	Unable to Determine
Frido	Smulders	Likely Male	Likely White	Likely Male	Unable to Determine
Lilian	Henze	Likely Female	Likely White	Likely Female	Likely White
Anita	Pak	Likely Female	Likely Asian	Likely Female	Likely Asian or Pacific Islander
Bernard	Choi	Likely Male	Likely Asian	Likely Male	Likely Asian or Pacific Islander

The raw accuracy of the predictions was calculated by finding the percentage of correct predictions divided by the total number of predictions. As a result, it was evaluated that with a total number of 28 authors, the gender and ethnicity inferences performed with 82.14% accuracy each for this sample. This is a sufficiently good accuracy for the algorithm implementation approaches proposed in this study. There are possibilities of improving the performance of these evaluations that are further discussed in the next chapter.

Recommendation Incorporation Analysis

In an effort to understand the performance of the recommendations suggested by the web application, three files, consisting of 15 different publications each, were compared. The files

uploaded were tested using references in the BibTeX citation style to get the inferences. With the recommendations that are provided by the platform, they are parsed back into BibTeX using the arXiv id provided from the python wrapper and are appended to the initial file which was tested. This approach is taken to understand the performance of the web application on the diverseness it is indented to provide.

Table 4 illustrates the results of the gender compositions before adding the recommendations and after adding the recommendations to text File 1. As visible, a large proportion of the references before recommendations, about 43.5% of them, were predicted as likely male. While the second majority of the composition belongs to the "unable to determine" category, the "hard to tell" and likely female contribute an equal amount. The ratio of male to female authors is quite imbalanced. The addition of recommendations to the text file result in similar proportions. Even though there is only a slight reduction the male contributors, there is an increase with the female contributors in the papers.

Table 4: File 1 - Gender inference - Before & After Recommendations Addition

File 1				
Before Recommendations			**After Recommendations**	
Gender	**%**		**Gender**	**%**
Likely Male	43.5		Likely Male	43.3
Unable to Determine	43.5		Unable to Determine	43.3
Hard to Tell	6.45		Hard to Tell	5.22
Likely Female	6.45		Likely Female	8.21

Similarly, *Table 5* depicts the outcomes of the race/ethnicity inferences before and after adding the recommendations to File 1. A large composition of the publications in this file corresponded to likely White or likely Asian/Pacific islanders. When comparing the before and after recommendations aspect, most of the additional recommendations have been categorized as "Hard to tell" which highlights some of the drawbacks of the ethnicity inference algorithm.

Table 5: File 1 - Ethnicity inference - Before & After Recommendations Addition

File 1				
Before Recommendations			**After Recommendations**	
Ethnicity	**%**		**Ethnicity**	**%**
Likely White	43.5		Likely White	41
Likely Asian or Pacific Islander	33.9		Likely Asian or Pacific Islander	33.6
Hard to Tell	21		Hard to Tell	24.6
Likely Black	1.6		Likely Black	0.746

Moving on to File 2, Tables 6 and 7 provide the compositions of gender and ethnicity inferences before and after appending the recommendations to the file. The likelihood compositions for gender highlight that there is a high domination of male authors in the reference list. After the recommendations have been added, the "unable to determine" label had an increase. There are slight increases in the likely female and male category indicating some of the additions classified to those two labels. This suggests that the increase in the number of references mostly contributed to the "Unable to Determine" attribute and decreased the amount of "Hard to Tell" label.

Table 6: File 2 - Gender inference - Before & After Recommendations Addition

File 2				
Before Recommendations			**After Recommendations**	
Gender	**%**		**Gender**	**%**
Likely Male	54.8		Likely Male	54.9
Unable to Determine	16.1		Unable to Determine	20.9
Hard to Tell	9.68		Hard to Tell	3.3
Likely Female	19.4		Likely Female	20.9

The ethnicity inference for file 2 indicated by *Table 7* indicate that the publications mostly are associated with Likely White, Likely Asian or Pacific Islander and Likely Hispanic. Given that the proportion of likely whites were 51.6% initially, the addition of recommendations decreased this value to 33% as shown in the table. There is also a good increase to the contributions of likely Asian or Pacific Islander from 19.4 % to 30.8 %. While there was a slight addition to the Hard to Tell category, there was also good raise in the amount of Likely Hispanic from 3.23% to 6.59%. This indicates that the references recommended by the application overall contributed to increasing the diversity in the race/ethnicity background of the authors that are being cited.

Table 7: File 2 - Ethnicity inference - Before & After Recommendations Addition

File 2				
Before Recommendations			**After Recommendations**	
Ethnicity	**%**		**Ethnicity**	**%**
Likely White	51.6		Likely White	33

Likely Asian or Pacific Islander	19.4		Likely Asian or Pacific Islander	30.8
Hard to Tell	25.8		Hard to Tell	29.7
Likely Hispanic	3.23		Likely Hispanic	6.59

Like the above, Tables 8 and 9 give the comparisons of the gender and ethnicity inferences on File 3. The gender inference suggests that the likely male class is still dominant in File 3, however after the recommendations decreased the percentage suggesting increase in other categories. Given there is a decrease in the Likely Female from 15.4% to 11.5% suggests that there is a higher contribution to another label. As seen, there is a drastic change to the "Unable to Determine" category from 20.5% to 34.6% suggesting that the algorithm implemented is unable to classify a large contribution of the added recommendations.

Table 8: File 3 - Gender inference - Before & After Recommendations Addition

File 3				
Before Recommendations			**After Recommendations**	
Gender	**%**		**Gender**	**%**
Likely Male	53.8		Likely Male	44.2
Unable to Determine	20.5		Unable to Determine	34.6
Hard to Tell	10.3		Hard to Tell	9.62
Likely Female	15.4		Likely Female	11.5

Table 9 indicates that the ethnicity compositions in File 3 mostly contributed to the Likely Asian or Pacific Islander and Likely White categories. The addition of the recommendations

indicates a decrease in Likely White and a slight increase in Likely Asian or Pacific Islander. This suggests that most of the contributions in File 3 have been categorized as Hard to Tell category, indicated by the increase of percentage composition from 12.8% to 18.3%.

Table 9: File 3 - Ethnicity inference - Before & After Recommendations Addition

File 3				
Before Recommendations			**After Recommendations**	
Ethnicity	**%**		**Ethnicity**	**%**
Likely Asian or Pacific Islander	56.4		Likely Asian or Pacific Islander	56.7
Likely White	30.8		Likely White	25
Hard to Tell	12.8		Hard to Tell	18.3

As discerned from the discussion above, the recommendation system has sometimes provided reference recommendations that benefit to only ethnicity or only gender compositions. Additionally, sometimes the recommendations provided have a neutral impact and don't give a lot of contribution to increasing the richness of the diversity in the gender and ethnicity backgrounds of the authors. This indicates that the system needs to be improved in terms of gathering the recommendations based on the proportions of initial gender and ethnicity compositions.

Chapter 5

Conclusion & Future Work

Conclusion

Integrating diversity and inclusion is an important part of academia given its history in various types of citation biases that can exist within this realm. Various types of citation bias could influence the outcomes of a research and in turn affect the way a publication is received. This is disadvantageous to the authors as academic publications are one of the factors that impact the researcher's academic success and the ability to gain funding and recognition within their academic domain.

This thesis serves to provide a web-based platform that moves a step closer to minimizing the bias by allowing one to evaluate the ratio of the gender inclusiveness and ratio of ethnicity diverseness within the references used in their academic research papers. Given the input via a text file or through the text area, the system parses the data into a BibTeX format and segments out the titles and author names. Using python packages, the algorithm separates the list of names based on first and last and passes it into the gender inference and race/ethnicity inference functions to obtain the likelihoods of the genders and ethnicities of the authors. The distributions of results are displayed using a donut plot on the web-based application to give a visual representation of the diversity that the user has with their citations. Simultaneously, the titles extracted are preprocessed to retrieve academic publications that are recommended by the application to the user in order to

minimize the uniformity in bias. Through this method, users could incorporate the recommended publications to increase the inclusiveness in the female authors to male authors ratios and diverseness in the ethnicities of the authors, reducing any existing citation bias.

Future Work

Much of this project focused on the development of the initial web application that could draw the inference of the gender and race/ethnicity of the authors used for references. Eventually, the future goal is to optimize it aiming to increase the ease of use, providing more accurate results and strengthen the diversity aspect of other research papers. Several approaches have been identified to increase the effectiveness of the interface.

One of the primary features to be refined is the dataset range that is being used by the function that infers ethnicity. Currently, the dataset used for race/ethnicity inference incorporates only the census data from a single country, i.e., the United States. The next step would be to enlarge the range of the dataset by integrating datasets from a diverse range of countries, identifying various ethnicities via the first and last names. Multiple datasets can be stacked together to create a global dataset that would be more inclusive of different cultures and names from this dataset would be utilized by the ethnicity inference function. Given there are not many past works based on inferring ethnicity, methods to improve the accuracy of the current model will be considered. One improvement mentioned earlier regarding enriching the dataset would help enhance the model learning thus boosting the accuracy.

In terms of the development of the algorithms, better models are intended to be implemented for inferring gender. Similarly, the python package that is used for the gender inference, i.e., gender guesser utilizes a model that trained on euro-centric dataset. This limits the ability of the model to learn and predict the gender for names belonging to regions that are non-European, such as areas of North and South America, Middle Asia, Africa and Australia. As explained previously, there are many approaches already experimented for understanding the gender of an individual based on their first names. These existing models can be used to conduct a benchmarking study that allows to examine which model gives a better accuracy. This will improve the way the gender is predicted, reducing the percent of the unknown or unable to predict aspect.

For the recommendation system, it is intended to bring improvements to the way the query is passed into the API so that the API retrieves the published papers. Features such as sorting the search type, i.e., based on relevance, date published, or date submitted would allow the user to get a better range of publications based on their preference. Additionally, the arXiv API will be either supported or replaced by a python package, scholarly which retrieves information from Google Scholar. This is because Google Scholar expands the ability of improving the range of diversity such as providing articles, journals, and publications.

Appendix A

File 1

@article{SANAEI2019108091,
title = "Defect characteristics and analysis of their variability in metal L-PBF additive manufacturing",
journal = "Materials and Design",
volume = "182",
pages = "108091",
year = "2019",
issn = "0264-1275",
doi = "https://doi.org/10.1016/j.matdes.2019.108091",
url = "http://www.sciencedirect.com/science/article/pii/S0264127519305295",
author = "Niloofar Sanaei and Ali Fatemi and Nam Phan",
keywords = "Additive manufacturing, Defect characterization, Computed tomography, Defect variability, Ti-6Al-4V, 17-4 PH stainless steel",
abstract = "Additive manufacturing (AM) has provided an opportunity for fabricating complex parts. Fabricating these parts without defects is currently a challenge. Therefore, understanding AM defects is fundamental to the structural integrity of load carrying components, failure analysis, and defect-based modeling of mechanical performance. This work investigates defect content of metal AM specimens and correlations between defect characteristics (size, sphericity/circularity, aspect ratio) using 2D and 3D defect characterization techniques. Distributions of defect characteristics based on location throughout AM specimens were analyzed and the variabilities of defect characteristics within these specimens were studied. Laser-Based Power Bed Fusion (L-PBF) specimens manufactured with different metals, different AM machines and built directions, different surface conditions, and different thicknesses were evaluated. Significant variability in defect characteristics based on location, especially in as-built surface specimens was observed. Well-optimized process parameters and post-processing reduced the overall volume fraction of defects, and the specified variabilities, and resulted in a more random dispersion of defects around the specimens. 2D and 3D defect analysis showed similar trends regarding correlations between defect characteristics and provided complementary information about the actual defect content based on their resolution."
}
@article{GONG201487,
title = "Analysis of defect generation in Ti–6Al–4V parts made using powder bed fusion additive manufacturing processes",
journal = "Additive Manufacturing",
volume = "1-4",
pages = "87 - 98",
year = "2014",
note = "Inaugural Issue",
issn = "2214-8604",
doi = "https://doi.org/10.1016/j.addma.2014.08.002",
url = "http://www.sciencedirect.com/science/article/pii/S2214860414000074",
author = "Haijun Gong and Khalid Rafi and Hengfeng Gu and Thomas Starr and Brent Stucker",
keywords = "Defect, Ti–6Al–4V, SLM, EBM, Additive manufacturing",
abstract = "Ti–6Al–4V parts made using additive manufacturing processes such as selective laser melting (SLM) and electron beam melting (EBM) are subject to the inclusion of defects. This study purposely fabricated Ti–6Al–4V samples with defects by varying process parameters from the factory default settings in both SLM and EBM systems. Process parameters are classified according to their tendency to create certain types of porosity. Finally, defect characteristics are discussed with respect to defect generation mechanisms; and effective process windows for SLM and EBM system are discussed."
}
@article{YABANSU201526,
title = "Representation and calibration of elastic localization kernels for a broad class of cubic polycrystals",
journal = "Acta Materialia",
volume = "94",
pages = "26 - 35",
year = "2015",
issn = "1359-6454",
doi = "https://doi.org/10.1016/j.actamat.2015.04.049",
url = "http://www.sciencedirect.com/science/article/pii/S1359645415003018",

author = "Yuksel C. Yabansu and Surya R. Kalidindi",
keywords = "Materials Knowledge Systems, Localization kernels, Generalized spherical harmonics, Legendre polynomials, Hierarchical multiscale modeling",
abstract = "Localization kernels play an important role in the study of hierarchical material systems with well separated length scales. They allow for a computationally efficient communication of critical information between the constituent length scales. They are particularly well suited for capturing how an imposed variable (e.g., stress or strain) at the higher length scale is spatially distributed at the lower length scale (i.e., localization linkages). In recent work, our research group has presented a novel framework called Materials Knowledge Systems (MKS) for the representation and calibration of the localization kernels, and demonstrated the viability of this approach on selected individual material systems. In this work, we present and demonstrate an important extension to the MKS framework that allows representation and calibration of the localization kernels for an entire class of materials (e.g., a selected class of single phase cubic polycrystalline materials)."
}
@article{WANG2019852,
title = "Influence of manufacturing geometric defects on the mechanical properties of AlSi10Mg alloy fabricated by selective laser melting",
journal = "Journal of Alloys and Compounds",
volume = "789",
pages = "852 - 859",
year = "2019",
issn = "0925-8388",
doi = "https://doi.org/10.1016/j.jallcom.2019.03.135",
url = "http://www.sciencedirect.com/science/article/pii/S0925838819309429",
author = "Panding Wang and Hongshuai Lei and Xiaolei Zhu and Haosen Chen and Daining Fang",
keywords = "A. Selective laser melting, B. AlSi10Mg, C. Manufacturing defect, D. Image-based finite element model, E. Additive manufacturing",
abstract = "The tensile behavior of bulk AlSi10Mg components, fabricated by selective laser melting (SLM), was investigated by uniaxial tensile testing and image-based finite element simulation. The initial morphological features of the structures were imaged by micro X-ray tomography. Moreover, the reconstructed model and as-designed model were compared to quantify the process-induced defects, which remained unavoidable due to complex manufacturing processes. The un-melted AlSi10Mg powders, sticking to the melting pool after condensation, intensified the deviation of side edges. Futhermore, the unevenly distributed process-induced defects resulted in anisotropic mechanical properties of AlSi10Mg alloy. Two finite element models were developed from X-ray tomography images and CAD model, which were simulated by finite element solver ABAQUS/Standard to discuss the effect of initial morphological features on the mechanical behavior of these samples. The geometric defects have slightly reduced Young's modulus and yield strength, but remarkably increased the equivalent plastic strain of the bulk structures. Furthermore, the ultimate strength and elongation, predicted by the image-based finite element model and the ductile failure criterion, was much lower than the values predicted by the as-designed model due to the influence of geometric defects."
}
@article{LIU2017160,
title = "Elastic and failure response of imperfect three-dimensional metallic lattices: the role of geometric defects induced by Selective Laser Melting",
journal = "Journal of the Mechanics and Physics of Solids",
volume = "107",
pages = "160 - 184",
year = "2017",
issn = "0022-5096",
doi = "https://doi.org/10.1016/j.jmps.2017.07.003",
url = "http://www.sciencedirect.com/science/article/pii/S0022509616307608",
author = "Lu Liu and Paul Kamm and Francisco García-Moreno and John Banhart and Damiano Pasini",
abstract = "This paper examines three-dimensional metallic lattices with regular octet and rhombicuboctahedron units fabricated with geometric imperfections via Selective Laser Sintering. We use X-ray computed tomography to capture morphology, location, and distribution of process-induced defects with the aim of studying their role in the elastic response, damage initiation, and failure evolution under quasi-static compression. Testing results from in-situ compression tomography show that each lattice exhibits a distinct failure mechanism that is governed not only by cell topology but also by geometric defects induced by additive manufacturing. Extracted from X-ray tomography images, the statistical distributions of three sets of defects, namely strut waviness, strut thickness variation, and strut oversizing, are used to develop numerical models of statistically representative lattices with imperfect geometry. Elastic and failure responses are predicted within 10% agreement from the experimental data. In addition, a computational study is presented to shed light into the relationship between the amplitude of selected defects and the reduction of elastic properties compared to their nominal values. The evolution of failure mechanisms is also explained with respect to strut oversizing, a parameter that can critically cause failure mode transitions that are not visible in defect-free lattices."
}

@article{ZHAO201876,
title = "Effect of building direction on porosity and fatigue life of selective laser melted AlSi12Mg alloy",
journal = "Materials Science and Engineering: A",
volume = "729",
pages = "76 - 85",
year = "2018",
issn = "0921-5093",
doi = "https://doi.org/10.1016/j.msea.2018.05.040",
url = "http://www.sciencedirect.com/science/article/pii/S0921509318306890",
author = "Junwen Zhao and Mark Easton and Ma Qian and Martin Leary and Milan Brandt",
keywords = "Additive manufacturing, Selective laser melting (SLM), Gas porosity, Aluminum alloy, Fatigue life prediction",
abstract = "Gas porosity is one of the most common defects in aluminum alloy parts manufactured by solidification processing, and can have a strong influence on fatigue properties. This study shows that gas pores with a fraction of 0.2–1.6% and an average size of 20–55 μm are present in the Al-Si alloy parts manufactured by Selective Laser Melting (SLM). Failure after fatigue testing was found to initiate from surface or subsurface gas pores and fatigue life prediction equations were developed considering the influence of pores. The building direction did not have a statistically verifiable effect on the average gas porosity fraction, size and distribution, although the scatter in porosity fraction was greater in the vertically built specimens. At the same applied stress, the fatigue life of SLM manufactured specimens decreased with an increase in pore size, and specimens built horizontally exhibited a greater fatigue life than those built vertically. The cause is attributed to greater propensity of cracks to propagate along lower strength melt pool boundary layers in vertically built specimens."
}
@article{Sci-rep_Prangnell_crack_tip,
title = "The Influence of Porosity on Fatigue Crack Initiation in Additively Manufactured Titanium Components",
journal = "Scientific Reports",
volume = "7",
number = "7308",
year = "2017",
author = "S. Tammas Williams and P. J. Withers and I. Todd and P. B. Prangnell",
}

@article{THIJS20103303,
title = "A study of the microstructural evolution during selective laser melting of Ti–6Al–4V",
journal = "Acta Materialia",
volume = "58",
number = "9",
pages = "3303 - 3312",
year = "2010",
issn = "1359-6454",
doi = "https://doi.org/10.1016/j.actamat.2010.02.004",
url = "http://www.sciencedirect.com/science/article/pii/S135964541000090X",
author = "Lore Thijs and Frederik Verhaeghe and Tom Craeghs and Jan Van Humbeeck and Jean-Pierre Kruth",
keywords = "Selective laser melting, Additive manufacturing, Laser treatment, Titanium alloys, Optical microscopy",
abstract = "Selective laser melting (SLM) is an additive manufacturing technique in which functional, complex parts can be created directly by selectively melting layers of powder. This process is characterized by highly localized high heat inputs during very short interaction times and will therefore significantly affect the microstructure. In this research, the development of the microstructure of the Ti–6Al–4V alloy processed by SLM and the influence of the scanning parameters and scanning strategy on this microstructure are studied by light optical microscopy. The martensitic phase is present, and due to the occurrence of epitaxial growth, elongated grains emerge. The direction of these grains is directly related to the process parameters. At high heat inputs it was also found that the intermetallic phase Ti3Al is precipitated during the process."
}
@article{MUKHERJEE2018442,
title = "Mitigation of lack of fusion defects in powder bed fusion additive manufacturing",
journal = "Journal of Manufacturing Processes",
volume = "36",
pages = "442 - 449",
year = "2018",
issn = "1526-6125",
doi = "https://doi.org/10.1016/j.jmapro.2018.10.028",
url = "http://www.sciencedirect.com/science/article/pii/S1526612518302603",
author = "T. Mukherjee and T. DebRoy",

keywords = "Powder bed fusion, Lack of fusion defect, Heat transfer and fluid flow, Marangoni convection, Non-dimensional temperature",
abstract = "Components manufactured by additive manufacturing often exhibit improper fusion among layers and hatches. Currently, there is no practical way to select process parameters and alloy systems based on scientific principles to mitigate these defects. Here, we develop, test and demonstrate a methodology to predict and prevent these defects based on a numerical heat transfer and fluid flow model for the laser powder bed fusion (PBF) additive manufacturing (AM). These defects are avoidable by adjusting laser power, scanning speed, layer thickness and hatch spacing. An easy to use parameter is proposed for practical use in shop floors. Relative susceptibilities of three widely used AM alloys are demonstrated using this parameter."
}
@article{VASTOLA2016231,
title = "Controlling of residual stress in additive manufacturing of Ti6Al4V by finite element modeling",
journal = "Additive Manufacturing",
volume = "12",
pages = "231 - 239",
year = "2016",
note = "Special Issue on Modeling and Simulation for Additive Manufacturing",
issn = "2214-8604",
doi = "https://doi.org/10.1016/j.addma.2016.05.010",
url = "http://www.sciencedirect.com/science/article/pii/S2214860416300951",
author = "G. Vastola and G. Zhang and Q.X. Pei and Y.-W. Zhang",
keywords = "Residual stress, Electron beam melting, Ti6Al4V, Additive manufacturing, Powder metallurgy",
abstract = "Minimizing the residual stress build-up in metal-based additive manufacturing plays a pivotal role in selecting a particular material and technique for making an industrial part. In beam-based additive manufacturing, although a great deal of effort has been made to minimize the residual stresses, it is still elusive how to do so by simply optimizing the manufacturing parameters, such as beam size, beam power, and scan speed. With reference to the Ti6Al4V alloy and manufacturing by electron beam melting, we perform systematic finite element modeling of one-pass scanning to study the effects of beam size, beam power density, beam scan speed, and chamber bed temperature on the magnitude and distribution of residual stresses. Our study elucidates both qualitative and quantitative features of the residual stress fields originated by these manufacturing parameters. Our findings can serve as useful guidelines for engineers and designers to deal with residual stress build-up during additive manufacturing of Ti6Al4V."
}
@inproceedings{liu2014,
author = {Liu, Qian Chu and Elambasseril, Joe and Sun, Shou Jin and Leary, Martin and Brandt, Milan and Sharp, Peter Khan},
title = {The Effect of Manufacturing Defects on the Fatigue Behaviour of Ti-6Al-4V Specimens Fabricated Using Selective Laser Melting},
year = {2014},
month = {5},
volume = {891},
pages = {1519--1524},
booktitle = {11th International Fatigue Congress},
series = {Advanced Materials Research},
publisher = {Trans Tech Publications Ltd},
doi = {10.4028/www.scientific.net/AMR.891-892.1519},
keywords = {Fatigue Crack Propagation, Fatigue Life, Fatigue Crack Initiation, Ti-6Al-4V Titanium Alloy, Defect, Selective Laser Melt (SLM), Lack-of-Fusion (LOF)},
abstract = {Additive Manufacturing (AM) technologies are considered revolutionary because they could fundamentally change the way products are designed. Selective Laser Melting (SLM) is a metal based AM process with significant and growing potential for the manufacture of aerospace components. Traditionally a material needs to be listed in the Metallic Materials Properties Development and Standardization (MMPDS) handbook if it is to be considered certified. However, this requires a considerable amount of test data to be generated on the materials mechanical properties. Therefore, the MMPDS certification process does not lend itself easily to the certification of AM components as the final component can have similar mechanical properties to wrought alloys combined with the defects associated with traditional casting and welding technologies. These defects can substantially decrease the fatigue life of a fabricated component. The primary purpose of this investigation was to study the fatigue behaviour of as-built Ti-6Al-4V (Ti64) samples. Fatigue tests were performed on the Ti-6Al-4V specimens built using SLM with a variety of layer thicknesses and build (vertical or horizontal) directions. Fractography revealed the presence of a range of manufacturing defects located at or near the surface of the specimens. The experimental results indicated that Lack-of-Fusion (LOF) defects were primarily responsible for fatigue crack initiation. The reduction in fatigue life appeared to be affected by the location, size and shape of the LOF defect.}
}

@Article{Rodgers2019,

author="Rodgers, Theron M.
and Lim, Hojun
and Brown, Judith A.",
title="Three-Dimensional Additively Manufactured Microstructures and Their Mechanical Properties",
journal="JOM",
year="2019",
month="Oct",
day="30",
abstract="Metal additive manufacturing (AM) allows for the freeform creation of complex parts. However, AM microstructures are highly sensitive to the process parameters used. Resulting microstructures vary significantly from typical metal alloys in grain morphology distributions, defect populations and crystallographic texture. AM microstructures are often anisotropic and possess three-dimensional features. These microstructural features determine the mechanical properties of AM parts. Here, we reproduce three ``canonical'' AM microstructures from the literature and investigate their mechanical responses. Stochastic volume elements are generated with a kinetic Monte Carlo process simulation. A crystal plasticity-finite element model is then used to simulate plastic deformation of the AM microstructures and a reference equiaxed microstructure. Results demonstrate that AM microstructures possess significant variability in strength and plastic anisotropy compared with conventional equiaxed microstructures.",
issn="1543-1851",
doi="10.1007/s11837-019-03808-x",
url="https://doi.org/10.1007/s11837-019-03808-x"
}

@article{damage_tolerant_Architectured_materials,
Author = {Pham, Minh-Son and Liu, Chen and Todd, Iain and Lertthanasarn, Jedsada},
Journal = {Nature},
Number = {7739},
Pages = {305--311},
Title = {Damage-tolerant architected materials inspired by crystal microstructure},
Volume = {565},
Year = {2019}}
@article{POLONSKY2020249,
title = "Solidification-driven orientation gradients in additively manufactured stainless steel",
journal = "Acta Materialia",
volume = "183",
pages = "249 - 260",
year = "2020",
issn = "1359-6454",
doi = "https://doi.org/10.1016/j.actamat.2019.10.047",
url = "http://www.sciencedirect.com/science/article/pii/S1359645419307189",
author = "Andrew T. Polonsky and William C. Lenthe and McLean P. Echlin and Veronica Livescu and George T. Gray and Tresa M. Pollock",
keywords = "Additive manufacturing, Tomography, Solidification, Microstructure, Tribeam",
abstract = "A sample of 304L stainless steel manufactured by Laser Engineered Net Shaping (LENS) was characterized in 3D using TriBeam tomography. The crystallographic, structural, and chemical properties of the as-deposited microstructure have been studied in detail. 3D characterization reveals complex grain morphologies and large orientation gradients, in excess of 10â^ ˜ , that are not easily interpreted from 2D cross-sections alone. Misorientations were calculated via a methodology that locates the initial location and orientation of grains that grow during the build process. For larger grains, misorientation increased along the direction of solidification. For grains with complex morphologies, K-means clustering in orientation space is demonstrated as a useful approach for determining the initial growth orientation. The gradients in misorientation directly tracked with gradients in chemistry predicted by a Scheil analysis. The accumulation of misorientation is linked to the solutal and thermal solidification path, offering potential design pathways for novel alloys more suited for additive manufacturing."
}

@article{doi:10.1146/annurev-matsci-070115-031816,
author = {Collins, P.C. and Brice, D.A. and Samimi, P. and Ghamarian, I. and Fraser, H.L.},
title = {Microstructural Control of Additively Manufactured Metallic Materials},
journal = {Annual Review of Materials Research},
volume = {46},
number = {1},
pages = {63-91},

year = {2016},
doi = {10.1146/annurev-matsci-070115-031816},

URL = { https://doi.org/10.1146/annurev-matsci-070115-031816},
eprint = {
https://doi.org/10.1146/annurev-matsci-070115-031816

}
,
abstract = { In additively manufactured (AM) metallic materials, the fundamental interrelationships that exist between composition, processing, and microstructure govern these materialsâ€ ™properties and potential improvements or reductions in performance. For example, by using AM, it is possible to achieve highly desirable microstructural features (e.g., highly refined precipitates) that could not otherwise be achieved by using conventional approaches. Simultaneously, opportunities exist to manage macro-level microstructural characteristics such as residual stress, porosity, and texture, the last of which might be desirable. To predictably realize optimal microstructures, it is necessary to establish a framework that integrates processing variables, alloy composition, and the resulting microstructure. Although such a framework is largely lacking for AM metallic materials, the basic scientific components of the framework exist in literature. This review considers these key components and presents them in a manner that highlights key interdependencies that would form an integrated framework to engineer microstructures using AM. }
}

File 2

@book{Kirton2003,
abstract = {Adaption-Innovation theory (A-I theory) is a model of problem solving and creativity, which aims to increase collaboration and reduce conflict within groups. A-I Theory and the associated Kirton Adaption-Innovation (KAI) inventory have been extensively researched and are increasingly used as tools for teambuilding and personnel management. In Adaption-Innovation: In the Context of Change and Diversity, Kirton outlines the central concepts of the theory, including the processes of problem solving, decision making and creativity.},
author = {Kirton, M. J.},
booktitle = {Adaption-Innovation: In the Context of Diversity and Change},
doi = {10.4324/9780203695005},
isbn = {0203695003},
issn = {0-415-29851-2},
mendeley-groups = {THRED Lab},
pmid = {21015},
title = {{Adaption-innovation: In the context of diversity and change}},
year = {2003}
}
@article{Collins1964,
abstract = {Marine microalgae and cyanobacteria are very rich in several chemical compounds and, therefore, they may be used in several biological applications related with health benefits, among others. This review brings the research up-to-date on the bioactive compounds produced by marine unicellular algae, directly or indirectly related to humanhealth. It covers and goes through themost studied applications of substances such asPUFA, sterols,proteins and enzymes, vitamins and pigments, in areas so diverse as humanand animal nutrition, therapeutics, and aquacul- ture. The great potential of marine microalgae and the biocoumpounds they produce are discussed in this review.},
author = {Collins, Barry E and Guetzkow, Harold},
isbn = {0471165816},
journal = {Industrial Management Review},
mendeley-groups = {THRED Lab},
title = {{A Social Psychology of Group Processes for Decision-Making}},
year = {1964}
}
@article{Kurtzberg2005,
abstract = {Two empirical studies explored objectively measured creative fluency and subjectively perceived creativity in cognitively diverse teams. Results indicate that cognitive diversity may be beneficial for objective functioning but may damage team satisfaction, affect, and members' impressions of their creative performance. Subjective ratings diverged greatly from more

objective measures and were more closely related to affective measures. The overall findings present creativity as a complex multidimensional construct, and cognitive diversity as an important predictor of both team emotions and outcomes. Arguments are presented for the value of subjectively perceived creativity, even in the absence of more concrete performance in the immediate time period.},
author = {Kurtzberg, Terri R.},
doi = {10.1207/s15326934crj1701_5},
isbn = {1040-0419},
issn = {10400419},
journal = {Creativity Research Journal},
mendeley-groups = {THRED Lab},
pmid = {16236864},
title = {{Feeling creative, being creative: An empirical study of diversity and creativity in teams}},
year = {2005}
}
@article{Levitt2012,
abstract = {The Virtual Design Team: Designing Project Organizations as Engineers Design Bridges},
annote = {comprehensive hitory of VDT, continue reading from p18},
author = {Levitt, Raymond E},
doi = {10.7146/jod.6345},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Levitt - 2012 - The Virtual Design Team Designing Project Organizations as Engineers Design Bridges.pdf:pdf},
issn = {2245-408X},
journal = {Journal of Organization Design},
keywords = {Virtual design team,organization design,project organization design,virtual design team},
mendeley-groups = {THRED Lab},
number = {2},
pages = {14},
title = {{The Virtual Design Team: Designing Project Organizations as Engineers Design Bridges}},
volume = {1},
year = {2012}
}
@article{Kilicay-Ergin2012,
abstract = {Cognitive architectures provide a promising means to model human behavior in complex systems and problem domains. The fields of social simulation and cognitive architectures can be linked more effectively if cognitive variability can be modeled in a realistic way. In particular, if individual differences in the ways humans solve problems can be captured in computational models, the dynamic patterns of change and diversity in human systems can be explored in greater depth. Kirton's Adaption-Innovation theory provides a robust foundation for the study of creativity, problem solving, and decision making based on individual differences in cognitive level (capacity) and cognitive style (preferred approach) of problem solving. This paper examines four well-known cognitive architectures (SOAR, ACT-R, CLARION, and DUAL) in light of Adaption-Innovation theory to explore if and how cognitive style and level variables are manifested within them. This analysis leads to a proposed cognitive style continuum for cognitive architectures, as well as other possible architectural mechanisms to incorporate problem-solving variability.},
annote = {compares 4 cognitive architectures in the context of A-I theory

three parts of cognitive problem solving: affect (need/value/belief/attitude), effect (style and potential), resources},
author = {Kilicay-Ergin, Nil H. and Jablokow, Kathryn W},
doi = {10.1109/TSMCC.2012.2201469},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Kilicay-Ergin, Jablokow - 2012 - Problem-solving variability in cognitive architectures.pdf:pdf},
isbn = {1094-6977$\backslash$n1558-2442},
issn = {10946977},
journal = {IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews},
keywords = {Adaption-Innovation (A-I) theory,cognitive architectures,cognitive style,creativity,problem solving},
mendeley-groups = {THRED Lab},
number = {6},
pages = {1231--1242},
pmid = {22107904},
title = {{Problem-solving variability in cognitive architectures}},
volume = {42},
year = {2012}
}

@article{Bonabeau2002,
abstract = {Agent-based modeling is a powerful simulation modeling technique that has seen a number of applications in the last few years, including applications to real-world business problems. After the basic principles of agent-based simulation are briefly introduced, its four areas of application are discussed by using real-world applications: flow simulation, organizational simulation, market simulation, and diffusion simulation. For each category, one or several business applications are described and analyzed.},
archivePrefix = {arXiv},
arxivId = {1709.03423},
author = {Bonabeau, Eric},
doi = {10.1073/pnas.082080899},
eprint = {1709.03423},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Bonabeau - 2002 - Agent-based modeling methods and techniques for simulating human systems.pdf:pdf},
isbn = {0027-8424 (Print)$\backslash$r0027-8424 (Linking)},
issn = {0027-8424},
journal = {Proceedings of the National Academy of Sciences},
mendeley-groups = {THRED Lab},
number = {suppl. 3},
pages = {7280--7287},
pmid = {12011407},
title = {{Agent-based modeling: methods and techniques for simulating human systems.}},
volume = {99},
year = {2002}
}
@article{Herath2017,
abstract = {Purpose: This paper aims at simulating on how “disorganization” affects team problem solving. The prime objective is to determine how team problem solving varies between an organized and disorganized environment also considering motivational aspects. Design/methodology/approach: Using agent-based modeling, the authors use a real-world data set from 226 volunteers at five different types of non-profit organizations in Southwest England to define some attributes of the agents. The authors introduce the concepts of natural, structural and functional disorganization while operationalizing natural and functional disorganization. Findings: The simulations show that “disorganization” is more conducive for problem solving efficiency than “organization” given enough flexibility (range) to search and acquire resources. The findings further demonstrate that teams with resources above their hierarchical level (access to better quality resources) tend to perform better than teams that have only limited access to resources. Originality/value: The nuanced categories of “(dis-)organization” allow us to compare between various structural limitations, thus generating insights for improving the way managers structure teams for better problem solving. {\textcopyright} 2017, {\textcopyright} Emerald Publishing Limited.},
annote = {teams of 7 volunteers, basic resource-access modeling

model is available online},
author = {Herath, Dinuka and Costello, Joyce and Homberg, Fabian},
doi = {10.1108/TPM-10-2015-0046},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Herath, Costello, Homberg - 2017 - Team problem solving and motivation under disorganization – an agent-based modeling approach.pdf:pdf},
isbn = {1855213656},
issn = {13527592},
journal = {Team Performance Management},
keywords = {Agent-based modeling,Disorganization,Problem solving},
mendeley-groups = {THRED Lab},
number = {1-2},
pages = {46--65},
title = {{Team problem solving and motivation under disorganization – an agent-based modeling approach}},
volume = {23},
year = {2017}
}
@article{McComb2017,
abstract = {The performance of a team with the right characteristics can exceed the mere sum of the constituent members' individual efforts. However, a team having the wrong characteristics may perform more poorly than the sum of its individuals. Therefore, it is vital that teams are assembled and managed properly in order to maximize performance. This work examines how the properties of configuration design problems can be leveraged to select the best values for team characteristics (specifically team size and interaction frequency). A computational model of design teams which has been shown to effectively emulate human team behavior is employed to pinpoint optimized team characteristics for solving a variety of configuration design

problems. These configuration design problems are characterized with respect to the local and global structure of the design space, the alignment between objectives, and the resources allotted for solving the problem. Regression analysis is then used to create equations for predicting optimized values for team characteristics based on problem properties. These equations achieve moderate to high accuracy, making it possible to design teams based on those problem properties. Further analysis reveals hypotheses about how the problem properties can influence a team's search for solutions. This work also conducts a cognitive study on a different problem to test the predictive equations. For a configuration problem of moderate size, the model predicts that zero interaction between team members should lead to the best outcome. A cognitive study of human teams verifies this surprising prediction, offering partial validation of the predictive theory.},
author = {McComb, Christopher and Cagan, Jonathan and Kotovsky, Kenneth},
doi = {10.1115/1.4035793},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/McComb, Cagan, Kotovsky - 2017 - Optimizing Design Teams Based on Problem Properties Computational Team Simulations and an Applied Empir.pdf:pdf},
issn = {1050-0472},
journal = {Journal of Mechanical Design},
mendeley-groups = {THRED Lab},
number = {4},
pages = {041101},
title = {{Optimizing Design Teams Based on Problem Properties: Computational Team Simulations and an Applied Empirical Test}},
volume = {139},
year = {2017}
}
@article{Zhang2017,
author = {Zhang, Guanglu and Mcadams, Daniel A},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Zhang, Mcadams - 2017 - PRODUCT PERFORMANCE EVOLUTION PREDICTION BY LOTKA-VOLTERRA EQUATIONS.pdf:pdf},
keywords = {DETC2017-67369,development planning,lotka-volterra equations,product,product performance,technology evolution,technology prediction},
mendeley-groups = {THRED Lab},
pages = {1--8},
title = {{PRODUCT PERFORMANCE EVOLUTION PREDICTION BY LOTKA-VOLTERRA EQUATIONS}},
year = {2017}
}
@article{Siggelkow2005,
abstract = {We use an innovative technique to examine an enduring but recently neglected question: How do environmental turbulence and complexity affect the appropriate formal design of organizations? We construct an agent-based simulation in which multidepartment firms with different designs face environments whose turbulence and complexity we control. The model's results produce two sets of testable hypotheses. One set pinpoints formal designs that cope well with three different environments: turbulent settings, in which firms must improve their performance speedily; complex environments, in which firms must search broadly; and settings with both turbulence and complexity, in which firms must balance speed and search. The results shed new light on longstanding notions such as equifinality. The other set of hypotheses argues that the impact of individual design elements on speed and search often depends delicately on specific powers granted to department heads, creating effects that run contrary to conventional wisdom and intuition. Ample processing power at the bottom of a firm, for instance, can slow down the improvement and narrow the search of the firm as a whole. Differences arise between our results and conventional wisdom when conventional thinking fails to account for the powers of department heads—powers to withhold information about departmental options, to control decision-making agendas, to veto firmwide alternatives, and to take unilateral action. Our results suggest how future empirical studies of organizational design might be fruitfully coupled with rigorous agent-based modeling efforts.},
annote = {semi-interesting paper on how a manager's level of power affects team},
author = {Siggelkow, Nicolaj and Rivkin, Jan W},
doi = {10.1287/orsc.1050.0116},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Siggelkow, Rivkin - 2005 - Speed and Search Designing Organizations for Turbulence and Complexity.pdf:pdf},
isbn = {1160210101},
issn = {1047-7039},
journal = {Organization Science},
keywords = {complexity,interactions,organizational design,simulation model,turbulence},
mendeley-groups = {THRED Lab},
number = {2},
pages = {101--122},
pmid = {16993879},

title = {{Speed and Search: Designing Organizations for Turbulence and Complexity}},
volume = {16},
year = {2005}
}
@article{Mccomb2015,
author = {McComb, Christopher and Cagan, Jonathan and Kotovsky, Kenneth},
doi = {10.1016/j.destud.2015.06.005},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Mccomb, Cagan, Kotovsky - 2015 - Lifting the Veil Drawing insights about design teams from a cognitively-inspired computational model.pdf:pdf},
issn = {0142-694X},
journal = {Design Studies},
keywords = {computational model,design cognition,engineering,teamwork},
mendeley-groups = {THRED Lab},
pages = {119--142},
publisher = {Elsevier Ltd},
title = {{Lifting the Veil: Drawing insights about design teams from a cognitively-inspired computational model}},
volume = {40},
year = {2015}
}
@article{Wu,
abstract = {Continuous technological innovation has been playing a vital role in ensuring the survival and development of an enterprise in today's economy. This paper studies the problem of technological innovation risk-based decision-making from an entrepreneurial team point of view. We identify the differences between this team decision-making and a traditional individual decision-making problem, where decisions are mainly affected by the decision-maker's risk and value perceptions, and risk preferences. We create a modeling framework for such a new problem, and use system dynamics theory to model it from the agent-based modeling perspective. The proposed approach is validated by a case study of the technological innovation risk decision-making in a Chinese automobile company. {\textcopyright} 2010 Elsevier Inc.},
annote = {ABM for team vs individual decision making},
author = {Wu, Desheng Dash and Kefan, Xie and Hua, Liu and Shi, Zhao and Olson, David L},
doi = {10.1016/j.techfore.2010.01.015},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Wu et al. - Unknown - Modeling technological innovation risks of an entrepreneurial team using system dynamics An agent-based perspectiv.pdf:pdf},
isbn = {0040-1625},
issn = {00401625},
journal = {Technological Forecasting and Social Change},
keywords = {Agent-based modeling,Entrepreneurial team,Risk-based decision-making (RDM),System dynamics,Technological innovation risk},
mendeley-groups = {THRED Lab},
number = {6},
pages = {857--869},
pmid = {20307717},
title = {{Modeling technological innovation risks of an entrepreneurial team using system dynamics: An agent-based perspective}},
volume = {77},
year = {2010}
}
@article{Chang,
annote = {general ABM for organizaitons, for economists},
author = {Chang, Myong-hun and {Harrington Jr}, Joseph E},
doi = {10.1016/S1574-0021(05)02026-5},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Chang, Harrington Jr - Unknown - AGENT-BASED MODELS OF ORGANIZATIONS.pdf:pdf},
mendeley-groups = {THRED Lab},
title = {{AGENT-BASED MODELS OF ORGANIZATIONS *}},
}
@article{Oliver2011,
abstract = {An existing gap in the literature on founding-leaders and small business growth is, "What happens to strategically important socially-created resources and performance as a micro-firm grows?" Using agent based modeling, we examine the development of such a resource, CONTEXT-FOR-LEARNING, and a performance potential across the first hierarchical structural change. The model that mimics a solo work group (2 levels: individual and group) and that which has two work groups (4 levels: individual, group, leadership team, and organization). Most of our hypotheses were confirmed but one interesting

difference was that some marginal conditions that supported maintaining a socially-created resource resulted in worse performance levels. [ABSTRACT FROM AUTHOR]},
author = {Oliver, Richard L and Black, Janice a},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Oliver, Black - 2011 - Micro-Business Hierarchies, Strategic Intangible Resources {\&} Performance.pdf:pdf},
issn = {21552843},
journal = {Journal of Marketing Development {\&} Competitiveness},
keywords = {BUSINESS enterprises,BUSINESS planning,EXPANSION (Business),SMALL business,UNITED States},
mendeley-groups = {THRED Lab},
number = {5},
pages = {9--28},
title = {{Micro-Business Hierarchies, Strategic Intangible Resources {\&} Performance.}},
volume = {5},
year = {2011}
}
@article{Garcia2005,
abstract = {Little has been written in the new product development literature about the simulation technique agent-based modeling, which is a by-product of recent explorations into complex adaptive systems in other disciplines. Agent-based models (ABM) are commonly used in other social sciences to represent individual actors (or groups) in a dynamic adaptive system. The social system may be a marketplace, an organization, or any type of system that acts as a collective of individuals. Agents represent autonomous decision-making entities that interact with each other and/or with their environment based on a set of rules. These rules dictate the behavioral choices of the agents. In these simulation models, heterogeneous agents interact with each other in a repetitive process. It is from the interactions between agents that aggregate macroscale behaviors or trends emerge. The simulated environment can be thought of as a "virtual" society in which actions taken by one agent may have an effect on the resulting actions of another agent. This article is an introduction to the ABM methodology and its possible uses for innovation and new product development researchers. It explores the benefits and issues with modeling dynamic systems using this methodology. Benefits of ABMs found in sociology and management studies have found that as the heterogeneity of individuals increase in a system or as network effects become more important in a system, the effectiveness of ABMs as a methodology increases. Additionally, the more adaptive a system or the more the system evolves over time, the greater the opportunity to learn more about the adaptive system using ABMs. Limitations to using this methodology include some knowledge of computer-programming techniques. Three potential areas of research are introduced: diffusion of innovations, organizational strategy, and knowledge and information flows. A common use of ABMs in the extant literature has been the modeling of the diffusion process between networked heterogeneous agents. ABMs easily allow the modeling of different types of networks and the impact of these networks on the diffusion process. A demonstrative example of an agent-based model to address the research question of how should manufacturers allocate resources to research (exploration) and development projects is provided. Future courses of study using ABMs also are explored.},
annote = {general paper on ABM},
author = {Garcia, Rosanna},
doi = {10.1111/j.1540-5885.2005.00136.x},
file = {:Users/samlapp/Library/Application Support/Mendeley Desktop/Downloaded/Garcia - 2005 - Uses of agent-based modeling in innovationnew product development research.pdf:pdf},
isbn = {0737-6782},
issn = {07376782},
journal = {Journal of Product Innovation Management},
mendeley-groups = {THRED Lab},
number = {5},
pages = {380--398},
title = {{Uses of agent-based modeling in innovation/new product development research}},
volume = {22},
year = {2005}
}

File 3

@article{Rismiller2021,
abstract = {Products must often endure challenging conditions while fulfilling their intended functions. Game-theoretic methods can readily create a wide variety of these conditions to consider when creating designs. This work introduces

Cognitively Inspired Adversarial Agents (CIAAs) that use a Stackelberg game format to generate designs resistant to these conditions. These agents are used to generate designs while considering a multidimensional attack. Designs are produced under these adversarial conditions and compared to others generated without considering adversaries to confirm the agents' performance. The agents create designs able to withstand multiple combined conditions.},
  author = {Sean C. Rismiller and Jonathan Cagan and Christopher McComb},
  doi = {10.1115/1.4049862},
  issn = {1050-0472},
  issue = {3},
  journal = {Journal of Mechanical Design},
  month = {3},
  title = {An Adversarial Agent-Based Design Method Using Stochastic Stackelberg Game Conditions},
  volume = {143},
  url = {https://asmedigitalcollection.asme.org/mechanicaldesign/article/doi/10.1115/1.4049862/1096685/An-Adversarial-Agent-Based-Design-Method-Using},
  year = {2021},
}
@article{Gyory2019,
  author = {Joshua T. Gyory and Jonathan Cagan and Kenneth Kotovsky},
  doi = {10.1007/s00163-018-00303-3},
  issn = {0934-9839},
  issue = {1},
  journal = {Research in Engineering Design},
  month = {1},
  pages = {85-102},
  title = {Are you better off alone? Mitigating the underperformance of engineering teams during conceptual design through adaptive process management},
  volume = {30},
  url = {http://link.springer.com/10.1007/s00163-018-00303-3},
  year = {2019},
}
@article{Khanolkar2021,
  author = {Pranav Milind Khanolkar and Christopher Carson McComb and Saurabh Basu},
  doi = {10.1016/j.commatsci.2020.110068},
  issn = {09270256},
  journal = {Computational Materials Science},
  month = {1},
  pages = {110068},
  title = {Predicting elastic strain fields in defective microstructures using image colorization algorithms},
  volume = {186},
  url = {https://linkinghub.elsevier.com/retrieve/pii/S0927025620305590},
  year = {2021},
}
@book{Box1987,
  author = {George E.P. Box and Norman R. Draper},
  title = {Empirical Model-Building and Response Surfaces},
  year = {1987},
}
@article{,
  abstract = {We provide a new unifying view, including all existing proper probabilistic sparse approximations for Gaussian process regression. Our approach relies on expressing the effective prior which the methods are using. This allows new insights to be gained, and highlights the relationship between existing methods. It also allows for a clear theoretically justified ranking of the closeness of the known approximations to the corresponding full GPs. Finally we point directly to designs of new better sparse approximations, combining the best of the existing strategies, within attractive computational constraints.},
  author = {Joaquin Quiñonero-Candela and Carl Edward Rasmussen},
  issn = {15337928},
  journal = {Journal of Machine Learning Research},
  keywords = {Bayesian committee machine,Gaussian process,Probabilistic regression,Sparse approximation},
  title = {A unifying view of sparse approximate Gaussian process regression},
  year = {2005},
}
@article{doi:10.1002/adts.201900177,

author = {Wang, Zhi-Lei and Ogawa, Toshio and Adachi, Yoshitaka},
title = {A Machine Learning Tool for Materials Informatics},
journal = {Advanced Theory and Simulations},

volume = {n/a},
number = {n/a},
pages = {1900177},
keywords = {image analysis, inverse analysis, machine learning, materials informatics, topological microstructures},
doi = {10.1002/adts.201900177},
url = {https://onlinelibrary.wiley.com/doi/abs/10.1002/adts.201900177},
eprint = {https://onlinelibrary.wiley.com/doi/pdf/10.1002/adts.201900177},
abstract = {Abstract In response to the increasing demand for the highly efficient design of materials, materials informatics has been proposed for using data and computational sciences to extract data features that provide insight into how properties track with microstructure variables. However, the general metrics of microstructural features often ignore the complexities of the microstructure geometry for many properties of interest. An independently developed machine learning tool called shiny materials genome integration system for phase and property analysis (ShinyMIPHA), which is designed with either standalone software or cloud system based on an R programing package of “Shiny”, is introduced. ShinyMIPHA provides topological microstructure analysis methods based on image processing technology by employing a two-point correlation function, persistent homology, and mean (H)–Gauss (K) curvature approaches, as well as sparse study and regression analysis methods that enable a data-driven properties-to-microstructure-to-processing inverse materials-design approach. The demo version is available at https://adachi-lab.shinyapps.io/demo/.}
}

@article{doi:10.1002/adts.201900056,
author = {Chen, Chun-Teh and Gu, Grace X.},
title = {Effect of Constituent Materials on Composite Performance: Exploring Design Strategies via Machine Learning},
journal = {Advanced Theory and Simulations},

volume = {2},
number = {6},
pages = {1900056},
keywords = {composites, finite elements, graphene, machine learning, molecular dynamics},
doi = {10.1002/adts.201900056},
url = {https://onlinelibrary.wiley.com/doi/abs/10.1002/adts.201900056},
eprint = {https://onlinelibrary.wiley.com/doi/pdf/10.1002/adts.201900056},
abstract = {Abstract Nature assembles a range of biological composites with remarkable mechanical properties despite being composed of relatively weak polymeric and ceramic components. However, the architectures of biomaterials cannot be considered as optimal designs for engineering applications since biomaterials are constantly evolving for multiple functions beyond carrying external loading. Here, it is aimed to develop an intelligent approach to design superior composites from scratch—starting from constituent materials. A systematic computational investigation of the effect of constituent materials (assumed to be perfectly brittle) on the behavior of composites using an integrated approach combining finite element method, molecular dynamics, and machine learning (ML) is reported. It is demonstrated that instead of using brute-force methods, machine learning is a much more efficient approach and can generate optimal designs with similar performance to those obtained from an exhaustive search. Furthermore, it is shown that the toughening and strengthening mechanism observed in composites at the continuum-scale by combining stiff and soft constituents is valid for nanomaterials as well. Results show that high-performing designs of graphene nanocomposites can be generated using our ML approach. This novel ML-based design framework can be applied to other material systems to study a variety of structure–property relationships over several length-scales.},
year = {2019}
}
@article{Liu_scientific_report,
title = "An efficient machine learning approach to establish structure-property linkages",
journal = "Scientific Reports",
volume = "5",
pages="11551",
year = "2015",
author = "Ruoqian Liu and Abhishek Kumar and Zhengzhang Chen and Ankit Agrawal and Veera Sundararaghavan and Alok Choudhary"
}

@article{JUNG201917,
title = "An efficient machine learning approach to establish structure-property linkages",
journal = "Computational Materials Science",
volume = "156",
pages = "17 - 25",
year = "2019",
issn = "0927-0256",
doi = "https://doi.org/10.1016/j.commatsci.2018.09.034",
url = "http://www.sciencedirect.com/science/article/pii/S0927025618306335",
author = "Jaimyun Jung and Jae Ik Yoon and Hyung Keun Park and Jin You Kim and Hyoung Seop Kim",
keywords = "Microstructure, Machine learning, Gaussian process regression, Optimization",
abstract = "Full-field simulations with synthetic microstructure offer unique opportunities in predicting and understanding the linkage between microstructural variables and properties of a material prior to or in conjunction with experimental efforts. Nevertheless, the computational cost restrains the application of full-field simulations in optimizing materials microstructures or in establishing comprehensive structure-property linkages. To address this issue, we propose the use of machine learning technique, namely Gaussian process regression, with a small number of full-field simulation results to construct structure-property linkages that are accurate over a wide range of microstructures. Furthermore, we demonstrate that with the implementation of expected improvement algorithm, microstructures that exhibit most desirable properties can be identified using even smaller number of full-field simulations."
}
@article{Choi2008,
abstract = {Background/Purpose. Multiple disciplinary efforts are increasingly encouraged in health research, services, education and policy. This paper is the third in a series. The first discussed the definitions, objectives, and evidence of effectiveness of multiple disciplinary teamwork. The second examined the promoters, barriers, and ways to enhance such teamwork. This paper addresses the questions of discipline, inter-discipline distance, and where to look for multiple disciplinary collaboration. Methods. This paper proposes a conceptual framework of the knowledge universe, based on a review of a number of key papers on the Global Brain. These key papers were identified during a literature review on multiple disciplinary teamwork, using Google and MEDLINE (1982-2007) searches. Results. A discipline is held together by a shared epistemology. In general, disciplines that are more disparate from one another epistemologically are more likely to achieve new insight for a complex problem. The proposed conceptual framework of the knowledge universe consists of several knowledge subsystems, each containing a number of disciplines. The inter-discipline distance can guide us to select appropriate disciplines for a multiple disciplinary team. Conclusion. If multiple disciplinarity is called for, the proposed view of the knowledge universe as a series of knowledge subsystems and disciplines, and the place of health sciences in the knowledge universe, will help researchers, practitioners, and policy makers to identify disciplines for multiple disciplinary efforts.},
author = {Bernard C. K. Choi and Anita W. Pak},
doi = {10.25011/cim.v31i1.3140},
issn = {1488-2353},
issue = {1},
journal = {Clinical & Investigative Medicine},
month = {2},
pages = {41},
title = {Multidisciplinarity, interdisciplinarity, and transdisciplinarity in health research, services, education and policy: 3. Discipline, inter-discipline distance, and selection of discipline},
volume = {31},
url = {http://cimonline.ca/index.php/cim/article/view/3140},
year = {2008},
}
@article{Braha2003,
author = {Dan Braha and Yoram Reich},
doi = {10.1007/s00163-003-0035-3},
issn = {0934-9839},
issue = {4},
journal = {Research in Engineering Design},
month = {11},
pages = {185-199},
title = {Topological structures for modeling engineering design processes},
volume = {14},
url = {http://link.springer.com/10.1007/s00163-003-0035-3},
year = {2003},
}
@article{Hatchuel2009,

author = {Armand Hatchuel and Benoit Weil},
doi = {10.1007/s00163-008-0043-4},
issn = {0934-9839},
issue = {4},
journal = {Research in Engineering Design},
month = {1},
pages = {181-192},
title = {C-K design theory: an advanced formulation},
volume = {19},
url = {http://link.springer.com/10.1007/s00163-008-0043-4},
year = {2009},
}
@article{Stompff2016,
author = {Guido Stompff and Frido Smulders and Lilian Henze},
doi = {10.1016/j.destud.2016.09.004},
issn = {0142694X},
journal = {Design Studies},
month = {11},
pages = {187-214},
title = {Surprises are the benefits: reframing in multidisciplinary design teams},
volume = {47},
url = {https://linkinghub.elsevier.com/retrieve/pii/S0142694X15300375},
year = {2016},
}
@article{Choi2006,
author = {Bernard C. K. Choi and Anita W. Pak},
journal = {Clinical & Investigative Medicine},
pages = {351-364},
title = {Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness},
volume = {6},
year = {2006},
}
@article{Suh1998,
author = {Nam P. Suh},
doi = {10.1007/s001639870001},
issn = {0934-9839},
issue = {4},
journal = {Research in Engineering Design},
month = {12},
pages = {189-209},
title = {Axiomatic Design Theory for Systems},
volume = {10},
url = {http://link.springer.com/10.1007/s001639870001},
year = {1998},
}

BIBLIOGRAPHY

Ambekar, A., Ward, C., Mohammed, J., Male, S., & Skiena, S. (2009). Name-ethnicity classification from open sources. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 49–58. https://doi.org/10.1145/1557019.1557032

Bourke, J., & Titus, A. (2019, March 29). *Why Inclusive Leaders Are Good for Organizations, and How to Become One*. https://hbr.org/2019/03/why-inclusive-leaders-are-good-for-organizations-and-how-to-become-one

Brand, C. O., Mesoudi, A., & Morgan, T. J. H. (2021). Trusting the experts: The domain-specificity of prestige-biased social learning. *PLoS ONE*, *16*(8), e0255346. https://doi.org/10.1371/journal.pone.0255346

Budden, A., Tregenza, T., Aarssen, L., Koricheva, J., Leimu, R., & Lortie, C. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution*, *23*(1), 4–6. https://doi.org/10.1016/j.tree.2007.07.008

Cain, D. M., & Detsky, A. S. (2008). Everyone's a Little Bit Biased (Even Physicians). *JAMA*, *299*(24), 2893–2895. https://doi.org/10.1001/jama.299.24.2893

Campbell, F. M. (1990). National bias: A comparison of citation practices by health professionals. *Bulletin of the Medical Library Association*, *78*(4), 376–382.

Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, *1*(6), 1–5.

Carpenter, C. R., Cone, D. C., & Sarli, C. C. (2014). Using Publication Metrics to Highlight Academic Productivity and Research Impact. *Academic Emergency Medicine*, *21*(10), 1160–1172. https://doi.org/10.1111/acem.12482

De Cruz, H. (2018). Prestige Bias: An Obstacle to a Just Academic Philosophy. *Ergo, an Open Access Journal of Philosophy*, *5*(20201214). https://doi.org/10.3998/ergo.12405314.0005.010

Drieschová, A. (2020). Failure, persistence, luck and bias in academic publishing. *New Perspectives*, *28*(2), 145–149. https://doi.org/10.1177/2336825X20911792

Frachtenberg, E., & McConville, K. S. (2022). Metrics and methods in the evaluation of prestige bias in peer review: A case study in computer systems conferences. *PLOS ONE*, *17*(2), e0264131. https://doi.org/10.1371/journal.pone.0264131

Galdas, P. (2017). Revisiting Bias in Qualitative Research: Reflections on Its Relationship With Funding and Impact. *International Journal of Qualitative Methods*. https://doi.org/10.1177/1609406917748992

gender-verification by forename (cmd-line-tool & db)—Utilities. (n.d.). AutoHotkey Community. Retrieved April 4, 2022, from http://www.autohotkey.com/board/topic/20260-gender-verification-by-forename-cmd-line-tool-db/

Ghiasi, G., Larivière, V., & Sugimoto, C. (2015). On the Compliance of Women Engineers with a Gendered Scientific System. *PloS One*, *10*, e0145931. https://doi.org/10.1371/journal.pone.0145931

Herrera, A. J. (1999). Language bias discredits the peer-review system. *Nature*, *397*(6719), 467–467. https://doi.org/10.1038/17194

Homaeipour, S. (2018). *Exploring citation patterns of male and female scholars in Physics*. 74.

Hu, Y., Hu, C., Tran, T., Kasturi, T., Joseph, E., & Gillingham, M. (2021). What's in a Name? -- Gender Classification of Names with Character Based Machine Learning Models. *ArXiv:2102.03692 [Cs]*. http://arxiv.org/abs/2102.03692

Kaufman, R. R., & Chevan, J. (2011). The Gender Gap in Peer-Reviewed Publications by Physical Therapy Faculty Members: A Productivity Puzzle. *Physical Therapy*, *91*(1), 122–131. https://doi.org/10.2522/ptj.20100106

Keil, S. (2022). *AnyStyle* [Ruby]. https://github.com/inukshuk/anystyle (Original work published 2011)

Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide. *EJIFCC*, *25*(3), 227–243.

King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. (2017). Men Set Their Own Cites High: Gender and Self-citation across Fields and over Time. *Socius*, *3*, 2378023117738903. https://doi.org/10.1177/2378023117738903

Lawani, S. M. (1986). Some bibliometric correlates of quality in scientific research. *Scientometrics*, *9*(1–2), 13–25. https://doi.org/10.1007/BF02016604

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*(1), 2–17. https://doi.org/10.1002/asi.22784

Leone, M. (2020). Diversity and Inclusion in the Workplace; Benefits, Challenges and Strategies for Success. *School of Professional Studies*. https://commons.clarku.edu/sps_masters_papers/42

Link, A. M. (1998). US and Non-US SubmissionsAn Analysis of Reviewer Bias. *JAMA*, *280*(3), 246–247. https://doi.org/10.1001/jama.280.3.246

Nair, N., & Vohra, N. (2015). *Diversity and Inclusion at the Workplace: A Review of Research and Perspectives*. *2015*, 36.

Nanchahal, K., Mangtani, P., Alston, M., & dos Santos Silva, I. (2001). Development and validation of a computerized South Asian Names and Group Recognition Algorithm (SANGRA) for use in British health-related studies. *Journal of Public Health*, *23*(4), 278–285. https://doi.org/10.1093/pubmed/23.4.278

Nichani, A. S. (2013). Whose manuscript is it anyway? The ‘Write’ position and number of authors…. *Journal of Indian Society of Periodontology*, *17*(3), 283–284. https://doi.org/10.4103/0972-124X.115630

Porterfield, S. (2021, August 30). *10 Diversity & Inclusion Statistics That Will Change How You Do Business*. https://blog.bonus.ly/diversity-inclusion-statistics

Reiter-Palmon, R., & Illies, J. J. (2004). Leadership and creativity: Understanding leadership from a creative problem-solving perspective. *The Leadership Quarterly*, *15*(1), 55–77. https://doi.org/10.1016/j.leaqua.2003.12.005

Rezek, I., McDonald, R. J., & Kallmes, D. F. (2012). Pre-residency Publication Rate Strongly Predicts Future Academic Radiology Potential. *Academic Radiology*, *19*(5), 632–634. https://doi.org/10.1016/j.acra.2011.11.017

Robertson, A. (2017, January 22). Pie Chart vs. Donut Chart: Showdown in the Ring. *Medium*. https://medium.com/@hypsypops/pie-chart-vs-donut-chart-showdown-in-the-ring-5d24fd86a9ce

Rothchild, J. (2007). Gender Bias. In *The Blackwell Encyclopedia of Sociology*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405165518.wbeosg011

Schachner, M. K. (2019). From equality and inclusion to cultural pluralism – Evolution and effects of cultural diversity perspectives in schools. *European Journal of Developmental Psychology*, *16*(1), 1–17. https://doi.org/10.1080/17405629.2017.1326378

Schwab, L. (2022). *Arxiv.py* [Python]. https://github.com/lukasschwab/arxiv.py (Original work published 2015)

Septiandri, A. A. (2017). Predicting the Gender of Indonesian Names. *ArXiv:1707.07129 [Cs]*. http://arxiv.org/abs/1707.07129

Streamlit • The fastest way to build and share data apps. (n.d.). Retrieved April 4, 2022, from https://streamlit.io/

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of Sciences*, *114*(48), 12708–12713. https://doi.org/10.1073/pnas.1707323114

Tvina, A., Spellecy, R., & Palatnik, A. (2019). Bias in the Peer Review Process: Can We Do Better? *Obstetrics & Gynecology*, *133*(6), 1081–1083. https://doi.org/10.1097/AOG.0000000000003260

What is the significance of academic journals? (2019, May 12). Editage Insights. https://www.editage.com/insights/what-is-the-significance-of-journals

ACADEMIC VITA

Venkata Sai Renusree Bandaru
sairenusree.bv@gmail.com
www.linkedin.com/in/venkata-sai-renusree-bandaru/

Education

B.S. Computer Science, College of Engineering
Graduation: May 2022
The Pennsylvania State University, PA Schreyer's Honors College, Dean's List

Technical Work Experience

Fixed Income Trading Technology Intern, Bank Of New York Mellon Capital Markets (remote)
Aug,2021 – Dec,2021

- Responsible for assisting in the development of software applications to ease the utilization of Capital Market data
- Improving skills in VBA, Python and Java and developing soft skills like presentation skills and team building

Information Technology Services Lab & Teaching Assistant – Tech TA and Tutor Supervisor
Oct,2019 – Present

- Responsible for and assisting students and faculty with usage of technology tools and software to accomplish teaching, learning and/or business requirements. Improves communicative and time management skills.
- Supervise 7 Tech TAs/Tutors and simultaneously support classes every week and one-time events.

Projects

Covid-19 Visualizer – HackHers2021 – Merck Best HealthCare Hack
(**Language** – Python, **Tools** – VSCode)

- Web based application that helps visualize which countries use a certain Covid-19 vaccine using a color-coded world map and a drop-down feature that allows one to view different global usages of various vaccines
- Using Python libraries including Pandas, NumPy, Dash, Plotly, and Matplotlib.

Flixter – iOS Application
(**Language** – Swift, **Tools** – Movie Database API, Cocoa Pods, XCode)

- Built a movie browsing app that lets users view and scroll through a list of movies; users can view the movies in table view or collection view and tap on the cell for more details on the movie.

Student Database Project - UI project that was collaborated with a diverse team of three members
(**Language** – C++)

- Created a UI project that is a database of students and their information, i.e., personal details, as well as course accomplishments for the semesters, that can be viewed in various formats.
- Included usage of file I/O, linked lists, multidimensional arrays, calculations, handling header files with various classes and constructors, sorting and searching.

Budget Helper – HackHers 2019 – Fiserv Best "Spend-Wise" Hack
(**Language** – Java)

- Worked on a GUI project prototype that allows the user to precisely plan and budget for a bank for the company Fiserv
- Used JSwing, Derby Database, Arrays, I/O functionalities.

Research Experience

Multi-Campus (MCREU) Undergraduate Research Assistant
May – August 2021, May – August 2020

- Summer 2021 – Integrated ML concepts by using python packages (OpenCV) for processing and performing image segmentation on a leaf dataset for a system of plant disease detection
- Summer 2020 - Analyzed Python code for developing code simulating Hidden Markov Models algorithm to input EEG data and identify cognitive characteristics of the learning stages

Skills

Languages: C++, Java, Python, C, Swift, SQL
Tools: Jupyter Notebook, XCode, GitHub, NetBeans, VS Code Microsoft Office, Adobe Illustrator, Autodesk Fusion 360, Zoom

Leadership and Involvement

- Society of Women Engineers Club (SWE) - Director of Student Transitions *2021 -2022*
- iOS Development Tech Fellow for GWC College Loop PSU *2020-2021*
- Peer STEM Tutor for Penn State Libraries *2019-2020*
- President of the Women in Engineering Club (WIE) *2019-2020*

Honors & Awards

- Student of the Year (Female) – Penn State Brandywine *2019-2020*
- President's Sparks Award *2019-2020*
- President's Freshman Award *2018-2019*