

Increasing the Visibility of AI Digital Humanities Projects

Jon Chun
Kenyon College

2024 MLA Annual Convention
January 4th-7th 2024 Philadelphia, PA
<https://github.com/jon-chun/mla-generative-ai>

Overview

- Introduction
- Human-Centered AI
- Research Methodology
- Information Design & Editing

Introduction

DH@Kenyon



**Research
Innovation
Collaboration
Communication**

AI DIGITAL HUMANITIES @ KENYON COLLEGE

Integrated Program for Humane Studies (IPHS)

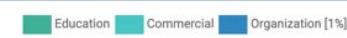
Digital humanities at Kenyon encompasses a dual lens. On the one hand, we empower our next generation of thinkers with the conceptual framework underlying our Age of Information — from dataism and algorithmic thinking to synthetic biology and artificial intelligence. Our unique approach to computational thinking positions students to engage with the many practical, theoretical and ethical issues surrounding technological innovation and social change.

DIGITAL HUMANITIES AT KENYON



2.793 Institutions

Institutions



Filter by type: [Show All](#)

Show 25 entries

Name

- Kenyon College
 - HT
 - Stanford University
 - Taiwan Academic Network
 - Chinese Academy of Social Sciences
 - University of Maryland
 - Ministry of Education Computer Center
 - New York University
 - University of Florida
 - Fundacao Para a Ciencia e a Tecnologia I.P.

- Stichting Duwo
 - The University of Melbourne
 - University of Arizona
 - NUS
 - University of Warwick
 - Massachusetts Institute of Technology
 - University of California at Berkeley
 - Korean Education Network
 - China Education and Research Network
 - Aalto University
 - Carnegie Mellon University
 - NTU
 - University of California Los Angeles
 - The University of Edinburgh
 - Columbia University



144
Total Papers

39,217
Total Downloads

22,139
Downloads in the past year

- Data Analytics
 - Machine Learning
 - DNN
 - NLP
 - Multimodal Affective AI
 - LLM/LMM Generative AI
 - FATE/XAI
 - Open Source Models
 - Metrics and Benchmarks
 - Human-AI Alignment and AI Safety

164 Countries

Show 25 entries

Regions

- [➤ United States](#)
 - [➤ United Kingdom](#)
 - [➤ Germany](#)
 - [➤ India](#)
 - [➤ Canada](#)
 - [➤ China](#)
 - [➤ Australia](#)
 - [➤ Philippines](#)
 - [➤ Singapore](#)
 - [➤ Netherlands](#)
 - [➤ Korea, Republic Of](#)
 - [➤ Hong Kong](#)
 - [➤ France](#)
 - [➤ Japan](#)
 - [➤ Turkey](#)
 - [➤ Italy](#)
 - [➤ Taiwan](#)
 - [➤ Indonesia](#)
 - [➤ Brazil](#)
 - [➤ Malaysia](#)



Featured



Student Scholarship



College Archives



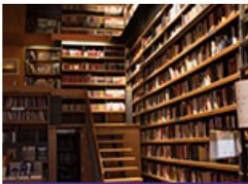
Special Collections



Educational Resources



Local History



All Collections

Jenna Nolt

Digital Initiatives Librarian

Campus

Chalmers Library, 314

P [\(740\) 427-5698](tel:(740)427-5698)

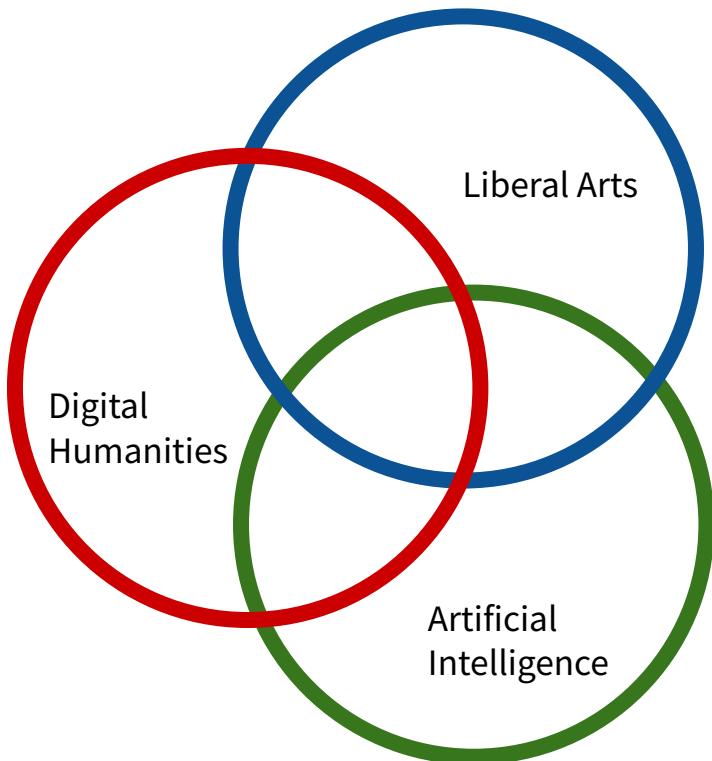
E noltj@kenyon.edu

Human-Centered AI



Kenyon Digital Humanities K{DH}

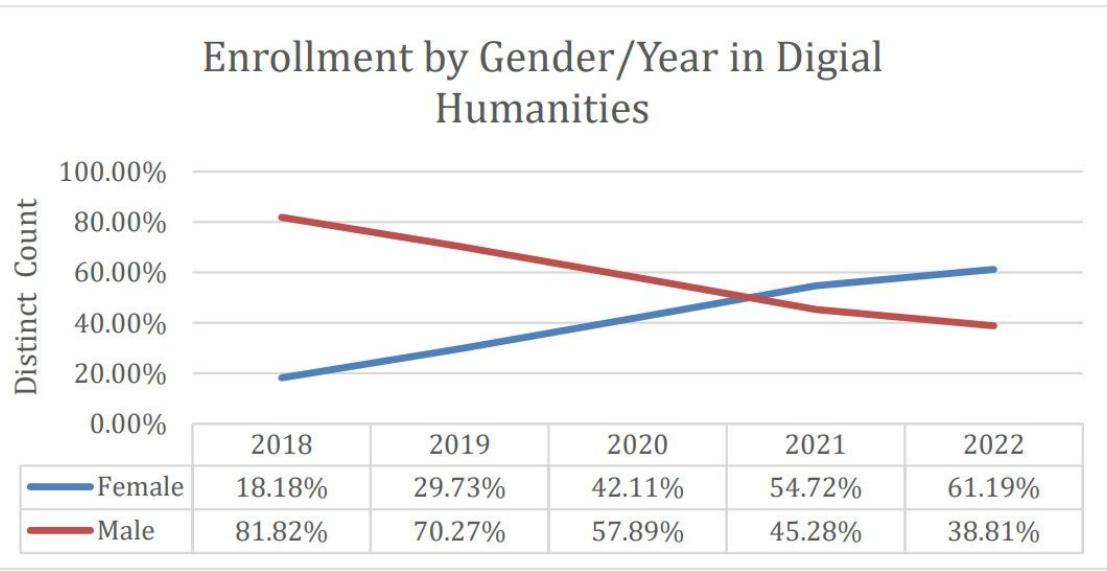
Human-First AI Curriculum



Digital humanities at Kenyon encompasses a ***dual lens***. We empower the next generation of thinkers with the ***conceptual framework*** underlying our Age of Information — from dataism and algorithmic thinking to synthetic biology and artificial intelligence. Our unique approach to ***computational thinking*** positions students to engage with the many ***practical, theoretical and ethical issues*** surrounding technological innovation and social change.

Digital humanities also explores the most recent computational approaches as a way ***to augment — rather than replace — more traditional humanist inquiry***.

Students imagine and pursue new avenues of research by identifying unexplored datasets of text, image and sound, while embracing new computational frameworks that are increasingly powerful and easy to use. We focus on finding ***interdisciplinary*** solutions to today's challenges.



- Non-STEM ~90%
- Women 61%
- Latine 11%
- African-American 13%

We serve students who may otherwise feel alienated by traditional CS or AI programs

Programming Humanity		2018	2019	2020	2021	2022
Non-Stem Major		93%	80%	95%	91%	91%
Stem Major		7%	20%	5%	9%	9%

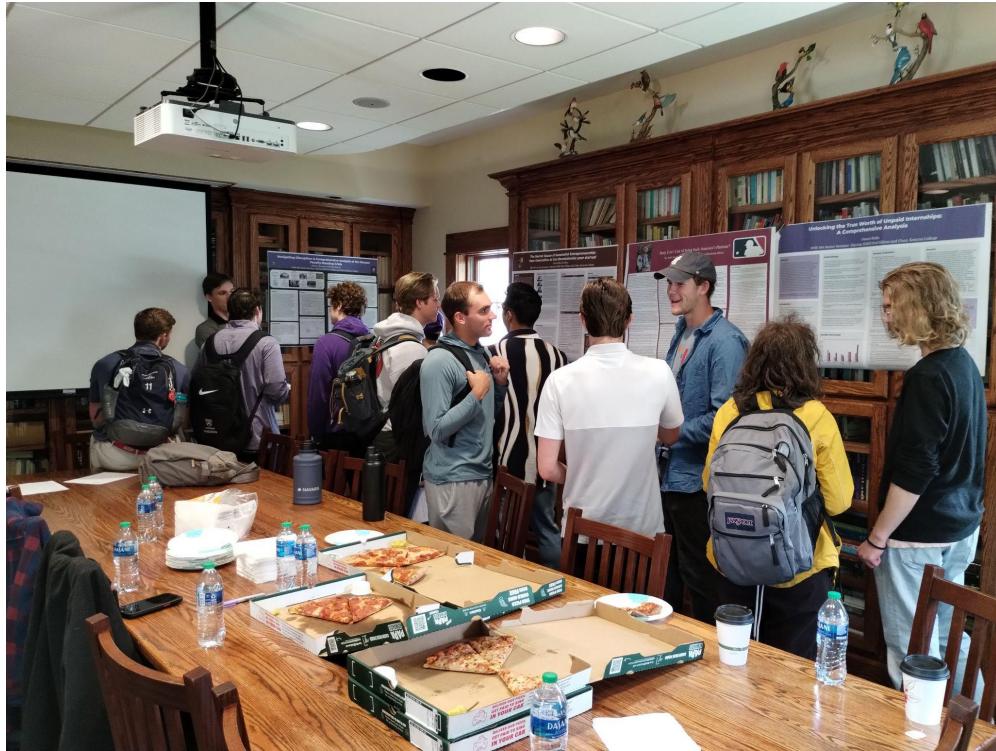
Goals & Outcomes

- Intellectual Framework
 - Key Abstractions
 - Many Model Thinkers
 - Think Critically & Thrive
- Synthesize/Applied Humanities
 - Find Stories in Data & Visualize
 - Testable Theories
 - Operationalize & Effect Change
- Grad School / Opportunities

Research Methodology



Poster Sessions



AI Research Paper

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Ilia Polosukhin* ‡
ilia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

1 Introduction

Sections	What to do?
Abstract	Write 150 words on the purpose of the study, research questions or suggestions, and main findings with conclusions.
Outline	Organize the map of the study.
Introduction	Provide the main information on the problem statement, the indication of methodology, important findings, and principal conclusion.
Literature Review	Analyze and incorporate scholarly sources on past studies.
Methodology or Materials and Methods	Explain the design of the research with techniques that are used for gathering information and other aspects related to the experiment.
Results	Present and illustrate the obtained findings.
Discussion	Review the information in the introduction part, evaluate their gained results, or compare it with past studies.
Recommendations	Propose potential solutions or new ideas based on the obtained results.
Limitations	Consider the weaknesses of the research and results to get new directions.
Conclusion	Provide final thoughts and the summary of the whole work.
Acknowledgments or Appendix	Include additional information on the research paper if it is necessary.
References	Provide and cite all used sources in the study.

AI Research Poster



Overcoming the Spectral-bias of Neural Value approximation

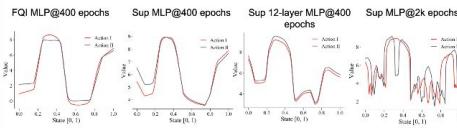
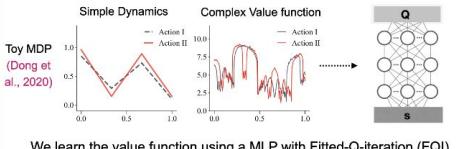


Ge Yang*, Anurag Ajay*, Pulkit Agrawal

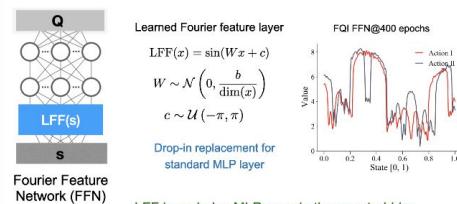
Improbable AI Lab, NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI), MIT CSAIL (* denotes equal contribution)

Spectral-bias in MLPs slows convergence in Neural Value approximation and causes instability in off-policy RL

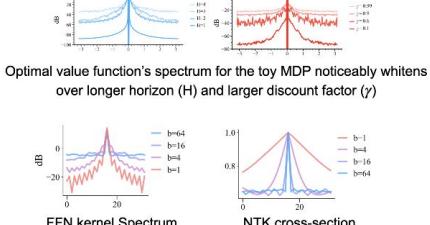
A Motivating Example



MLPs have **spectral-bias** towards low-frequency function which slows down value function convergence !!



Spectral-bias in Value estimation

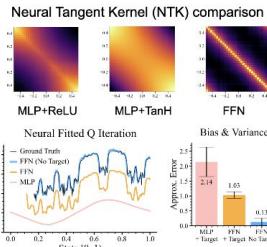


FFN kernel spectrum has larger eigenvalues on the high-frequency harmonics, thus improving convergence rate of the network

Analyzing off-policy stability via NTK

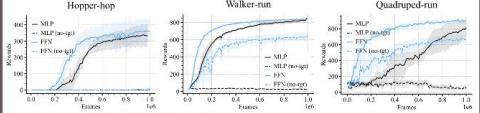
Cross-talks between the gradients of near-by states, as measured by NTK, leads to off-policy divergence (Achiam et al., 2019)

The NTK of FFN is **localized** with small off-diagonal elements showing less cross-talk.

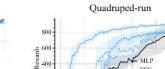


The improved learning stability allows us to **remove the target network** on a number of domains while retaining substantial performance.

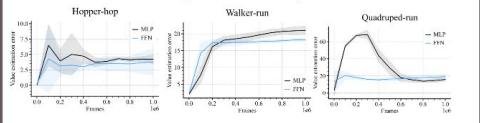
Scaling up to Deepmind Control Suite



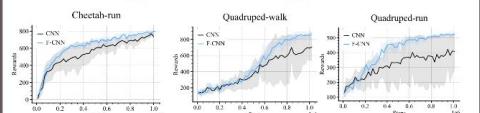
FFN leads to **faster convergence** than SAC and improves stability thereby retaining substantial performance on some tasks despite removing target-network



FFN improves compute efficiency and matches MLP's performance with only a fraction of compute

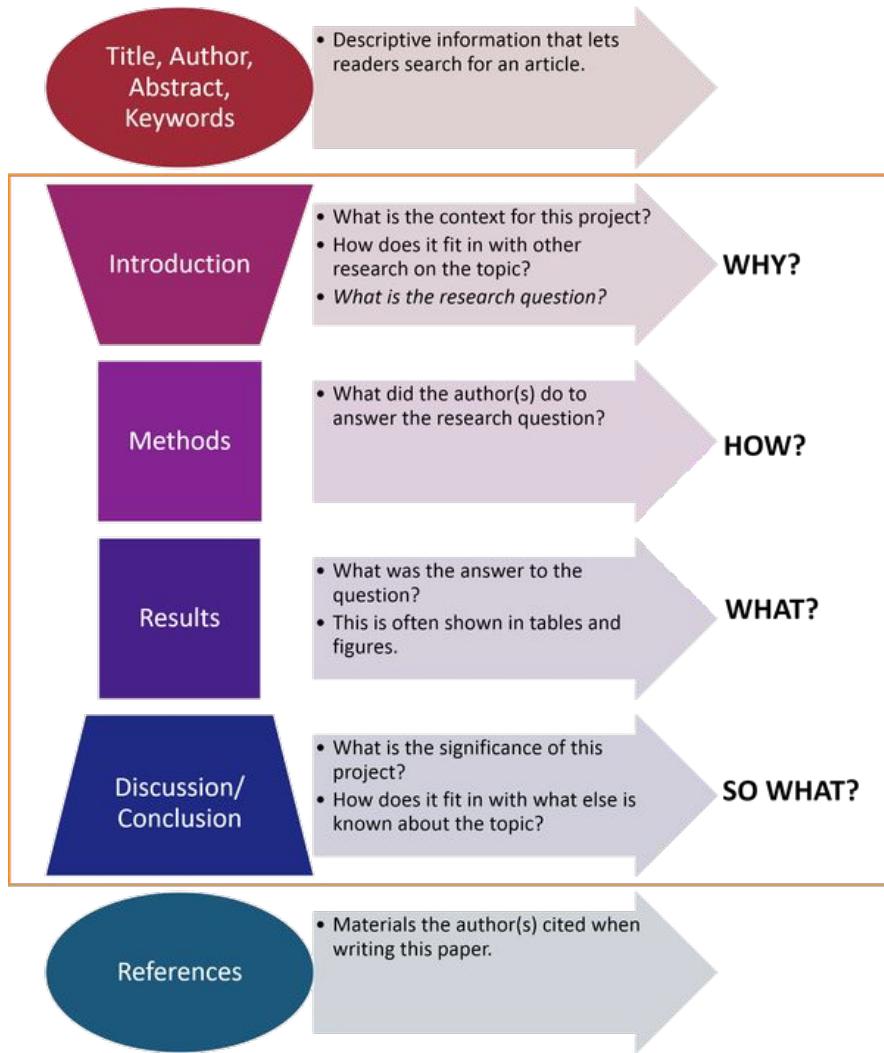


FFN consistently produces smaller value approximation error than MLP



FFN can be extended to image domains in form of Fourier CNN (F-CNN)

Distilling Papers To Posters

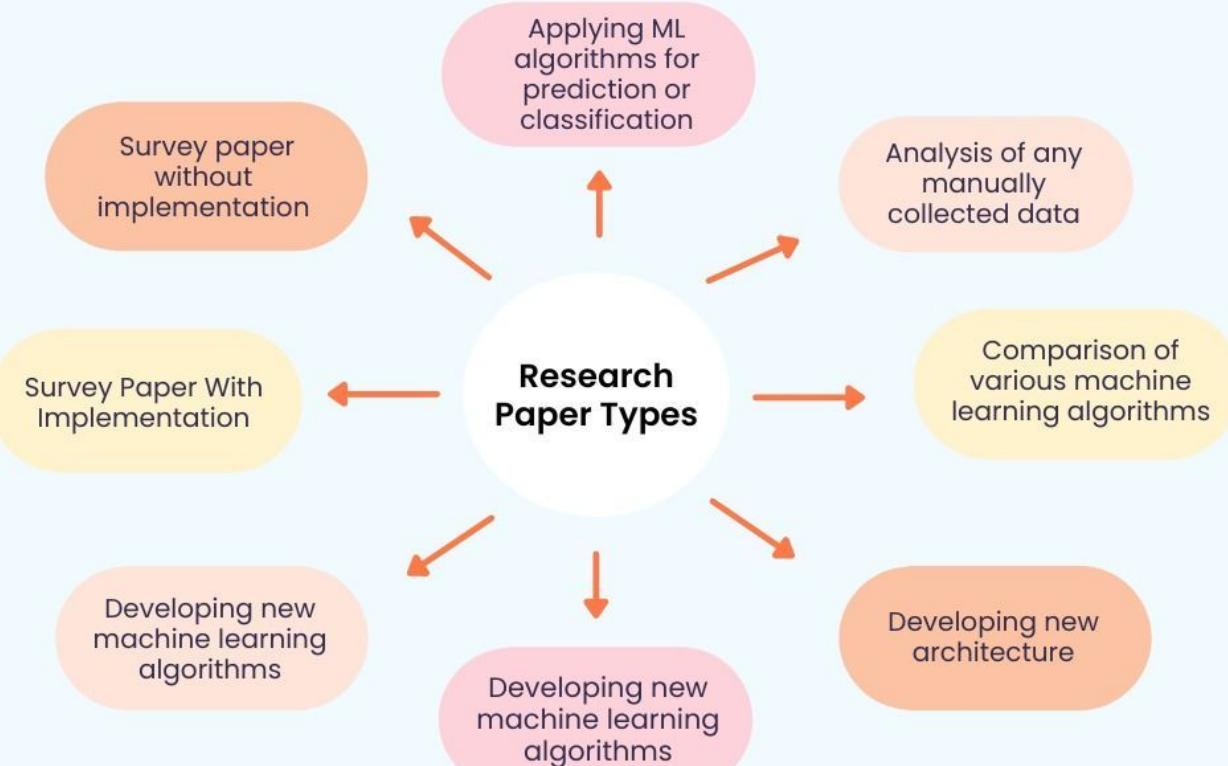


Process to Write A Good ML Research Paper



A New Writing Process

Topics & Genres



Information Architecture & Editing



Poster Title

Names of Scientists, Department, Institution, Address, Email

Abstract



Materials & Methods



Introduction



Figure 1

Results



Results



Figure 4



Discussion



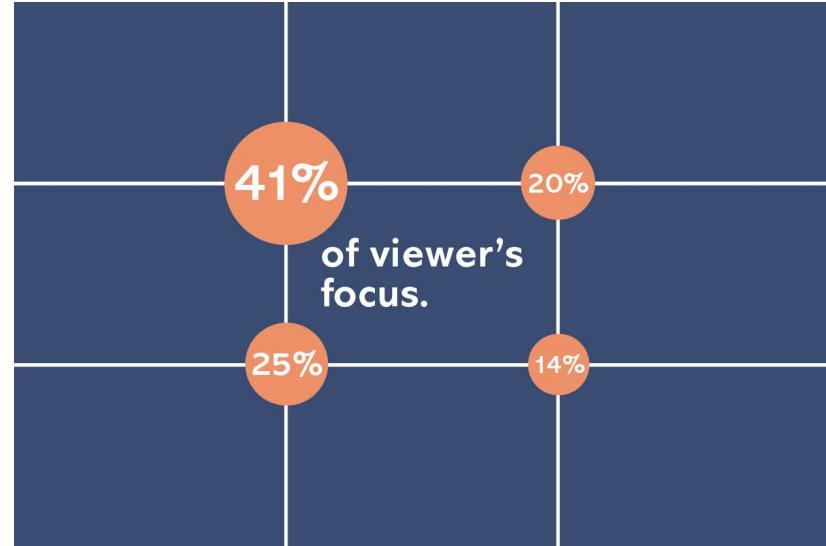
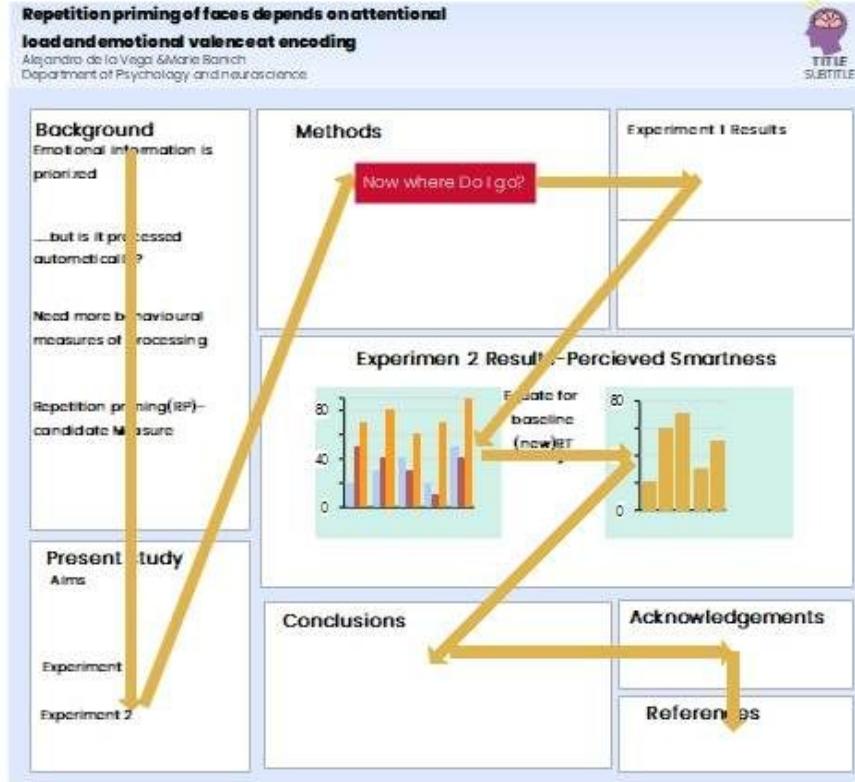
Figure 5

References



Acknowledgements

Psychology & Information Architecture



Graphic Design Principles

CONTRAST



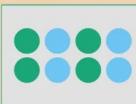
BALANCE



EMPHASIS



REPETITION



PROPORTION



HIERARCHY



RHYTHM



PATTERN



WHITE SPACE



MOVEMENT



VARIETY



UNITY



Key parts of a scientific poster



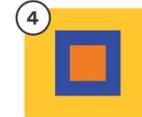
Content: Only include information that supports the take-home message of your research.



Layout: Poster sections should have a logical visual flow, with ample white space between elements.



Typography: Choose large font sizes so that the text of the poster is easy to read.



Colors: Help the viewer make contrasts by choosing colors carefully.

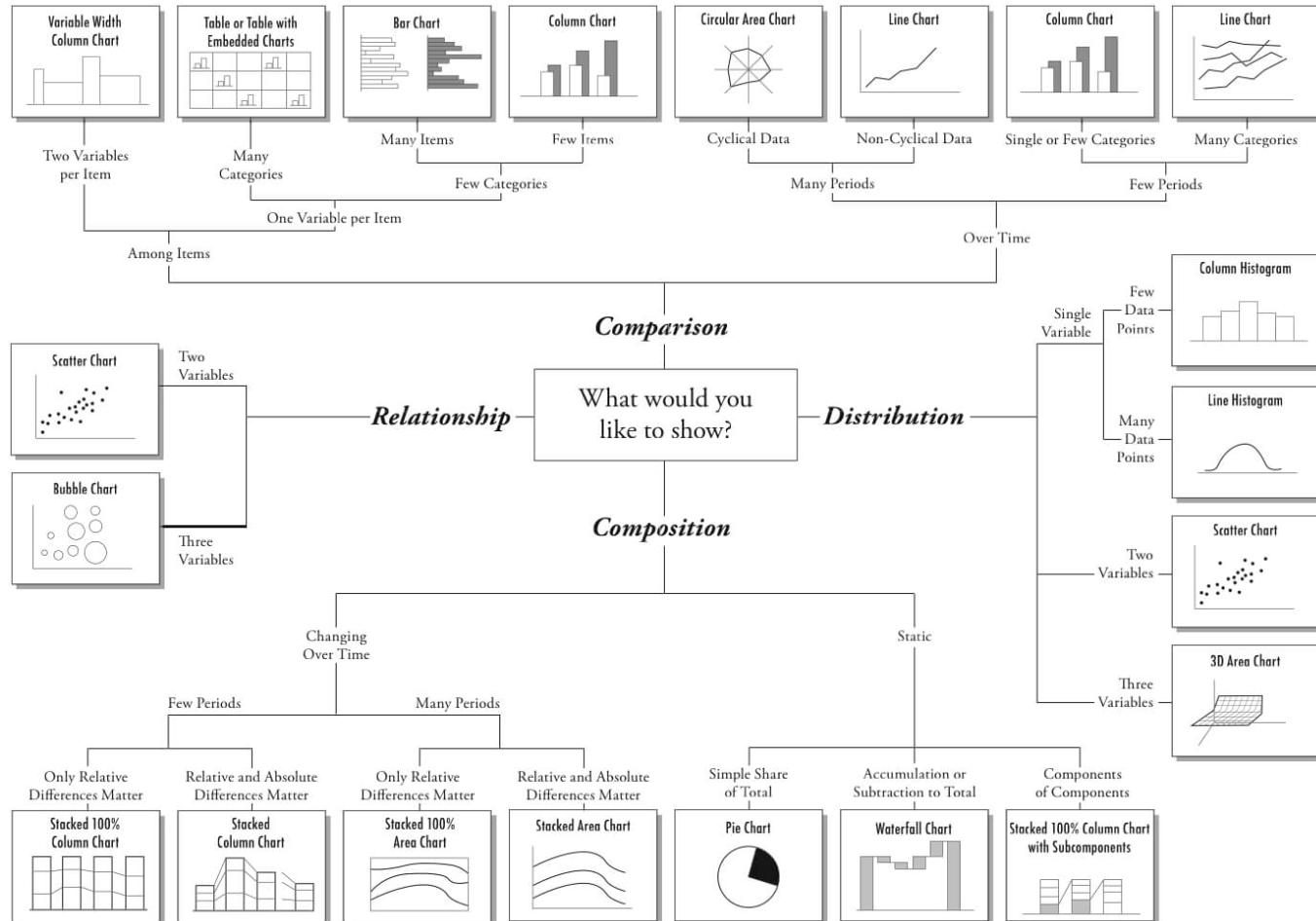


Images and Illustrations: Use good quality images with high resolution.



Design: Choose one design tool to avoid formatting errors.

DataViz Best Practices



World's Biggest Data Breaches & Hacks

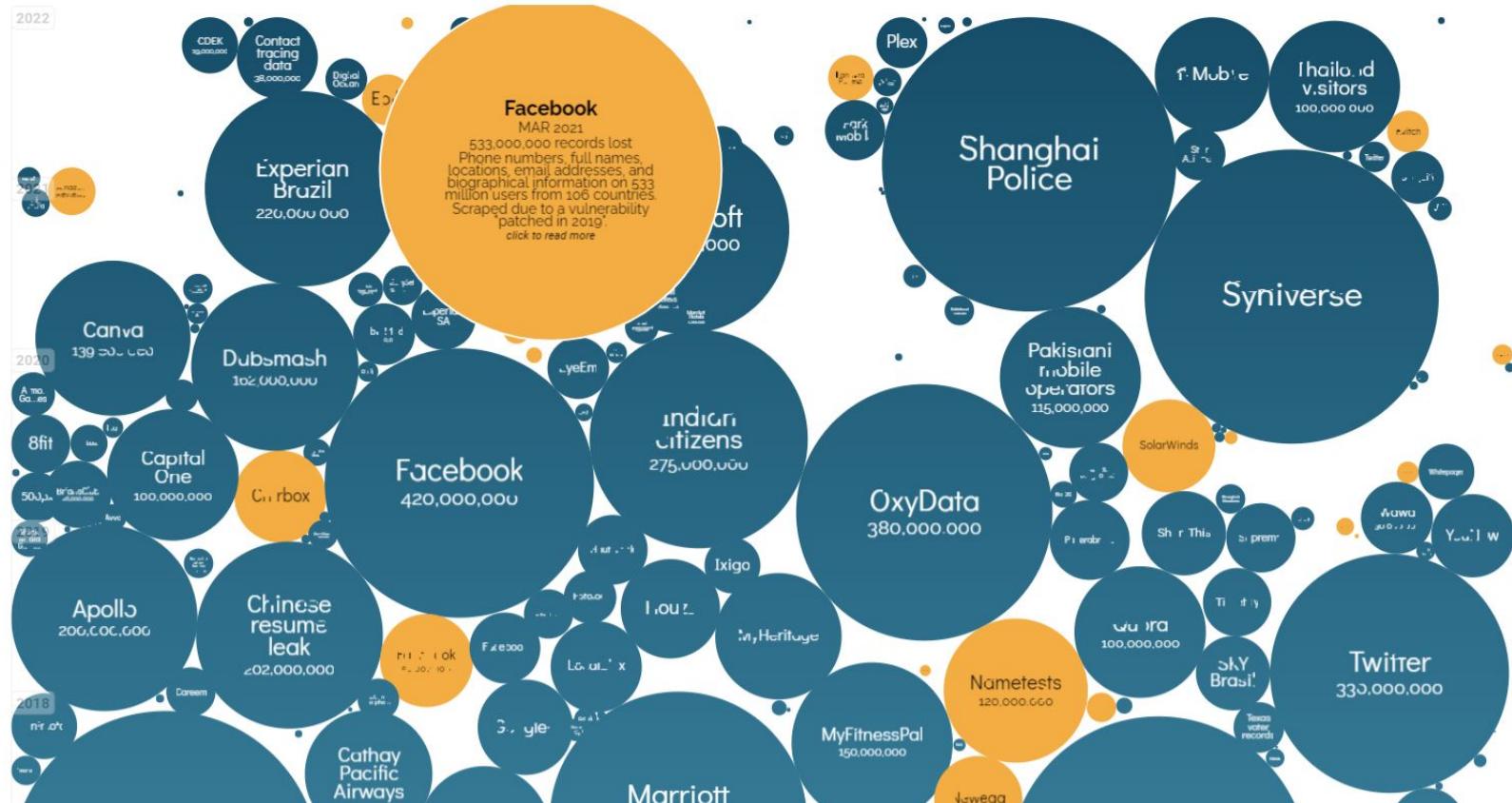
Selected events over 30,000 records

UPDATED: Sep 2022



size: records lost **filter**

search



Infographic Makeovers

- ▲ LAYOUT
- COLOURS
- WHITESPACE
- TYPOGRAPHY
- INFORMATION HIERARCHY
- DATA ANALYSIS
- DATA GROUPING
- FILTERS
- HOVERS
- ICONS

BEFOREAFTER

SALES

Salesman	Revenue (\$)	Change (%)
SALESMAN A	\$5,000	+3.06%
SALESMAN B	\$8,000	+2.12%
SALESMAN C	\$10,000	-2.67%
SALESMAN D	\$12,000	+3.21%
SALESMAN E	\$5,000	+3.06%

PAID MARKETING EFFECTIVE

	SPEND	VISITORS/\$
TOTAL	\$4,987	8.76
LinkedIn	\$2,567	17.76 +2.4
Youtube	\$376	9.36 -9.7
Google ads	\$1,487	8.45 -2.4
Facebook	\$376	9.36 -3.5
Twitter	\$977	7.56 +6.5

FORM FILLOUTS BY PRODUCT

Category	Product	Percentage
By Email	IV	22%
	BD	16%
	MA	18%
	AV	7%
	DP	11%
By Call	IV	22%
	BD	16%
	MA	18%
	WS	2%
	AV	7%
By Web Form	IV	22%
	BD	16%
	MA	18%
	WS	2%
	AV	7%

REVENUE IN PIPELINE

Category	Value
IN PIPELINE	\$25,000
PROJECTS SITTING	8
AVERAGE DAYS	6.3
SCOPE INCREASE	\$4,000
SECOND INVOICE	\$6,000

SALES

Salesman	Revenue (\$)
SALESMAN A	\$10,000
SALESMAN B	\$30,000
SALESMAN C	\$5,000

PAID MARKETING EFFECTIVE

Platform	Spent	Visitors/\$
LinkedIn	\$2,567	17.76
YouTube	\$376	9.36
Google ads	\$1,487	8.45
Facebook	\$376	9.36
Twitter	\$977	7.56

FORM FILLOUTS BY PRODUCT

Category	Product	Value
Email	IV	22%
	BD	16%
	MA	18%
	AV	7%
	DP	11%
	WS	2%
	MS	17%
Call	IV	22%
	BD	16%
	MA	18%
	WS	2%
	AV	7%
	IV	22%
	MS	17%
Web Form	IV	22%
	BD	16%
	MA	18%
	WS	2%
	AV	7%
	IV	22%
	MS	17%

REVENUE IN PIPELINE

Category	Value
IN PIPELINE	\$25,000
PROJECTS SITTING	8
AVE. DAYS	6.3
SCOPE INCREASE	\$4,000
SECOND INVOICE	\$6,000

<https://www.datalabsagency.com/2023/07/06/brand-guidelines-for-data/?v=7516fd43adaa>

(i)

Human-Centered AI

VS

STEM

1. Title and Introduction:

- **Interdisciplinary Title:** Unlike typical STEM titles, which are often highly technical, titles for human-centered AI research should *balance technical accuracy with an understanding of social or cultural context*. For instance, "Analyzing Social Media Sentiment Pre-Economic Crisis in Sri Lanka: A Human-Centered AI Approach" bridges technical and contextual aspects.
- **Broader Research Question:** These questions should encompass both the *technical and humanistic elements*. For instance, "How can AI-driven sentiment analysis provide insights into socio-economic conditions?" reflects both AI and societal concerns.

2. Overall Writing Style:

- **Balanced Technical and Humanistic Language:** Human-centered AI research demands a *blend of technical language and narrative styles* that *reflect human experiences and societal impacts*, more so than typical STEM research which might focus solely on technical aspects.
- **Cultural and Ethical Considerations:** *Explicitly discuss the cultural, ethical, and societal implications* of the research, a facet often less prominent in standard STEM posters.

(ii) Human-Centered AI VS STEM

3. **Organization and Layout:**
 - **Interdisciplinary Sections:** In addition to the standard sections (introduction, methods, results, conclusion), include sections on *ethical considerations, societal impact, or cultural relevance*.
 - **User-Centric Design:** Design the poster with a focus on *accessibility and inclusivity*, reflecting the *human-centered nature* of the research.

4. **Data Visualizations and Graphics:**
 - **Inclusive and Diverse Visuals:** Use visuals that are not only *informative* but also *culturally sensitive and inclusive*, considering the diverse audiences that human-centered AI research addresses.
 - **Representative Data Presentation:** Ensure that data visualization reflects the *diversity and complexity of human experiences*, which is a unique aspect compared to more straightforward STEM data presentation.

(iii) Human-Centered AI

VS

STEM

5. Engagement and Interaction:

- **Broader Engagement Questions:** Pose questions that *encourage viewers to think* about *both the technical and humanistic aspects* of the research.
- **Interactive Elements:** Consider including *interactive elements like QR codes* linking to simulations or models, which can provide a deeper understanding of both the AI technology and its human-centered applications.

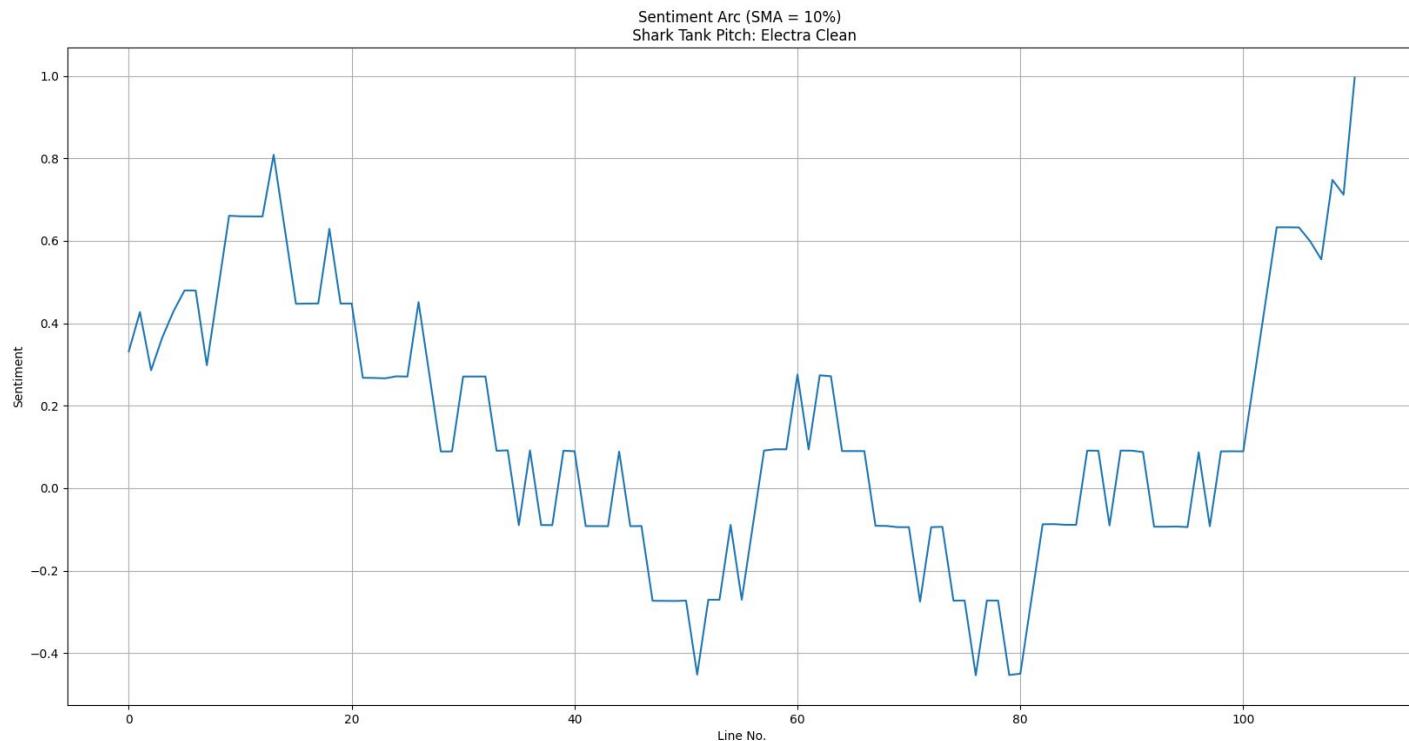
6. Presentation and Delivery:

- **Interdisciplinary Summary Speech:** Prepare a *summary* speech that not only explains the technical aspects but also addresses the *human, cultural, and ethical* dimensions of the research.
- **Empathetic Communication:** Emphasize a *communication style* that is not only *clear and concise* but also *empathetic and inclusive*, reflecting the human-centric nature of the research.

Examples

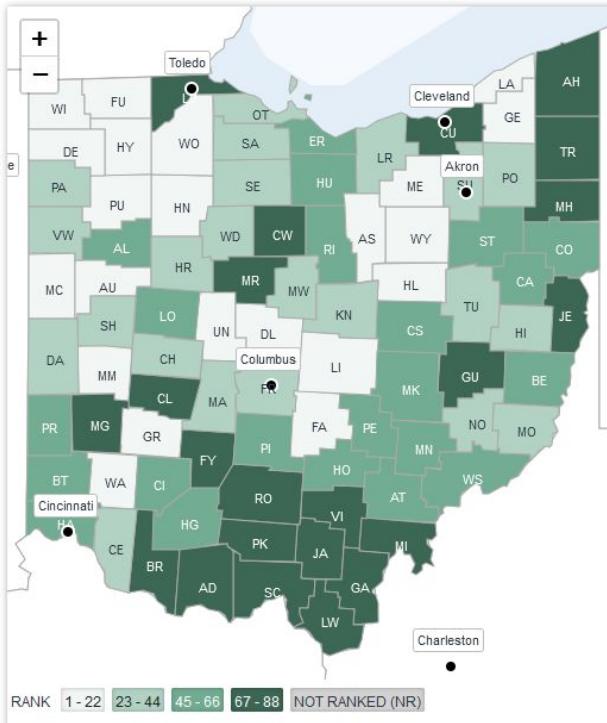
- Papers
 - Websites / Blogs
 - Plays
 - Films
 - Art Installations
 - Gallery Designs
 - Community Engagements
 - Independent Research
 - Internships / Jobs
 - Startups
 - Guest Speakers
 - Campus Roundtables
-

Shark Tank Pitch: Alexander Gow

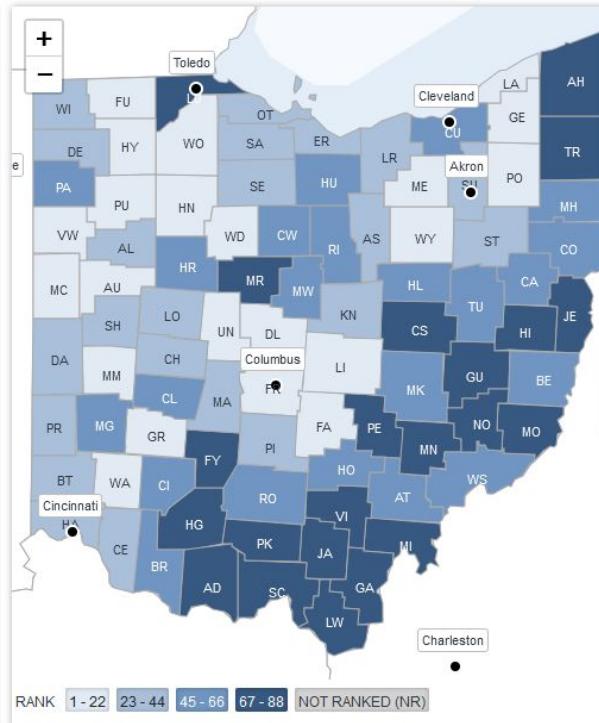


Ohio Opioid Crisis: Zach Elsawy

Overall Rankings in Health Outcomes [i](#)

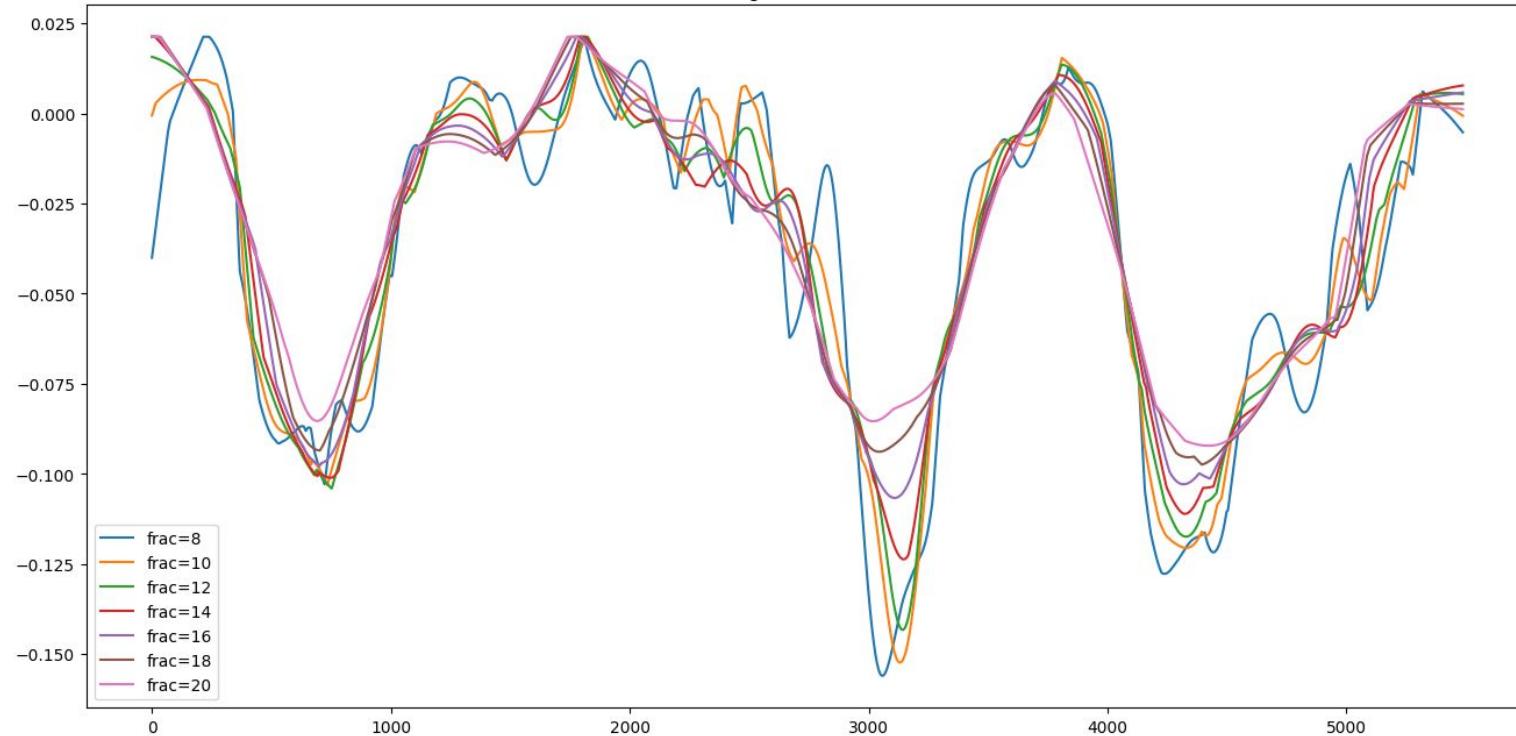


Overall Rankings in Health Factors [i](#)



Harry Potter Screen Adaptations: Erin Shaeen

The Sorcerers Stone by: J.K. Rowling
LOWESS Smoothing Grid Search (frac=0.08 to 0.2)



Generative AI Art & Design: Jill Noorily



Sponsored by Integrated Program in Humane Studies, Art History, and Studio Art

HOW AI IS CHANGING ART, DESIGN, AND MUSEUM CURATION

October 26, 2021 11:10AM-12:00PM

Community Foundation Theater

Yakera: Raul Romero & Tomas Munoz Reyes

yakera ABOUT US SUPPORT US CAMPAIGNS FAQ LOG IN

Changing how humanitarian aid is delivered around the world, one story at a time.

[Donate Now!](#)

Explore the stories and Donate to a Campaign Share the campaign See your Impact Join the Yakera Community

Doubles and Reflections: Sentiment Analysis and Vladimir Nabokov's *Pale Fire*

Catherine Perloff, Jon Chun, Katherine Elkins

Comparative World Literature Senior Seminar at Kenyon College

Introduction

In his novel, *Pale Fire*, Vladimir Nabokov uses a non-linear narrative composed in three parts: Foreword, Poem, and Commentary, apparently written by two distinct authors, John Shade and Charles Kinbote. Storylines appear to be separate at surface level yet work in interconnected ways at a deeper level. The novel deals with life after death, the story arc of a disguised, exiled king, and provides meta commentary on the art of commentary.

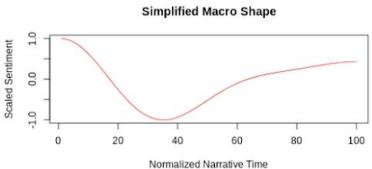
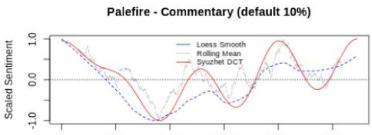
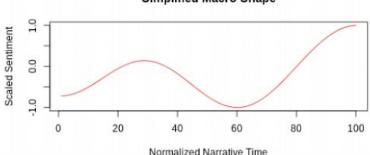
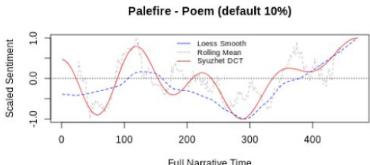
The text was tricky to analyze without sentiment analysis principally because of its seemingly disjointed structure. Thus, there's a bit of irony in that we found a repeated structure to the text despite the preconceived notion that the text is "plotless" overall.

Both the simplified macro structure to the Poem and to the Foreword and Commentary show the "Man in a Hole" narrative model, which illustrates a mirroring or doubling in structural self. This is strengthened by the larger structures of Foreword and Commentary and Poem being mirrors of one another as well; when compared, the two are reflections of one another in terms of structure. This was not apparent prior to sentiment analysis and is important, as it mirrors the reflective nature of the text.

Methodology

After cleaning the text, we used the Syuzhet and Sentiment.R packages in R to plot the emotional valence of *Pale Fire* over narrative time. We plotted the narrative arc using a DCT with a low pass of 10, a Loess smoothing model, and a rolling mean of 0.1.

Syuzhet and Sentiment.R sentiment analysis were used to map the narrative arc of *Pale Fire*. We first ran the full text through Syuzhet and Sentiment.R, then isolated the poem from the commentary and foreword and ran those sections through Sentiment.R as well. We did this to see if the narrative arc matched an already established narrative arc and to explore high and low emotional valences. We then merged more traditional literary analysis with these new tools; we looked at the emotional peaks and valleys and used literary analysis to uncover possible connections between the Foreword, Commentary, and Poem.



Nabokov Scholarship

We first looked to relevant secondary reading and scholarly articles on Nabokov and *Pale Fire*. Much of the literary scholarship on *Pale Fire* concerns the novel's authorship, which was not relevant to this project, but some of the articles proved vital resources.

In *Nabokov's Pale Fire: The Magic of Artistic Discovery*, author Brian Boyd establishes his belief that there are connections to be made between the commentary and foreword; that winding paths in the narrative that ask the reader to jump back and forth can, while appearing to be nothing, actually lead to discovery and understanding. Readers should, therefore, trust in their curiosity to "follow the trail." As for the poem, Boyd writes some commentary and analysis, though focuses on authorship in the novel. In "Shade and Shape in *Pale Fire*," Boyd argues that the echoes and patterns of *Poem* do indeed interlock into a Nabokovian "key" that will, when found and used, unlock the answers to the novel's riddles. He again, however, focuses on authorship instead of delving deeper into these "keys."

In "Bolt from the Blue," Mary McCarthy attempts to untangle the tricks and riddles within *Pale Fire* by picking apart the text and finding keys to understanding connections. In this way, she goes beyond Boyd in her analysis of the poem, her identification of such Nabokovian "keys," and analysis of the connections between the three sections to the novel. She argues that there are multiple levels to the novel, "planes in fictive space," and likens the novel to a chess game, a "mirror-game." She believes that the "real," "real" story underneath can be accessed by the attentive reader and that the novel deals with echoes, with mirrors, and with doubles.

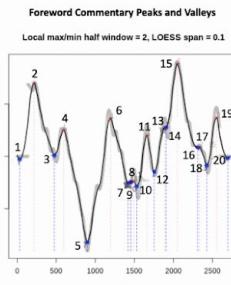
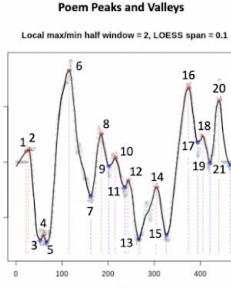
McCarthy's impressive attempt comes closest to an understanding of how the novel works at a deeper level instead of focusing on scholarship. Her article, however, came out in June 1962, far before Boyd's articles. Thus, this kind of analysis of *Pale Fire* has remained unanswered since; there is much that remains to be answered.

Results

There can be a lot of interpretation with the results. The computer is locating moments of emotional inflection and it's our job to ascertain if they're credible by looking at close reading, considering noise, and then considering the plot points as a whole with an underlying structure. And, the emotional valence was determined by humans, as the dictionary being used was created by humans.

Not all the crux points are talked about here, as some are too noisy for clear analysis and others are on or around the neutral line.

The crux points of the commentary/foreword and poem overlap in a few places, whether it be with the same scene shared in both graphs or high and low emotional valence points for a single event.



The crux points 7, 8, 9, and 10 of the Commentary occur very close together and illustrate that the scene has both a high and a low emotional valence. This is when Hazel goes to the haunted barn and sees the spectral "pale light" that she can't quite reach.

There is also doubling occurring within the connection between the Commentary and Poem. The crux point 5 of the Poem reflects the scene in which Charles, the King of Zembia, is looking into a lake and seeing his doubled reflection. Yet, this reflection is in a different location and turned a different way. As such, this isn't his genuine reflection and illustrates a sense of separation from self.

"In its limpid tintarn he saw his scarlet reflection but, oddly enough... this reflection was not as his feet but much further... his red-sweatered, red-capped double-ganger turned and vanished, whereas he, the observer, remained immobile."

This point is linked to crux points 3, 4, 5, and 15 of the Poem, as both occur as the lowest valleys, when Shade experiences his seizures:

"And then black night / That blackness was sublime. / I felt distributed through space and time."
"And then it happened—the attack, the trance, or one of my old fits... / I can't tell you how I knew—but I did know that I had crossed / The border"

The seizures cause Shade to experience a separation of self as well and link him and the King together / link their experiences together.

These points of doubling are further linked to crux point 19 of the Commentary, when Shade experiences a doubling of self in a sleep state – him seeing himself outside of his body, seeing himself waken:

"Once overheard / Myself awakening while half of me / Still slept in bed. I tore my spirit free. / And caught up with myself—upon the lawn / Where clover leaves / Cupped the tops of dawn. / And where Shade stood... / And then I realized that this half too / Was fast asleep"

There's a sense of disengagement with these linked scenes.

Conclusion

This proves to be a good proof of concept for this kind of approach; it makes the case for this kind of analysis. One would normally not expect *Pale Fire* to exhibit a narrative structure, yet it does. So, then the question remains: what does this mean? Is *Pale Fire* not hypertext, but reflections of the same story? This would be a very different notion of hypertext, as the paths are not totally different, but mirrored reflections of one another, as seen in the Results section.

This begs a larger meta question: is there a shared experience in reading? Or are our reading processes totally different from other's reading experiences? Kinbote's reading experience is different from ours, yet his experience mirrors Shade's story through the poem.

As Mary McCarthy stated, *Pale Fire* exists in a mirror world, a world of reflections. Thus, on a meta level, we all read our own story, but our stories are connected. Our stories are all "pale fires" of one another; they're not perfect, nor identical — they're just reflections.

Acknowledgements

Many thanks to Professor Chun for the coding and to Professor Elkins for supporting the literary analysis and for leading this CWL senior seminar.

References

- Boyd, Brian. *Nabokov's Pale Fire: The Magic of Artistic Discovery*. Core Textbook ed. Princeton University Press, 2001. Project MUSE
- Boyd, Brian. "Shade and Shape in *Pale Fire*." *Nabokov Studies* 4, 1997, pp. 173-224.

Christopher Innes, "The Two Translations of the Author's Note in *Pale Fire*," *Nabokov Studies* 10, 2002, pp. 1-16.

Lost in Translation: Using Sentiment Analysis to Analyze Translations of Homer's *Odyssey*

Erin Shaheen

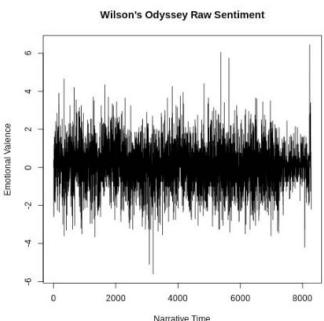
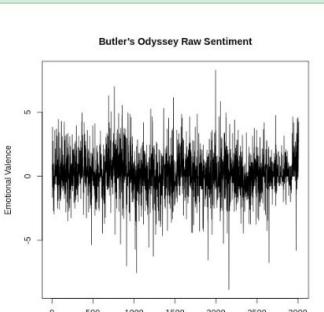
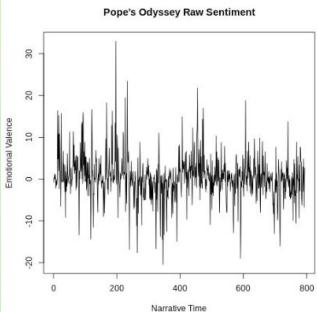
Introduction

For my final project in AI for the Humanities, I knew that I wanted to examine the creative role of translators in classic literature. I ran three different translations of Homer's *The Odyssey* through Syuzhet.R, a sentiment analysis tool that has grown increasingly popular in the Digital Humanities field. Syuzhet tracks the emotional arc (also called "emotional valence") of a text by giving each word a different score. For example, on a scale of -1 to 1, "terrible" might score -1 and "amazing" might score 1. A more neutral word like "okay" would score closer to 0.

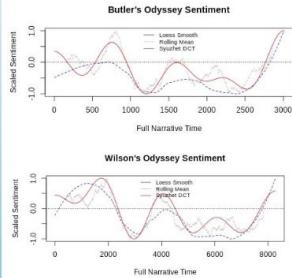
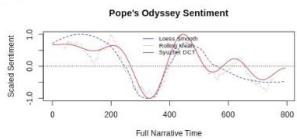
I chose translations by Alexander Pope, Samuel Butler, and Emily Wilson, published in 1725, 1900, and 2017 respectively. I wanted to explore how the translators' working in different time periods and having different life experiences might affect their interpretations of the text.

Methodology

The first thing I did was run the three translations through Syuzhet and look at the raw sentiment analysis. As you can see in the following graphs, Pope's translation is by far the least noisy, followed by Butler's, and then by Wilson's.



Though these graphs are certainly useful, the amount of noise can make it difficult to get accurate results. Thankfully, Syuzhet offers an option that simplifies the graphs, and allows for easier reading.



Once I had made the simplified graphs, I was able to pinpoint what are known as "cruxes," the specific high and low points of emotional valence. I focused on the Syuzhet DCT line, marked in red, for all three graphs, both for clarity's sake and for the fact that it appears to be closest to an average.

Interestingly enough, all three of the texts had approximately the same lowest point—in Book 11, when Odysseus sees the ghosts of dead comrades and his recently deceased mother in the House of Hades.

I found slightly more variation in the highest points. While both Butler and Wilson's translations mark the highest point as during Odysseus' time spent visiting the home of King Alcinous, Pope marks it as when Odysseus visits the swineherd Eumeaus.

Closer inspection of the above graphs reveals that all three of them dip slightly towards the beginning of the text, but only Butler's translation drops before the neutral line. When I pinpointed the cruxes, I found that while Butler and Wilson mark this low point in Book 4, where Telemachus visits Menelaus, Pope, who has the shallowest curve, marks it in Book 6, where Odysseus leaves Calypso.

Results

My initial hypothesis that the sentiment analysis would shift based on translator was correct—but only to a certain extent. While you can see from the graphs that different translations had different highs and lows, the overall arcs remained relatively similar. All three of the translations had roughly the same point of lowest valence, and while the points of highest valence had more variation, they all involved the theme of hospitality. Pope's translation had the most variations, with the most positive sentiment appearing at a different crux point (the visit to Eumeaus) and with variations at the beginning of the text.

Conclusion

Syuzhet proved to be extremely useful in tracing the differences between translators. Alexander Pope's work has different crux points than Butler and Wilson's, which correlates with how that his graph contains significantly less noise than the others. It's worth noting that his work was originally published in the early sixteenth century, while Butler worked in the eighteenth and nineteenth, and Wilson in the twenty-first.

Samuel Butler's translation has the most negative valence, being the only one to dip below average at the start of the text, and remaining there for the longest amount of narrative time.

I initially selected Emily Wilson's translation because of how recent it was, and the reviews I read online lead me to believe that it drastically differed from previous translations. In terms of Sentiment Analysis, Wilson's translation actually provided the closest thing I had to a middle ground between Butler and Pope. Of course, I only looked at the texts through the perspective of Sentiment Analysis, and I only chose three translations out of a large body of canonical work.

If I had the chance to do this project again, with the knowledge I have now, and significantly more time on my hands, I would do a few things differently. I would start by adding in other translations, to get a larger basis for comparison. Then I'd supplement my work with a variety of other tools, like Voyant, which includes features that note the frequency of words and their connections to each other, and could potentially shed light on the overall scores. I'd also go into a more in depth analysis of the texts themselves, particularly at the crux points I mentioned above, offering a human perspective to compare to the digital ones.

Acknowledgements

I would like to thank Professor Chun for his help with coding, Professor Elkins for her recommendations and her help with Syuzhet analysis, and Miles Shebar for helping me with the technical side of working with Wilson's translation.

References

- *The Odyssey* by Homer, translated by Samuel Butler
- *The Odyssey* by Homer, translated by Alexander Pope
- *The Odyssey* by Homer, translated by Emily Wilson
- Pope and Butler's translations came from Project Gutenberg
- <https://github.com/shaheene/OdysseyTrans>

Meditations in an Emergency: a Regional Analysis of the Relationship Between Union Membership and Minimum Wage in an Age of Labor Decline

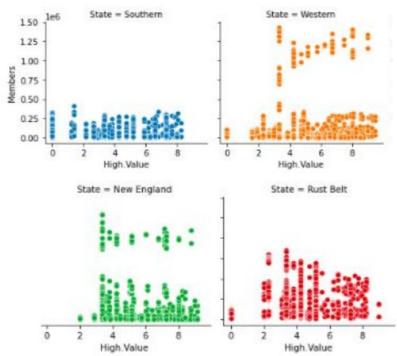
Dante Kanter

Professors Jon Chun and Katherine Elkins

Introduction

The general consensus amongst labor historians is that the election of Ronald Reagan led to a marked decline in Union Membership that, with a few exceptions, has continued steadily up until the present day. Beginning with the Air Traffic controller strike in 1981, where the Reagan administration fired and permanently replaced striking union controllers, and continuing with Reagan members of the National Right to Work Committee to the National Labor Relations Board, a legal entity which is hostile to union organizing created in the U.S. In my research my hope was to take popular and agreed upon census data and, by doing specific grouping and analysis of membership data by region, highlight particularities that offer a more complex view of the decline of American Labor.

In addition, I attempted a comparative analysis between changes in union membership and changes in the mean minimum wage throughout these regions, in order to highlight what the private sector had to gain by disempowering union networks.



Materials

For this project, I used Python as my base language, with Seaborn and Pandas as my two libraries. Seaborn is a language used to visualize datasets, and Pandas is a language used to manipulate datasets.

I used two datasets for this project— one a survey of the minimum wage, by state, from 1968 to 2017. Several wages were included each year—the low estimate of minimum wage in that state and the high estimate of minimum wage. The dataset also included wages adjusted to match a

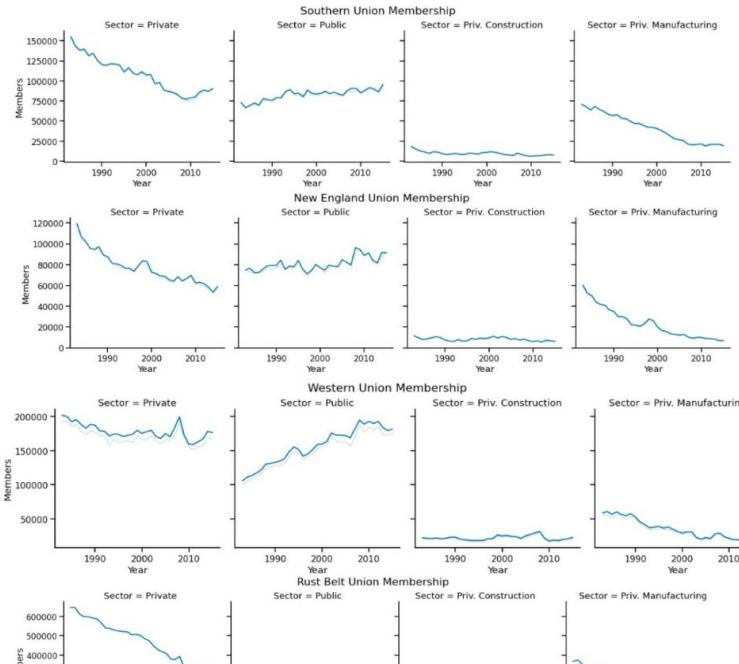
Methodology

I began by separating out the dataset by region, quartering the country into Southern, Western, New England, and Rust Belt quarters. I plotted the sum of memberships in each collection of states along a line graph, separating the “col” segment by the type of industry in which the union operated, testing out the differences between industries in the public and private sectors.

I performed much the same method on the dataset pertaining to minimum wage. I then aggregated the two datasets and plotted out a scatter graph meant to show the relationship between union membership and minimum wage.

Results

All regions of the country are marked by a steep decline in the minimum wage across the country, although recovery looks very different across regions, with the West having the steepest difference and the rustbelt looking to be on the precipice of a second decline. The south, with low union membership, experienced a step-ladder decline, while the Industrial Rust Belt took a nose-dive. The western states, especially California, were the few to spike after the 2008 recession. New England saw the benefits of the NAFTA and the Clinton administration. Public Union membership has increased everywhere except the Rust Belt. The greatest correlation between minimum wage and union membership appeared to be in the Western States and New England.



Conclusion

The scatter graph is telling, as those regions that include the western states with the highest union membership—California and New York— are “bastions” with high correlation between high minimum wage and high membership. The south, which does not have a particularly strong history of labor unions, remains in a zone of low wages and low union membership, while in the Rust Belt, people appear to be earning less, but existing on the same middle plane of membership and wages.

In the beginning of the Reagan administration, the South had an influx of new unions as established organizations, such as United Steelworkers, began to ciphon out of the region. This explains the halting decline in the early years, that later gives way to a steady decline.

The correlation between high union membership and high wages is difficult to miss, however, and the sharp drop in the minimum wage just as unions began to be disenfranchised speaks to how this process benefited private industries, who felt justified in paying their workers a lower wage without a union to protect them.

Citations

- Furukawa, H.S. and Western, B. (2002). Ronald Reagan and the Politics of Declining Union Organization. *British Journal of Industrial Relations*, 40, 385–401. <https://doi-org.libproxy.kenyon.edu/10.1111/j.1467-8543.00240>
- Bennett, Sam. "LABOR ORGANIZING IN THE SOUTH, 1970-1979: THE CASE OF THE INTERNATIONAL BROTHERHOOD OF TEAMSTERS." *Southeastern Geographer*, vol. 24, no. 1, 1984, pp. 1-13. JSTOR, www.jstor.org/stable/44370761. Accessed 17 Dec. 2020.
- Ruth Milkman, and Stephanie Luce. "Labor Unions and the Great Recession." *RSP: The Russell Sage Foundation Journal of the Social Sciences*, vol. 3, no. 3, 2017, pp. 145–165. JSTOR, www.jstor.org/stable/10.7758/rsp.2017.3.3.07. Accessed 17 Dec. 2020.



TikTok's Non-Inclusive Beauty Algorithm & Why We Should Care

Priya Melonio

IPHS 200 Programming Humanity



Introduction

This project will be a meta-analysis of how the popular social media app known as TikTok takes into account image recognition in their machine learning algorithms through the data it analyzes from its users. It will also see how it identifies and pushes the most beautiful to fame and virality. Though we don't have access to the actual TikTok algorithm, we are going to use a very similar dataset known as SCUT-FBP5500. We will see how it identifies specific western and eastern beauty standards that are only based on far too simple analysis of what is considered beautiful. We will also use a separate study through a scientific study, which analyzes men and women stimulus in response to beauty. We will lastly use an article, which explores the Chinese app called Alipay, and how it uses beauty filters that perpetuate patriarchal ideals over women. This dataset, study, and article will uncover how human nature and sociology can contribute to how algorithms are truly being fed our want to see beauty. We will also see how the beauty standard that is considered inherently bad is false, but that human society around the world needs new establishments of what true beauty is instead. Overall, the goal of this project is to understand these examples of beauty algorithms, how they work, the reason they are used in human society, and how we can improve or discourage use of them in our social media apps.

Background

The history of TikTok goes far back to the app's owner known as Bytedance, and its history of being originally two apps, Musical.ly and Douyin. Bytedance is a Beijing based internet and technology company, which has a long history of accusations of being a political and allegedly anti-western company. It also means that United States even open anti-trust investigation against the company and app. These accusations are for another report, but it can contribute to our understanding of how TikTok uses the data its users give to it. Musical.ly was a video-based app very similar to TikTok, which mainly focused on dancing videos. Douyin was an app based in China, which is the Chinese name for TikTok, that was known for being much more product based. This meant it uses visuals through uploaded videos, where users could buy products in those videos, or plan events to go to, and interact with other users through comments provided through the app. In short, the decline in Musical.ly and Douyin for a couple of years before their official merge in August of 2018 led Bytedance to become one of the most renowned companies, having its net worth sit around \$100 billion US dollars.

We will be mainly evaluating the two studies provided by Semantic Scholar. One will be with the SCUT-FBP5500 machine learning dataset provided by the South China University of Technology, and the other with a visual activity study provided by Dr. Daniel Gill et al. of UCR. The primary reason for this will reference is through a political and cultural lens from Alman Yuzhu Peng, who looks into the Chinese app Alipay. Peng explains how beauty rating technology that takes data from users is socially harmful. We will also see how this affects eastern culture surrounding beauty in the economic sense with hiring those based on looks, and how it has even caused a drastic rise in plastic surgery. We will be addressing the numerous apps of like TikTok relying on beauty-based machine learning algorithms, which are creating beauty standards that are similar to TikTok's operate. These issues include how the damaging use of algorithms using incomplete datasets fails to push minority people to fame. We will also see how it also pushes toxic standards of western and eastern beauty, all while destroying the youths' self image and expectations of what it means to be beautiful around the world.

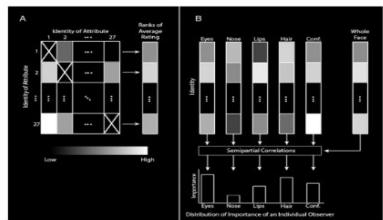
Methodology

TABLE I
REPRESENTATIVE DATABASES FOR FACIAL BEAUTY PREDICTION

Dataset	Image Num	Labels/Image	Beauty Class	Face Property	Face Landmarks	Publicly Available
V. Escorcia et al. [19]	104	21 ± 10	7	Caucasian Female	✓	x
F. Cen et al. [22]	23412	2	Asian Male/Female	✓	x	
H. Gao et al. [23]	215	46	19	Female	✓	x
J. Fan et al. [25]	432	30	7	General Female	✓	x
M. Reh et al. [34]	10141	78-540	10	Multiple Sampled from AVA [36]	x	x
SCUT-FBP [1]	500	70	5	Asian Female	✓	x
SCU14-FBP500	5500	60	5	Asian/Caucasian, Male/Female	✓	x

The SCUT-FBP5500 machine learning dataset analysis from the South China University of Technology was mainly centered around creating a new and improved dataset that is based on calculating through a machine learning model. This particular new dataset, called SCUT-FBP5500, used 5500 images of diverse demographic backgrounds. The images were taken so that facial features were placed on the faces between caucasian and asian males and females. Table one explains multiple different datasets and compares them to SCUT-FBP5500, where it succeeds compared to all of them in terms of facial recognition technology, as seen in the figure above. The authors, Lingyu Liang, Luojun Lin, Jianlin Jin*, Duorui Xie and Mengru Li, of the new dataset's creation relied on hand-crafting new features derived from visual recognition. This included the use of facial feature detection, including the location of face, nose, eyes, mouth, forehead and much more. It also checked for turned faces, decollagenation, signs of aging like defined lines, and the depth of each face due to bone, muscle, and fat structures. With the data collection of all these facial features, the authors then evaluated their own hand-crafted versions of the most to least beautiful faces on a 1-5 "beauty scale". The goal of making this new and improved beauty dataset was to include more diversity of faces, though diversity as we will later see in the Alipay analysis is far from completed for an app as big as Tik Tok.

The visual activity study conducted by Dr. Daniel Gill takes a very similar approach, methodology wise, to analyzing faces. He also used beauty scores and a grayscale, as seen in the figure below, which analyzed the contrasts in the face. This looked at mainly the nose, lips, hair, eyes, the whole face, and facial configuration, or otherwise in a grayscale. This particular study used a visual activity analysis to determine caucasian faces only, mainly observing only if the observed community had a certain set off facial features, and then comparing that with other facial attributes if the face is truly beautiful. The visual activity had six face observers, who were given various separate images of the facial features listed above separately and were asked to rate each image separately. Some of the results concluded that the differences in how women and men rated parts of the face, such as how women placed more importance on lips, while men placed more importance on eyes. Relating this with our last article we will mention in the next section, we will see how this analysis still has many problems, culturally and socially in western societies, and relating it all to TikTok's beauty algorithm.



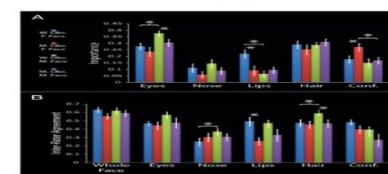
Analysis



Fig. 1 - The images with different facial properties and beauty scores from the proposed SCUT-FBP5500 benchmark dataset. The dataset download URL is shown below the title.

We will now compare the three stated issues with comparing them to a more direct article on specific social and cultural issues. The article in question is by Alman Yuzhu Peng, who explored a payment and wallet app called Alipay. This successful app in China introduced face filters in 2019, which saw a dramatic increase in use by the women especially in the way the filters beautify female faces. These filters smooth out the face and lighten it, but they are also known to change the face shape to be rounder and softer, while also enlarging the eyes and shrinking the nose. The design behind these filters were originally created for men, and then modified for women. This is a clear example of a patriarchal capitalist that is so prominent in Chinese socio-economic infrastructure. These patriarchal capitalist values persist in eastern technology, making it all the more noticeable how social standards of beauty over women in technology promote the use of the app even more. This ultimately attracts male and female consumers to this false narrative of beauty.

This target to promote the perfect appearance of what women should strive for and look up to can be related to the South China University of Technology data we mentioned before. Looking at the above image, the ideal women in asian and caucasian with the highest beauty scores center around highly edited faces of women; meaning they had very rounded faces, large eyes and more. The limitations of datasets set in place by limited examples proved by the male engineers is a stark problem in eastern societies in general, culturally and politically. We see some men and more than people in eastern countries that are more accepting of the visual activity of women, and then others who believe the narrative that the beauty provided by the technology can be mistaken as real. This hurts and promotes the cultural stigma that women have to go through in order to be deemed beautiful, including how we see many eastern cultures, especially in South Korea, rely on plastic surgery to obtain unnatural beauty standards provided by technology. In the Cosmetic Surgery and Self-esteem in South Korea: A Systematic Review and Meta-analysis, Dr. Daniel Gill et al. of UCR found that the desire for cosmetic surgery is so related with social networking sites, that many women and men have described self esteem, forcing them to get surgery for better job opportunities. When consumers and job recruiters expect more and more of their employees to look like the examples given to them by beauty standards set by datasets like SCUT-FBP5500, there is room for discrimination in peoples appearance, including skin color, skin conditions, physical deformities, and many more attributes that many can't control without facial filters. Now looking at the Gill study, we see that the bottom on the bottom used limitations in determining beauty.



Analysis

Though this study was more focused on how western caucasian women and men viewed each part of the face on a beauty scale, there were still patriarchal and toxic beauty opinions that showed through in this study. These opinions were seen to be much more critical of features like the nose, eyes, and lips based on how critical western beauty standards are in similarities to eastern beauty standards seen in the Alipay analysis. The Alipay analysis was much more focused on western beauty standards, while this study was more focused on eastern beauty standards. Both western and eastern societies are concluded to rely heavily on technological datasets to put a set-in stone beauty standards that go off preexisting discriminations and assumptions that hate against people who do not fit the pre established facial features that one should have for reaching fame or obtaining well-paying jobs.

Conclusion

Through all of these examples of datasets and studies relating to how the app TikTok operates based on beauty, we can determine that there are clear issues in how TikTok's beauty algorithm operates. Though with all of these analyses used in the examples given, we must conclude on the bigger issue at hand. We have been able to see how each of these examples use data obtained from real humans to determine beauty. Linking these issues back to TikTok, and its parent company, Bytedance, there is a much more pressing issue over just the visible issues with social media apps relying heavily on plastic surgery and what we see as real. With how Bytedance controls TikTok's beauty algorithm compared to the dataset we analyzed before; we can see how this company intentionally thrives off user data that we feed to it. Since western and eastern standards of beauty are highly sought after in social media, are really feeding the app our own toxic ideals of beauty. These ideals instill unrealistic expectations of physical beauty in society as a whole. With these toxic ideals and unrealistic expectations pushed through face filters, like how we saw through the Alipay article. The further social and cultural issues are also brought forth with how these studies also do not include people of color, facial irregularities, and physical deformities, and not taking their standards into consideration. We can link these studies with TikTok's algorithm based on the sheer number of women and men that use the app. Dr. Daniel Gill et al. of UCR found that those surrounding social media apps like TikTok in our society is to not destroy the algorithm, but to make the algorithms more inclusive by providing it new inclusive data provided by many racially different engineers. We must work further towards promoting other aspects of beauty that can be provided throughout the world, which is reachable due to how large the TikTok user base is. We can see that the reason why we turn to due to many cultural norms and differences successful businesses like Bytedance will need to address this, there is still a possibility to improve on algorithms by including inclusive images in TikTok's dataset. This will then help the beauty algorithm promote diversity throughout the app, and further help destroy negative stigmas and toxicity around racial differences in faces throughout the western and eastern societies.

References

- Domaika, F., Wang, K., Arganda-Carreras, L., Elorza, A., & Moujahid, A. (2020). Toward graph-based semi-supervised face beauty prediction. *Expert Syst. Appl.*, 142.
2. Gill, D. (2017). Women and men integrate facial information differently in appraising the beauty of a face. *Evolution and Human Behavior*, 38, 756-760.
3. Peng, A.Y. (2020). Alipay adds "beauty filters" to face-scan payments: a form of patriarchal control over women's bodies. *Feminist Media Studies*, 20, 582 - 585.
4. Yoon, S., & Kim, Y.A. (2020). Cosmetic Surgery and Self-esteem in South Korea: A Systematic Review and Meta-analysis. *Aesthetic plastic surgery*, 44(1), 229-238. <https://doi.org/10.1007/s00266-019-01515-1>

Digital | Kenyon

Home > STUDENT > DH



Digital Humanities @ Kenyon

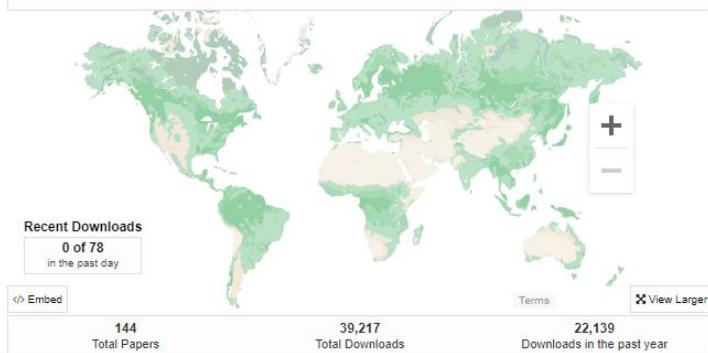


DIGITAL HUMANITIES

Gambier, Ohio, United States

Digital Humanities

This map shows recent readership activity for Digital Humanities.



Browse

[Collections](#)

Search

Enter search terms:

Search

▼

[Advanced Search](#)

Notify me via email or [RSS](#)

External Resources

[Digital Humanities at Kenyon](#)

[Digitization Lab](#)

[Library Services](#)

[Kenyon College](#)

digital.kenyon.edu/dh/

Fin