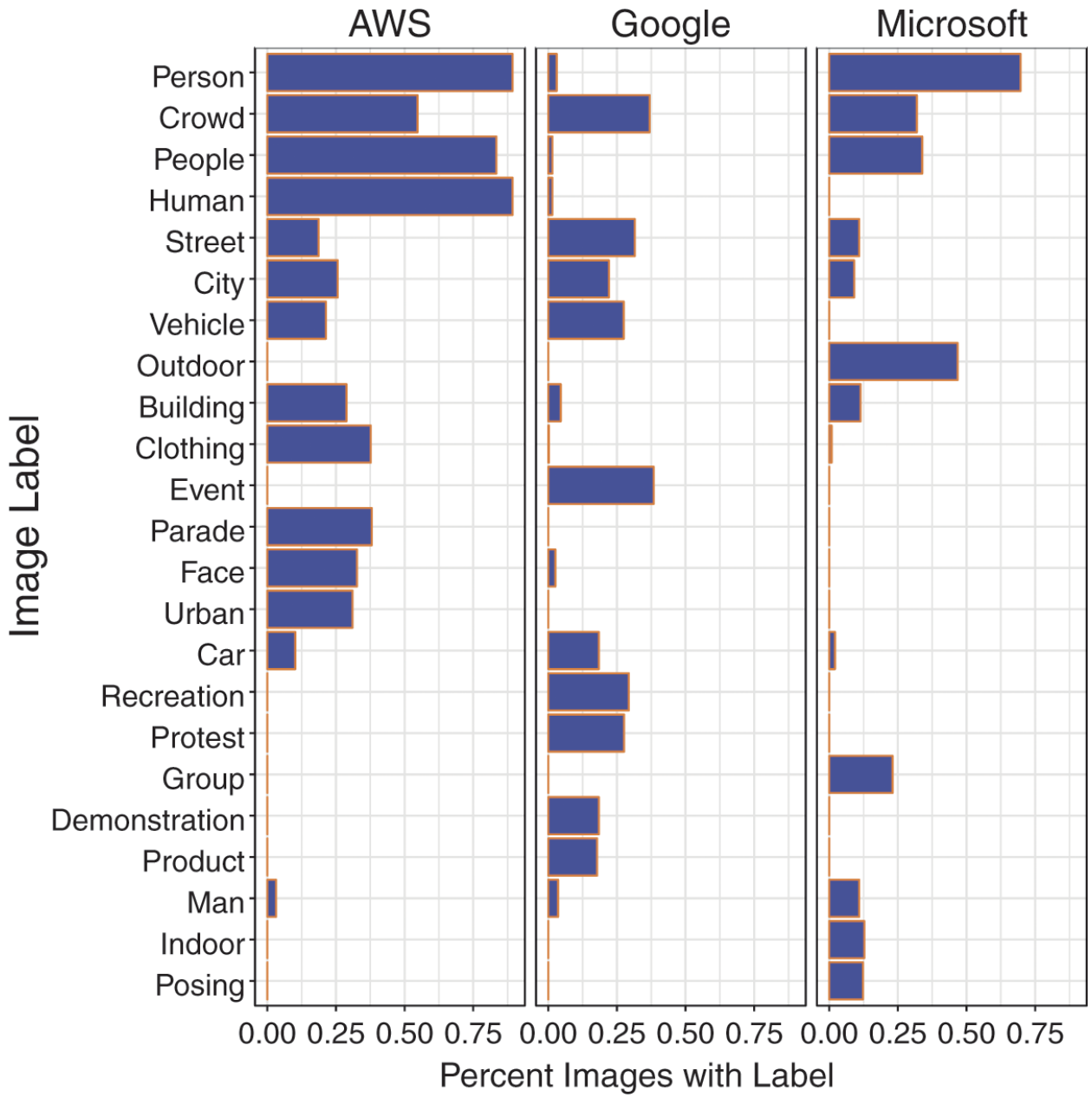
Being Critical With AI Investigations

Matti Nelimarkka
University of Helsinki & Aalto University
matti.nelimarkka@helsinki.fi

Structure

1. Using image recognition for analysis – issues and solutions
2. LLMs and values and Marxist LLM



Black live matters images from different commercial image recognition systems.

Webb Williams, N., Casas, A., & Wilkerson, J. D. (2020). *Images as Data for Social Science Research*. Cambridge University Press.


DO NOT USE IMAGE
RECOGNITION TOOLS
FOR SCHOLARLY WORK

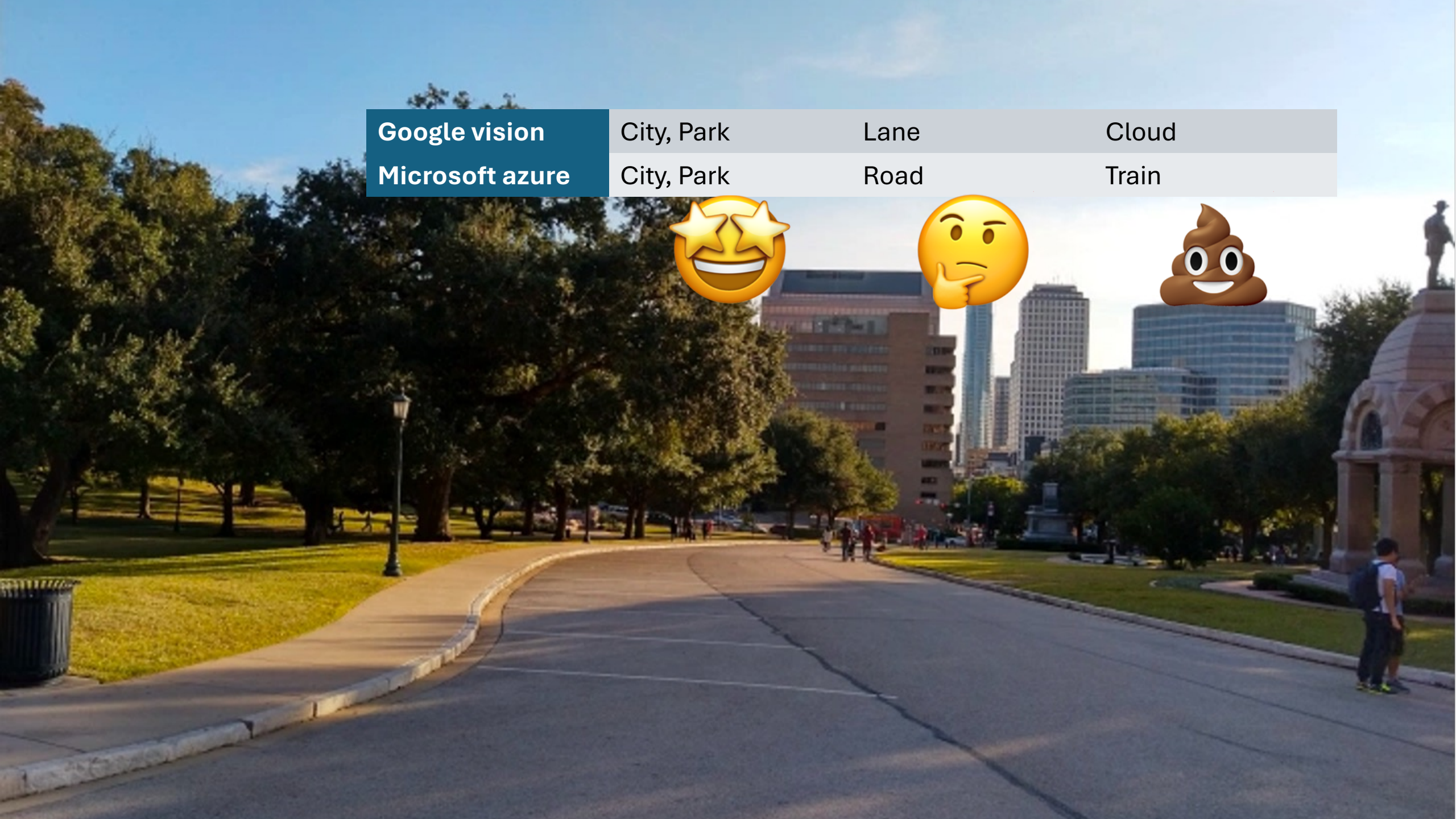
Why do we need an intercoder reliability?

Google Vision

Daytime, Sky, City, Public Space, Human Settlement, Road, Residential Area, Urban Area, Asphalt, Metropolitan Area, Tree, Infrastructure, Park, Road Surface, Downtown, Architecture, Neighbourhood, Skyline, Real Estate, Thoroughfare, Building, Suburb, Urban Design, Street, Lane, Walkway, Cloud, Recreation, Plaza, Town Square, Sidewalk, Nonbuilding Structure.

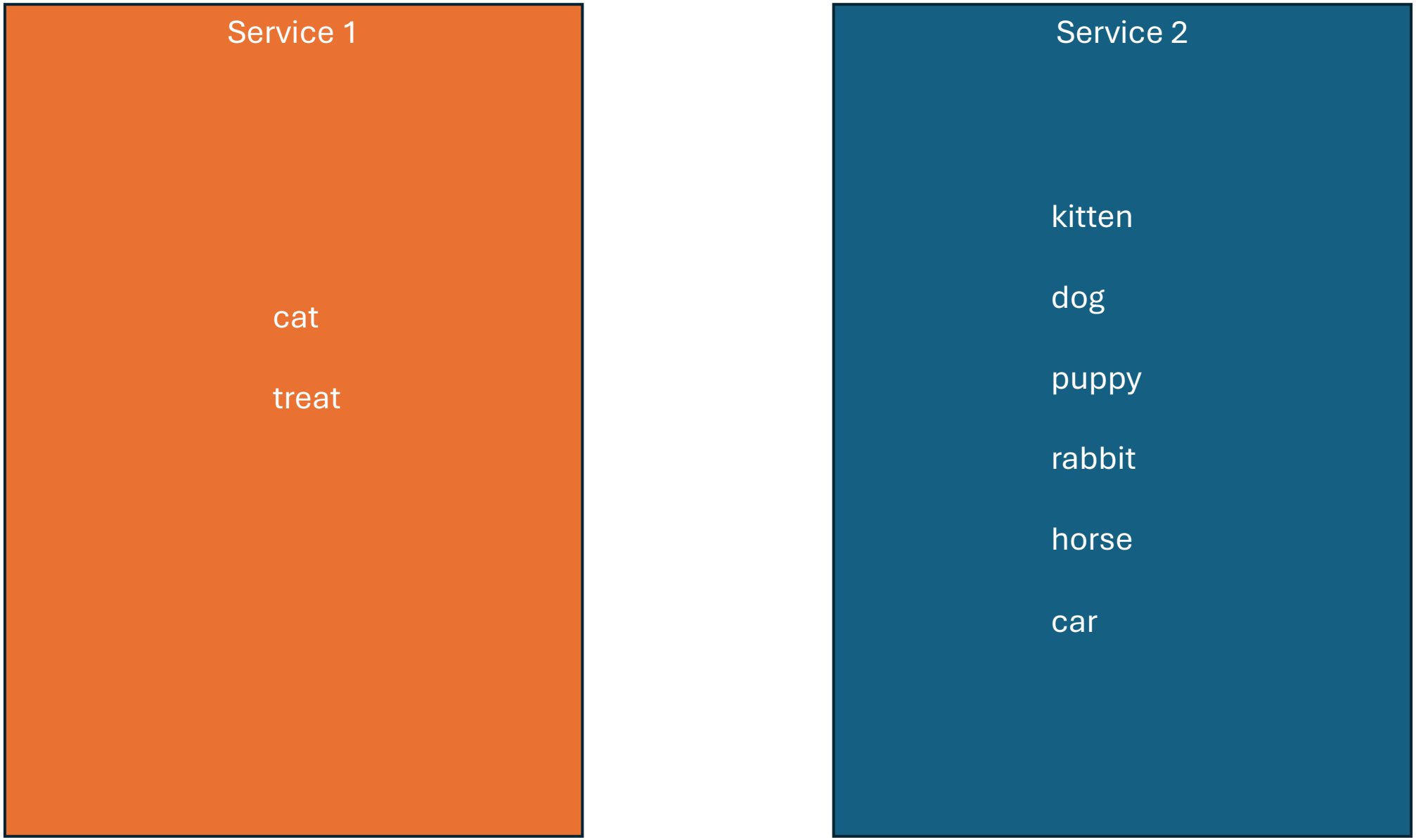
Google Vision
Daytime, Sky, City, Public Space, Human Settlement, Road, Residential Area, Urban Area, Asphalt, Metropolitan Area, Tree, Infrastructure, Park, Road Surface, Downtown, Architecture, Neighbourhood, Skyline, Real Estate, Thoroughfare, Building, Suburb, Urban Design, Street, Lane, Walkway, Cloud, Recreation, Plaza, Town Square, Sidewalk, Nonbuilding Structure.

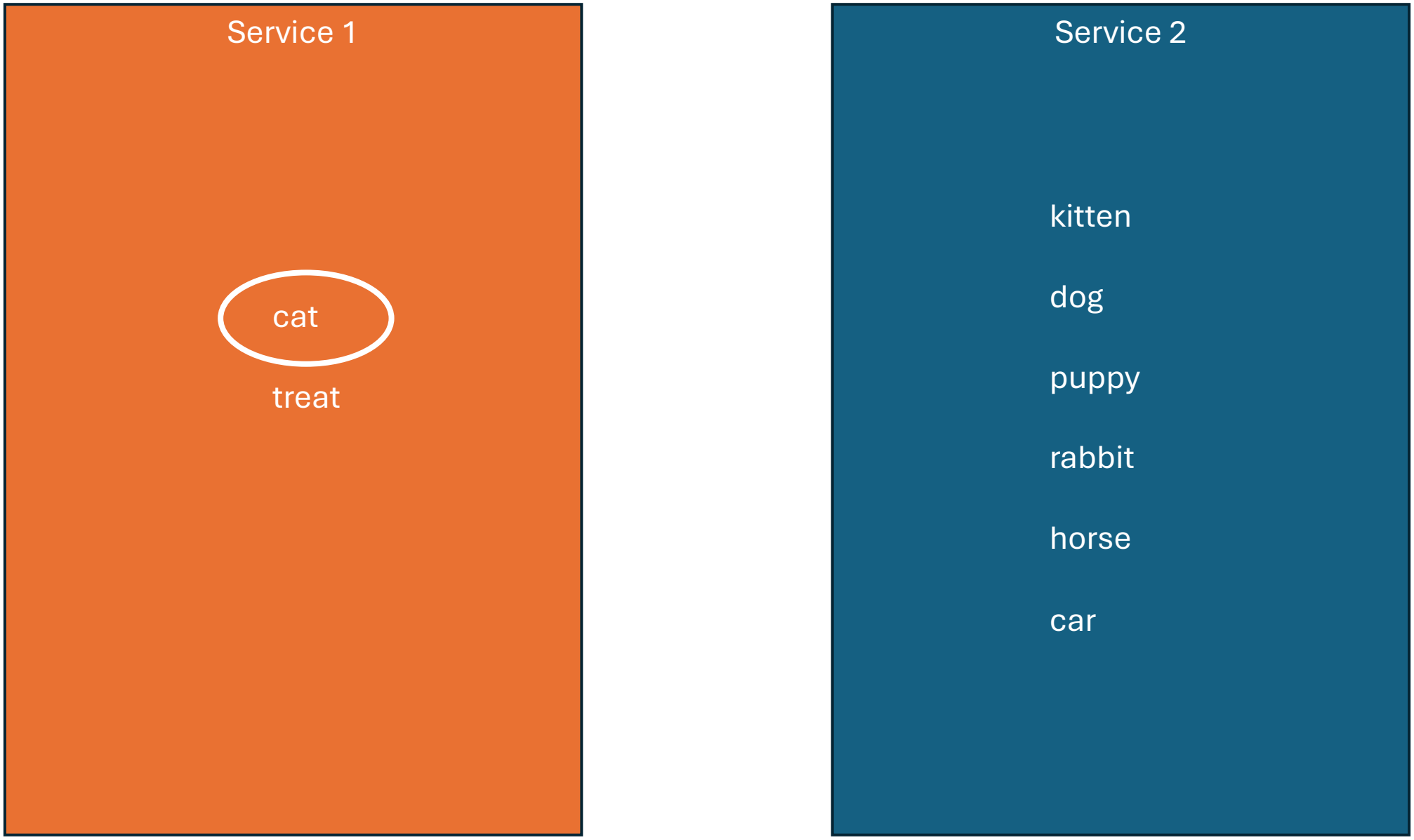
Microsoft Azure
Outdoor, Road, Street, City, Grass, Sitting, Empty, Red, View, Side, Stop, Green, Traffic, Sign, Track, Yellow, Park, Tall, Bench, Parked, Train, Hydrant, Riding, Bus

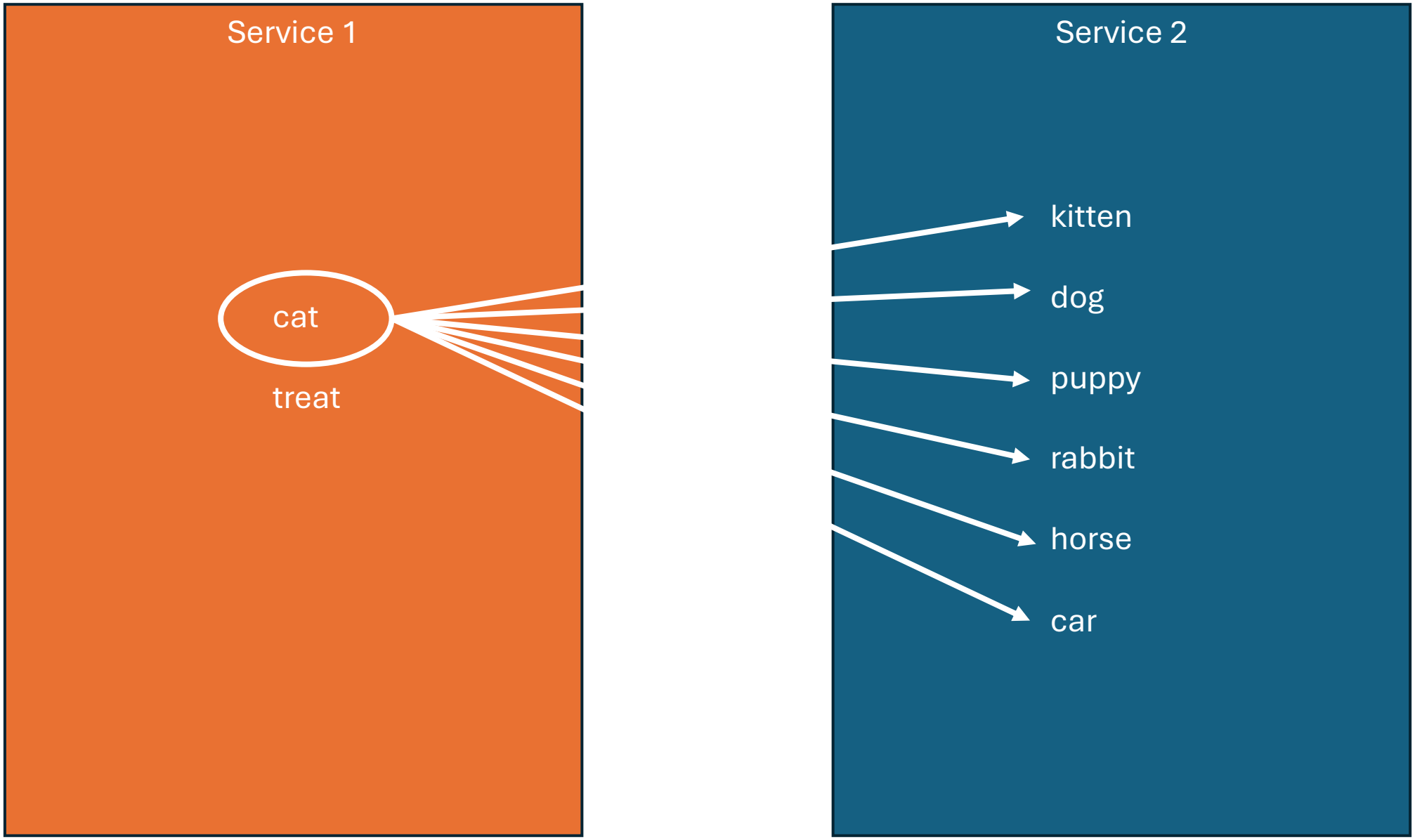

Google vision
City, Park
Lane
Cloud
Microsoft azure
City, Park
Road
Train

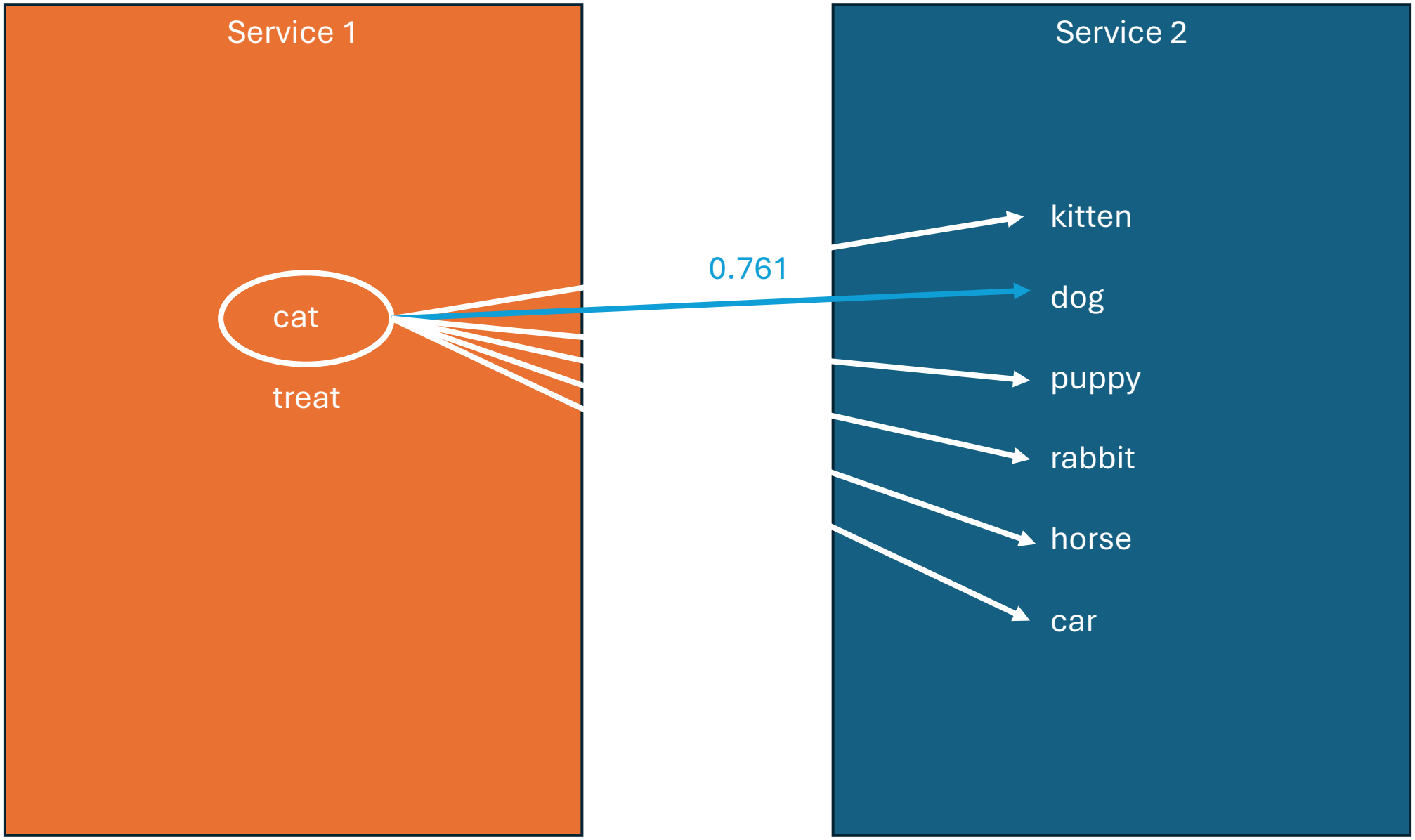
Corpus	Azure			Google			AWS		
	n	$\bar{x}$	$\overline{x_c}$	n	$\bar{x}$	$\overline{x_c}$	n	$\bar{x}$	$\overline{x_c}$
Pentzold et al. (2018)	4384	9.64	0.87	11,817	25.97	0.70	54,513	119.81	0.35
YLE News	139,342	11.43	0.87	369,735	30.33	0.71	1,396,434	114.54	0.36
Reddit: r/aesthetic	4010	9.44	0.84	15,027	35.36	0.72	51,251	120.59	0.36
Political party leaders' most recent	11,744	9.66	0.88	27,720	22.8	0.71	147,101	120.97	0.37
Thelwall et al. (2016)	8099	10.02	0.87	22,427	27.76	0.71	96,998	120.05	0.36
Hokka and Nelimarkka (2020)	77,381	8.24	0.88	207,579	22.11	0.70	1,147,848	122.27	0.35
Reddit: r/memes	6371	7.60	0.89	21,891	26.12	0.71	100,338	119.74	0.35
Political party leaders, pre-election	1996	10.62	0.89	4599	24.46	0.70	22,904	121.83	0.38

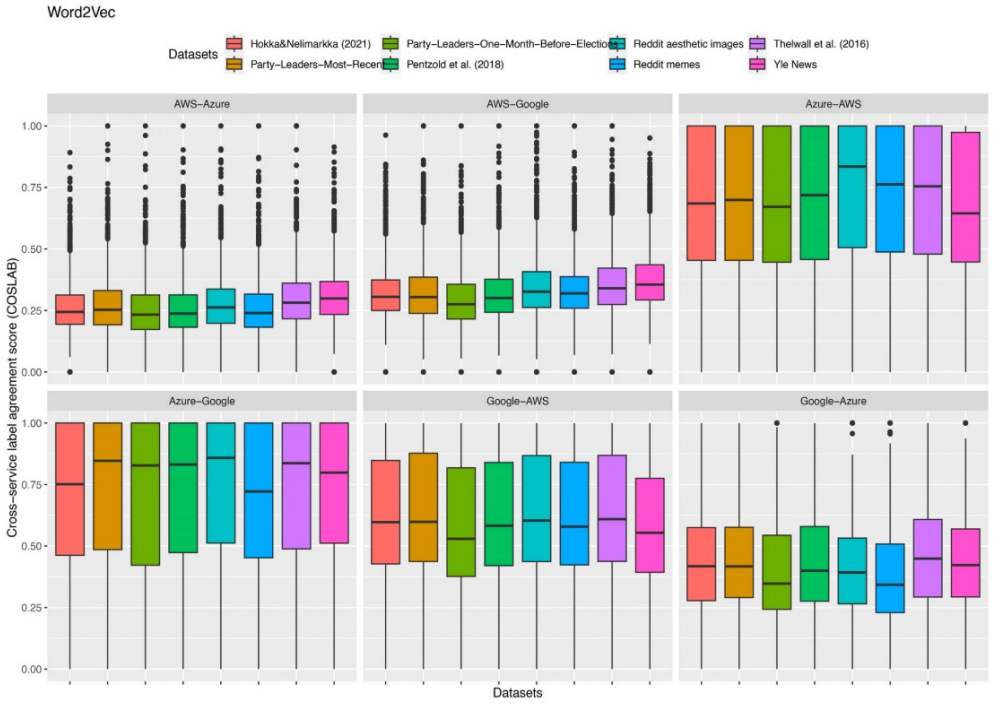
Berg, A., & Nelimarkka, M. (2023). Do you see what I see? Measuring the semantic differences in image-recognition services' outputs. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.24827

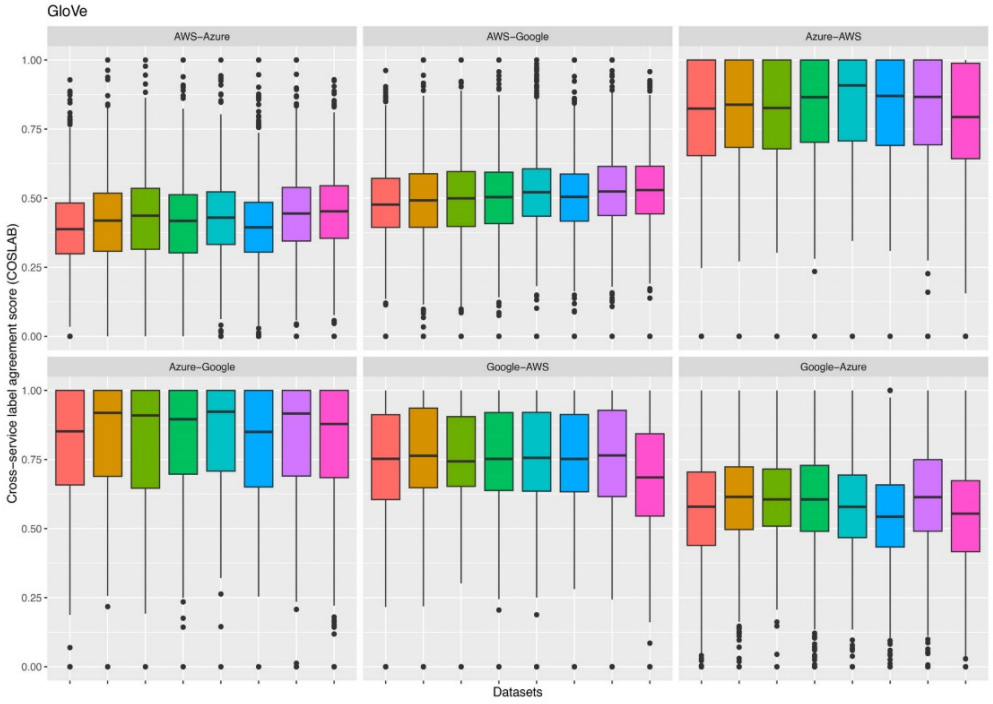

Service 1
cat
treat
Service 2
kitten
dog
puppy
rabbit
horse
car


Service 1
cat
treat
Service 2
kitten
dog
puppy
rabbit
horse
car


Service 1
cat
treat
Service 2
kitten
dog
puppy
rabbit
horse
car

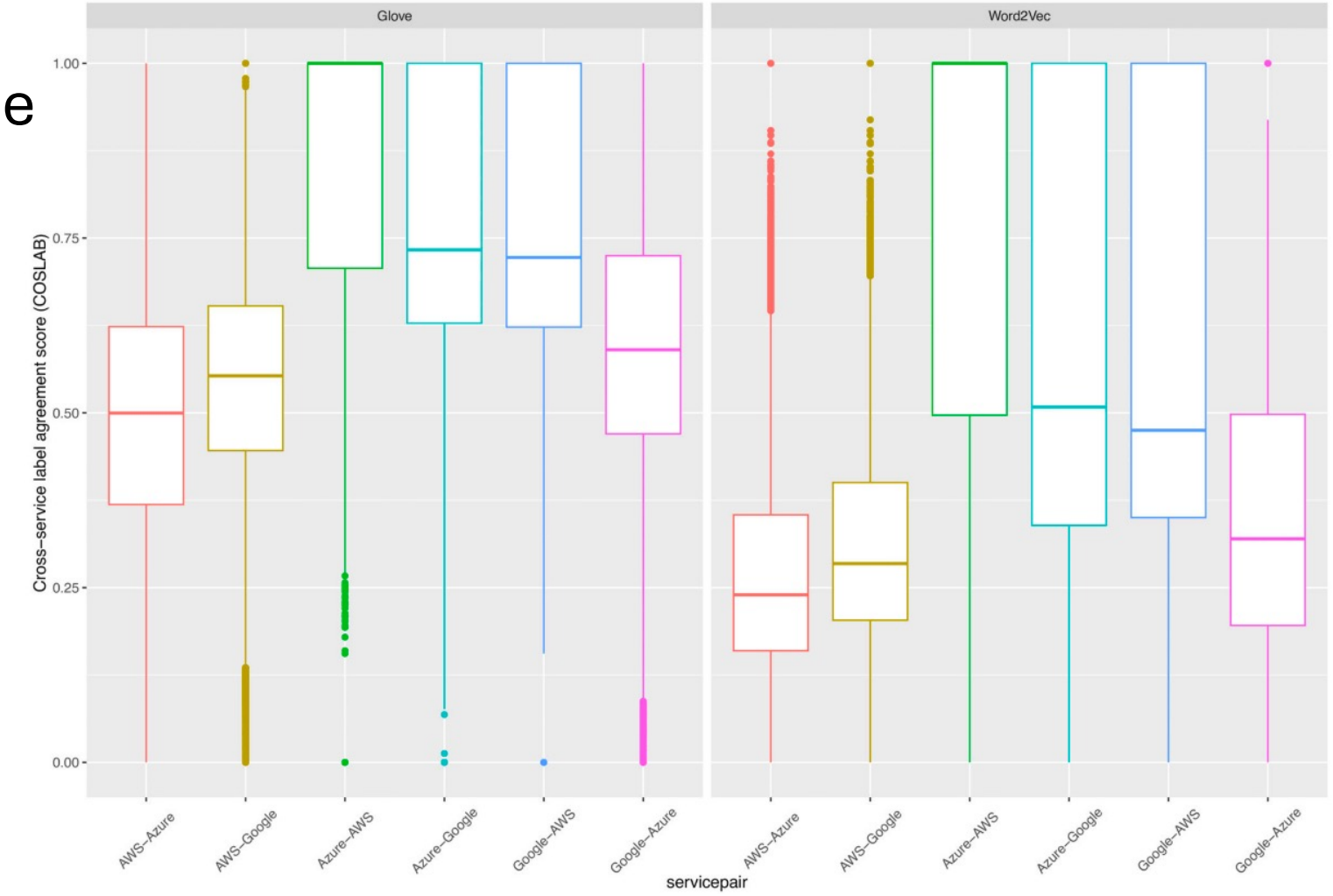

Service 1
Service 2
cat
treat
0.761
kitten
dog
puppy
rabbit
horse
car



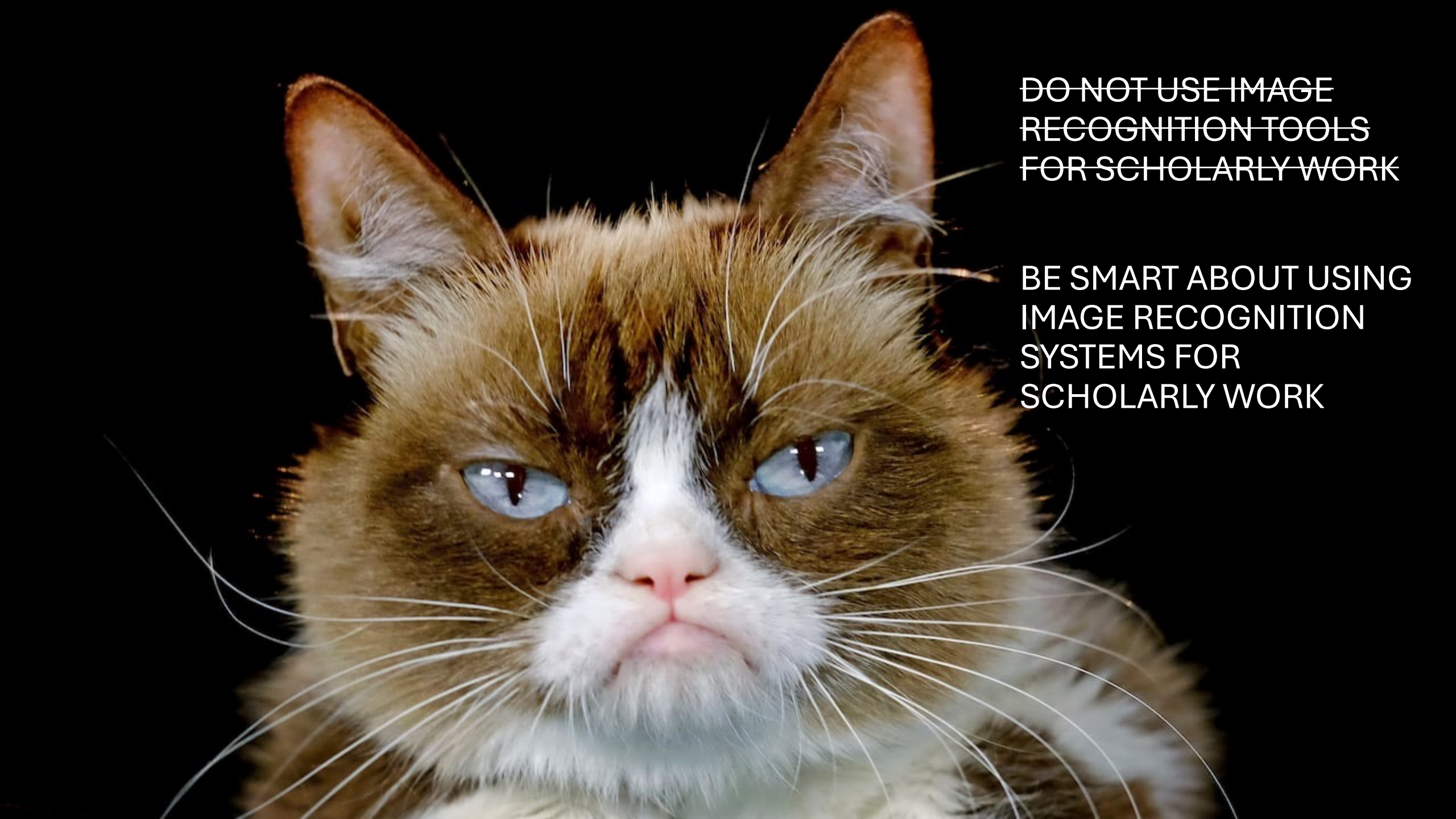


No clear differences between image catalogues.

Some service pairs seem to be better than others.




DO NOT USE IMAGE
RECOGNITION TOOLS
FOR SCHOLARLY WORK



~~DO NOT USE IMAGE RECOGNITION TOOLS FOR SCHOLARLY WORK~~

BE SMART ABOUT USING IMAGE RECOGNITION SYSTEMS FOR SCHOLARLY WORK

Strategy one

Use verythinng, accept mistakes

Strategy two

Conduct cross-service label agreement score to choose labels with sufficient agreement

Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics

Aylin Caliskan
University of Washington
Information School
Seattle, WA, USA
aylin@uw.edu

Pimparkar Parth Ajay
Birla Institute of Technology and Science
Department of Computer Science
Pilani, Rajasthan, India
f20180229@goa.bits-pilani.ac.in

Tessa Charlesworth
Harvard University
Department of Psychology
Cambridge, MA, USA
tessa_charlesworth@fas.harvard.edu

Robert Wolfe
University of Washington
Information School
Seattle, WA, USA
rwolfe3@uw.edu

Mahzarin R. Banaji
Harvard University
Department of Psychology
Cambridge, MA, USA
mahzarin_banaji@harvard.edu

ABSTRACT

Word embeddings are numeric representations of meaning derived from word co-occurrence statistics in corpora of human-produced texts. The statistical regularities in language corpora encode well-known social biases into word embeddings (e.g., the word vector for family is closer to the vector women than to men). Although efforts have been made to mitigate bias in word embeddings, with the hope of improving fairness in downstream Natural Language Processing (NLP) applications, these efforts will remain limited until we more deeply understand the multiple (and often subtle) ways that social biases can be reflected in word embeddings. Here, we focus on gender to provide a comprehensive analysis of group-based biases in widely-used static English word embeddings trained on internet corpora (GloVe 2014, fastText 2017). While some previous research has helped uncover biases in specific semantic associations between a group and a target domain (e.g., women – family), using the Single-Category Word Embedding Association Test, we demonstrate the widespread prevalence of gender biases that also show differences in: (1) frequencies of words associated with men versus women; (b) part-of-speech tags in gender-associated words; (c) semantic categories in gender-associated words; and (d) valence, arousal, and dominance in gender-associated words. We leave the analysis of non-binary gender to future work due to the challenges in accurate group representation caused by limitations inherent in data.

First, in terms of word frequency: we find that, of the 1,000 most frequent words in the vocabulary, 77% are more associated with men than women, providing direct evidence of a masculine default in the everyday language of the English-speaking world. Second, turning to parts-of-speech: the top male-associated words are typically verbs (e.g., fight, overpower) while the top female-associated words are typically adjectives and adverbs (e.g., giving, emotionally). Gender biases in embeddings also permeate parts-of-speech. Third, for semantic categories: bottom-up, cluster analyses of the top 1,000 words associated with each gender. The top male-associated concepts include roles and domains of big tech, engineering, religion, sports, and violence; in contrast, the top female-associated concepts are less focused on roles, including, instead, female-specific slurs and sexual content, as well as appearance and kitchen terms. Fourth, using human ratings of word valence, arousal, and dominance from a ~20,000 word lexicon, we find that male-associated words are higher on arousal and dominance, while female-associated words are higher on valence. Ultimately, these findings move the study of gender bias in word embeddings beyond the basic investigation of semantic relationships to also study gender differences in multiple manifestations in text. Given the central role of word embeddings in NLP applications, it is essential to more comprehensively document where biases exist and may remain hidden, allowing them to persist without our awareness throughout large text corpora.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Natural language processing**; **Learning latent representations**; **Learning paradigms**; **Cognitive science**.

KEYWORDS

word embeddings, AI bias, gender bias, psycholinguistics, representation, masculine default

ACM Reference Format:
Aylin Caliskan, Pimparkar Parth Ajay, Tessa Charlesworth, Robert Wolfe, and Mahzarin R. Banaji. 2022. Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics. In *AIES '22: AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, August 2022, Oxford, England.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3514094.3534162

AIES '22, August 2022, Oxford, England

ACM ISBN 978-1-4503-9247-1/22/08.
https://doi.org/10.1145/3514094.3534162

1 INTRODUCTION

Today, the vast majority of our daily tasks are facilitated and enhanced through the application of Natural Language Processing (NLP), from simple machine translation to automated resume screening to auto-complete in emails [7]. The core component of many of

What social attitudes about gender does BERT encode? Leveraging insights from psycholinguistics

Julia Watson[1] **Barend Beekhuizen**[2] **Suzanne Stevenson**[1]

[1]Department of Computer Science
University of Toronto
{jwatson, suzanne}@cs.toronto.edu

[2]Department of Language Studies
University of Toronto, Mississauga
barend.beekhuizen@utoronto.ca

Abstract

Much research has sought to evaluate the degree to which large language models reflect social biases. We complement such work with an approach to elucidating the connections between language model predictions and people's social attitudes. We show how word preferences in a large language model reflect social attitudes about gender, using two datasets from human experiments that found differences in gendered or gender neutral word choices by participants with differing views on gender (progressive, moderate, or conservative). We find that the language model BERT takes into account factors that shape human lexical choice of such language, but may not weigh those factors in the same way people do. Moreover, we show that BERT's predictions most resemble responses from participants with moderate to conservative views on gender. Such findings illuminate how a language model: (1) may differ from people in how it deploys words that signal gender, and (2) may prioritize some social attitudes over others.

1 Introduction

Language choices are revealing about speakers' social attitudes – their (evaluative) beliefs, views, and expectations about social phenomena. If a café advertises "gingerbread people," instead of "gingerbread men" (example adapted from Papineau et al., 2022), people may make inferences about the social views of the café owners based on their avoidance of the traditional masculine term. Social attitudes typically surface in less "pointed" but higher stakes scenarios, such as a speaker using the pronoun *they* to refer to a colleague who identifies as nonbinary, reflecting the speaker's acceptance of nonbinary identities.

Much work on the social knowledge encoded in language technology has focused on evaluating whether models encode stereotypical/harmful associations (e.g., Caliskan et al., 2017; Rudinger et al., 2018), and if so, removing them to "de-bias" NLP (e.g., Bolukbasi et al., 2016; Zhao et al., 2018). However, social knowledge permeates language (e.g., Nguyen et al., 2021), and what counts as harmful depends on one's perspective (e.g., Blodgett et al., 2020). To deal effectively with potentially harmful associations in NLP, we need a clear understanding of how social attitudes are linked to the language choices people make, so that we can assess the language choices of our technologies.

Here we seek to understand **what social attitudes a large language model encodes**, specifically social attitudes about gender. To address this question, we draw on datasets from two psycholinguistics studies, both of which included language tasks involving gendered and gender neutral language choices, *and* surveys eliciting the same participants' social attitudes on gender. By explicitly linking people's language choices with their social attitudes, this data enables us to evaluate how social attitudes are reflected in the language choices encoded in an NLP model, and to quantify the extent to which a language model propagates certain views over others (cf. Bender et al., 2021).

In the first study we draw on, Papineau et al. (2022) elicited preferences for feminine, masculine, and gender neutral variants of role nouns, such as *firewoman/fireman/firefighter*, and found that choices to use gendered over gender neutral variants can reflect more rigid views about men's and women's social roles. In the second study, Camilliere et al. (2021) elicited acceptability judgements of singular *they* pronouns in contexts like *My friend$_i$ said they$_i$ would be coming late to dinner*. They found that lower acceptability ratings of singular *they* are associated with less acceptance of nonbinary people. It is important to determine if language models make similar choices to these, since if they do, they may spread and reinforce such attitudes, which may contribute to gender stereotyping (Sczesny et al., 2016), or nonbinary erasure

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics
Volume 1: Long Papers, pages 6790–6809
July 9-14, 2023

Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development

MORGAN KLAUS SCHEUERMAN*, University of Colorado Boulder, USA
EMILY DENTON, Google Research, USA
ALEX HANNA, Google Research, USA

Data is a crucial component of machine learning. The field is reliant on data to train, validate, and test models. With increased technical capabilities, machine learning research has boomed in both academic and industry settings, and one major focus has been on computer vision. Computer vision is a popular domain of machine learning increasingly pertinent to real-world applications, from facial recognition in policing to object detection for autonomous vehicles. Given computer vision's propensity to shape machine learning research and impact human life, we seek to understand disciplinary practices around dataset documentation — how data is collected, curated, annotated, and packaged into datasets for computer vision researchers and practitioners to use for model tuning and development. Specifically, we examine what dataset documentation communicates about the underlying values of vision data and the larger practices and goals of computer vision as a field. To conduct this study, we collected a corpus of about 500 computer vision datasets, from which we sampled 114 dataset publications across different vision tasks. Through both a structured and thematic content analysis, we document a number of values around accepted data practices, what makes desirable data, and the treatment of humans in the dataset construction process. We discuss how computer vision datasets authors value efficiency at the expense of care; universality at the expense of contextuality; impartiality at the expense of positionality; and model work at the expense of data work. Many of the silenced values we identify sit in opposition with social computing practices. We conclude with suggestions on how to better incorporate silenced values into the dataset creation and curation process.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**; • **Computing methodologies** → *Artificial intelligence.*

Additional Key Words and Phrases: Datasets, computer vision, work practice, machine learning, values in design

ACM Reference Format:
Morgan Klaus Scheuerman, Emily Denton, and Alex Hanna. 2021. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 317 (October 2021), 37 pages. https://doi.org/10.1145/3476058

1 Introduction

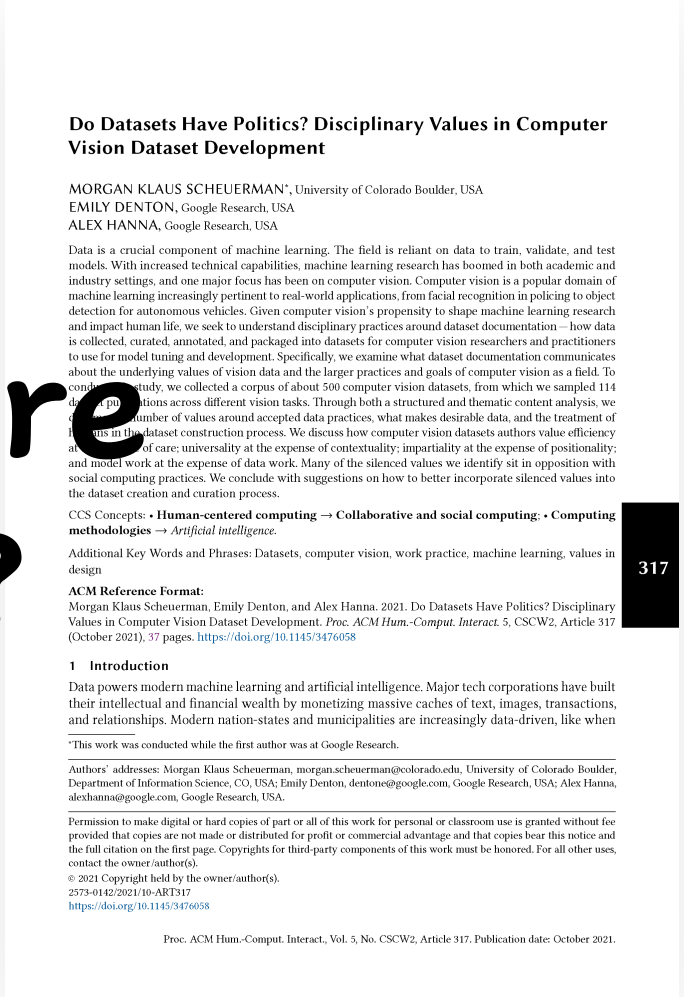
Data powers modern machine learning and artificial intelligence. Major tech corporations have built their intellectual and financial wealth by monetizing massive caches of text, images, transactions, and relationships. Modern nation-states and municipalities are increasingly data-driven, like when

*This work was conducted while the first author was at Google Research.

Authors' addresses: Morgan Klaus Scheuerman, morgan.scheuerman@colorado.edu, University of Colorado Boulder, Department of Information Science, CO, USA; Emily Denton, dentone@google.com, Google Research, USA; Alex Hanna, alexhanna@google.com, Google Research, USA.

2573-0142/2021/10-ART317
https://doi.org/10.1145/3476058

What if we measure things with these?



The Meaning of 'Theory'*

GABRIEL ABEND
Northwestern University

'Theory' is one of the most important words in the lexicon of contemporary sociology. Yet, their ubiquity notwithstanding, it is quite unclear what sociologists mean by the words 'theory,' 'theoretical,' and 'theorize.' I argue that confusions about the meaning of 'theory' have brought about undesirable consequences, including conceptual muddles and even downright miscommunication. In this paper I tackle two questions: (a) what does 'theory' mean in the sociological language?; and (b) what ought 'theory' to mean in the sociological language? I proceed in five stages. First, I explain why one should ask a *semantic* question about 'theory.' Second, I lexicographically identify seven different senses of the word, which I distinguish by means of subscripts. Third, I show some difficulties that the current lack of semantic clarity has led sociology to. Fourth, I articulate the question, 'what ought "theory" to mean?,' which I dub the 'semantic predicament' (SP), and I consider what one can learn about it from the theory literature. Fifth, I recommend a 'semantic therapy' for sociology, and advance two arguments about SP: (a) the principle of practical reason—SP is to a large extent a political issue, which should be addressed with the help of political mechanisms; and (b) the principle of ontological and epistemological pluralism—the solution to SP should not be too ontologically and epistemologically demanding.

1. INTRODUCTION

'Theory' is one of the most important words in the lexicon of contemporary sociology. I am not referring only—in fact, not principally—to the subfield of sociological theory. The words 'theory,' 'theoretical,' and 'theorize' are constantly and consequentially used by all sociologists. For instance, one way of describing what sociologists of social movements do is to say that they develop 'theories' about social movements. What sociologists of the family do is to develop 'theories' about the family. And so on. Moreover, it is a widespread belief that empirical sociological research should be driven or informed by 'theory.' Thus, sociology journals tend to reject 'atheoretical' and 'undertheorized' papers, as well as papers that fail to make a 'theoretical

*Direct correspondence to: Gabriel Abend, Department of Sociology, Northwestern University, 1810 Chicago Ave., Evanston, IL 60208 (g-abend@northwestern.edu). The origins of this paper lie in an invitation to reflect on the present and future of sociological theory. I am thankful to the organizers of the Junior Theorists Symposium 2005—Mathieu Deflem, Marion Fourcade, and Neil Gross—for this invitation, and to my discussant, Charles Camic. I also benefited from conversations with fellow "junior theorists" Pierre Kremp, Simone Polillo, Isaac Reed, Erika Summers-Effler, Jonathan VanAntwerpen, and Robb Willer. I presented a slightly different version of the argument at the 2005 Annual Retreat of the Society for Comparative Research, hosted by Central European University. At this conference I received useful suggestions from my discussant, Jack Goldstone, as well as from Carsten Schneider and Robin Stryker. Finally, I am indebted to Sareeta Amrute, Charles Camic, Mathieu Deflem, Marion Fourcade, Neil Gross, Carol Heimer, Adam Kissel, Donald Levine, Richard Morales, Michael Sauder, Arthur Stinchcombe, Devin B. Terhune, and the *Sociological Theory* editors and reviewers for their comments and criticisms on earlier drafts of this paper.

Sociological Theory 26:2 June 2008

Applying Computational Analysis to Textual Data from the Wild: A Feminist Perspective

Shauna Julia Concannon
Open Lab, Newcastle University
Newcastle, United Kingdom
shauna.concannon@ncl.ac.uk

Madeline Balaam
Media Technology and Interaction Design
KTH Royal Institute of Technology
Stockholm, Sweden
balaam@kth.se

Emma Simpson
Open Lab, Newcastle University
Newcastle, United Kingdom
emma.simpson@ncl.ac.uk

Rob Comber
RISE SICS
Stockholm, Sweden
rob.comber@ri.se

ABSTRACT
With technologies that afford much larger-scale data collection than previously imagined, new ways of processing and interpreting qualitative textual data are required. HCI researchers use a range of methods for interpreting the 'full range of human experience' from qualitative data, however, such approaches are not always scalable. Feminist geography seeks to explore how diverse and varied accounts of place can be understood and represented, whilst avoiding reductive classification systems. In this paper, we assess the extent to which unsupervised topic models can support such a research agenda. Drawing on literature from Feminist and Critical GIS, we present a case study analysis of a Volunteered Geographic Information dataset of reviews about breastfeeding in public spaces. We demonstrate that topic modelling can offer novel insights and nuanced interpretations of complex concepts such as privacy and be integrated into a critically reflexive feminist data analysis approach that captures and represents diverse experiences of place.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation: Miscellaneous; H.3.1 Information Storage and Retrieval: Content Analysis and Indexing

Author Keywords
Topic modelling; feminism; human-data-interaction; data analysis; geodata; GIS; feminist GIS; critical GIS; text analysis.

CHI 2018 April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04.
DOI: https://doi.org/10.1145/3173574.3173800

INTRODUCTION

Researchers in Human-Computer-Interaction (HCI) have increasingly focused their efforts on in the wild deployments of technologies that aim to empower communities. Working in this way can support extended engagements and garner insights from in-situ user experiences. However, the data collected can be large and unwieldy and finding scalable ways of working with such rich data can be challenging [7, 3, 19].

Recent discourse within HCI has brought feminist perspectives to bear upon the theory, methodologies and practices of interactive systems research, calling for more direct engagement with gender and the development of a socio-technical theory of gender [47]. Bardzell highlights that "feminist HCI entails critical perspectives that could help reveal unspoken values within HCI's dominant research and design paradigms and underpin the development of new approaches, methods and design variations" [4]. Bardzell and Bardzell outline key methodological positions characteristic of a feminist HCI methodology, such as a commitment to understanding participants' experiences in their own terms, the development of mixed-methods approaches which support dialectal information gathering, and reflexivity [5].

Geographical information systems (GIS) have become prevalent in mobile devices, and are applied at various levels of urban place-making, from urban planning and crisis informatics to individual, persuasive and aesthetic interventions. Such systems have become instrumental in decision-making processes for cities, organisations and citizens, and researchers have grappled with the mechanisms through which to make sense of GIS and its relationship to place-making. For instance, although prominent accounts of place-making in HCI focus on experiential accounts (e.g. [18]), it is less evident that such approaches account for systemic roots of such experiences [51]. On the other hand, computational accounts such as crisis informatics and urban informatics for smart cities, draw our attention to the infrastructures of urban spaces, but may not

What is the role of Weltanschauung in (computational) social science?

The Meaning of 'Theory'*

GABRIEL ABEND
Northwestern University

'Theory' is one of the most important words in the lexicon of contemporary sociology. Yet, their ubiquity notwithstanding, it is quite unclear what sociologists mean by the words 'theory,' 'theoretical,' and 'theorize.' I argue that confusions about the meaning of 'theory' have brought about undesirable consequences, including conceptual muddles and even downright miscommunication. In this paper, I tackle two questions: (a) what does 'theory' mean in the sociological language? and (b) what ought 'theory' to mean in the sociological language. I proceed in five steps. First, I explain why one should ask a *semantic* question about 'theory.' Second, I lexicographically identify seven different senses of the word which I distinguish by means of descriptive labels. Third, I show some conflicts that the current lack of semantic clarity has led sociology to. Fourth, I articulate the question, 'what ought "theory" to mean?,' which I dub the 'semantic predicament' (SP), and I consider what one can learn about it from the theory literature. Fifth, I recommend a 'semantic therapy' for sociology, and advance two arguments about (a) the principles of practical reason—SP is to a large extent a political issue, which [illegible] be addressed [illegible] and (b) the principles of ontological and epistemological pluralism—the solution to SP should not be sought ontologically a priori epistemological demanding.

1. INTRODUCTION

'Theory' is [illegible] the most important words in the lexicon of contemporary sociology. I am not referring only to the fact that principles of the subfields of sociological theory [illegible] words 'theory,' 'theoretical,' and 'theorize' are frequently and commonly used by all sociologists. For instance, the way of describing what sociologists of social movements do is to say that they develop 'theories' about social movements. What sociologists of the family do is to develop 'theories' about the family. And so on. Moreover, it is a widespread belief that empirical sociological research should be driven or informed by 'theory.' Thus, sociology journals tend to reject 'atheoretical' and 'undertheorized' papers, as well as [illegible] to make 'theoretical

*Direct correspondence to: Gabriel Abend, Department of Sociology, Northwestern University, 1810 Chicago Ave., Evanston, IL 60208 (gabend@northwestern.edu). I presented a version of the paper lie in an invitation to reflect on the present and future of sociological theory. I am thankful to the organizers of the Junior Theorists Symposium 2005—Mathieu Deflem, Marion Fourcade, and Neil Gross—for this invitation, and to my discussant, Charles Camic. I also benefited from conversations with fellow "junior theorists" Pierre Kremp, Simone Polillo, Isaac Reed, Erika Summers-Effler, Jonathan VanAntwerpen, and Robb Willer. I presented a slightly different version of the argument at the 2005 Annual Retreat of the Society for Comparative Research, hosted by Central European University. At this conference I received useful suggestions from my discussant, Jack Goldstone, as well as from Carsten Schneider and Robin Stryker. Finally, I am indebted to Sareeta Amrute, Charles Camic, Mathieu Deflem, Marion Fourcade, Neil Gross, Carol Heimer, Adam Kissel, Donald Levine, Richard Morales, Michael Sauder, Arthur Stinchcombe, Devin B. Terhune, and the *Sociological Theory* editors and reviewers for their comments and criticisms on earlier drafts of this paper.

Sociological Theory 26:2 June 2008

Applying Computational Analysis to Textual Data from the Wild: A Feminist Perspective

Shauna Julia Concannon
Open Lab, Newcastle University
Newcastle, United Kingdom
shauna.concannon@ncl.ac.uk

Madeline Balaam
Media Technology and Interaction Design
KTH Royal Institute of Technology
Stockholm, Sweden
balaam@kth.se

Emma Simpson
Open Lab, Newcastle University
Newcastle, United Kingdom
emma.simpson@ncl.ac.uk

Rob Comber
RISE SICS
Stockholm, Sweden
rob.comber@ri.se

ABSTRACT
With [illegible]es that afford [illegible]ch large data collection that [illegible]eviously imagined, [illegible] ways of processing and interpreting qualitative textual data are required. [illegible] researchers use [illegible] methods for interpreting the full range of human experience' from qualitative data, however, such approaches [illegible] not always scalable. Feminist geography seeks to explore how diverse and varied accounts of place can be understood and represented, whilst avoiding reductive classification systems. In this paper, we assess the extent to which unsupervised topic models can support such a research agenda. Drawing on literature from feminist and Critical GIS, we present a case study analysis of a [illegible]d Guardian Information dataset of views about breastfeeding in public spaces. We demonstrate that topic modelling can offer novel insights and nuanced [illegible]ions of complex concepts such as agency and be integrated into a critically reflexive feminist data analysis approach that captures and represents diverse experiences of place.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation: Miscellaneous; H.3.1 Information Storage and Retrieval: Content Analysis and Indexing

Author Keywords
Topic modelling; feminism; human-data-interaction; data analysis; geodata; GIS; feminist GIS; critical GIS; text analysis.

CHI 2018 April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04.
DOI: https://doi.org/10.1145/3173574.3173800

INTRODUCTION

Researchers in Human-Computer-Interaction (HCI) have increasingly focused their efforts on in the wild deployments of technologies that aim to empower communities. Working in this way can support extended engagements and garner insights from in-situ user experiences. However, the data collected can be large and unwieldy and finding scalable ways of working with such rich data can be challenging [7, 3, 19].

Recent discourse within HCI has brought feminist perspectives to bear upon the theory, methodologies and practices of interactive systems research, calling for more direct engagement with gender in the development of a socio-technical theory of gender [4]. Bardzell highlights that "feminist HCI entails critical perspectives that could help reveal unspoken values within HCI research and design paradigms and underpin insights into the development of new approaches, methods and design variations" [4]. Bardzell and Bardzell outline key methodological positions characteristic of a feminist HCI methodology, such as a commitment to understanding participants' experiences in their own terms, the development of mixed-methods approaches which support dialectal information gathering, and reflexivity [5].

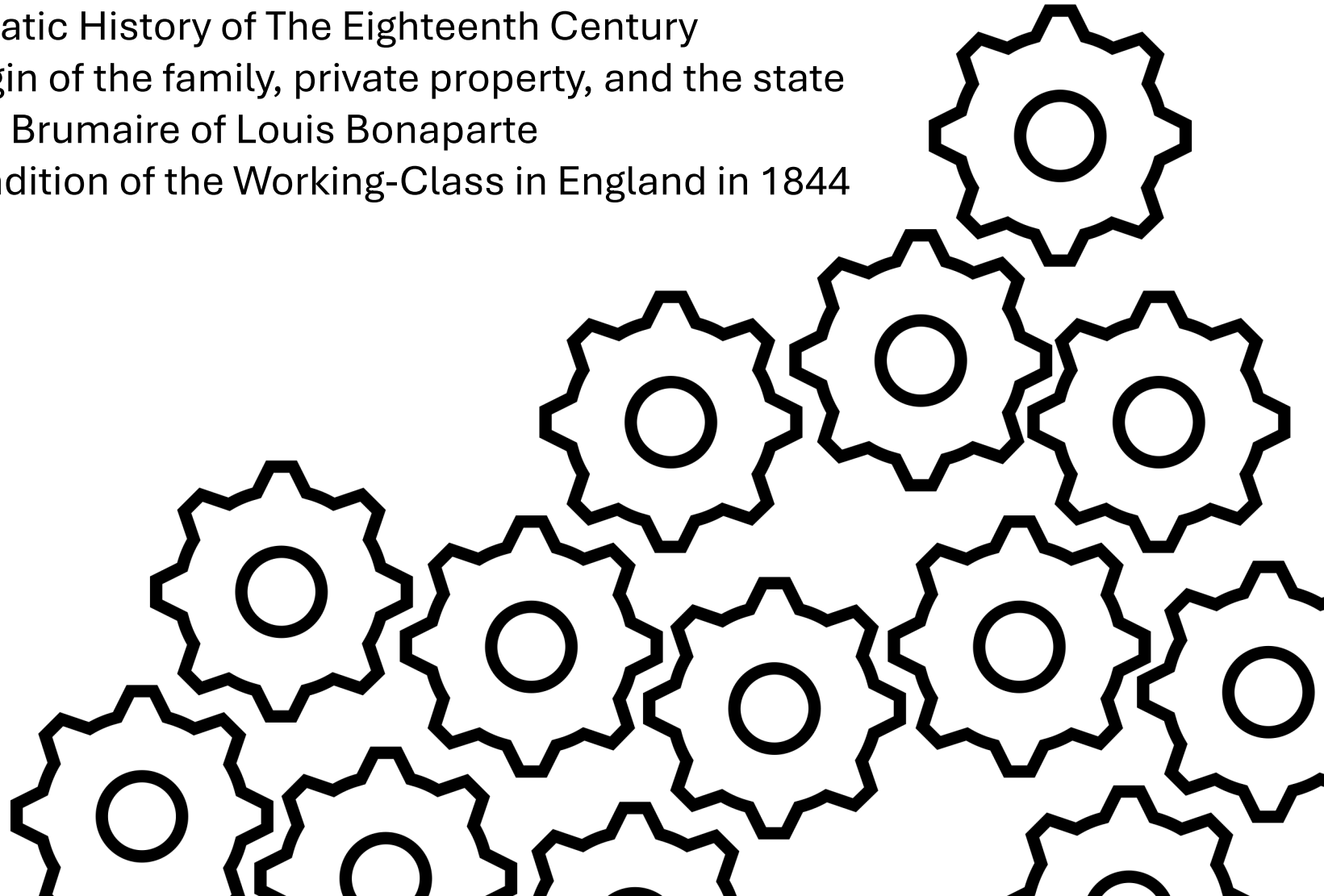
Geographical information systems (GIS) have become prevalent in mobile devices, and are applied at various levels of urban place-making, from urban planning and crisis informatics to individual, persuasive and aesthetic interventions. Such systems have become instrumental in decision-making processes for cities, organisations and citizens, and researchers have grappled with the mechanisms through which to make sense of GIS and its relationship to place-making. For instance, although prominent accounts of place-making in HCI focus on experiential accounts (e.g. [18]), it is less evident that such approaches account for systemic roots of such experiences [51]. On the other hand, computational accounts such as crisis informatics and urban informatics for smart cities, draw our attention to the infrastructures of urban spaces, but may not

- Karl Marx & Friedrich Engels: The Communist Manifesto
- Karl Marx: A Contribution to the Critique of Political
- Karl Marx & Friedrich Engels: Revolution and Counter-Revolution; Or, Germany in 1848
- Karl Marx : Secret Diplomatic History of The Eighteenth Century
- Friedrich Engels: The origin of the family, private property, and the state
- Karl Marx: The Eighteenth Brumaire of Louis Bonaparte
- Friedrich Engels: The Condition of the Working-Class in England in 1844

- Karl Marx & Friedrich Engels: The Communist Manifesto
- Karl Marx: A Contribution to the Critique of Political
- Karl Marx & Friedrich Engels: Revolution and Counter-Revolution; Or, Germany in 1848
- Karl Marx : Secret Diplomatic History of The Eighteenth Century
- Friedrich Engels: The origin of the family, private property, and the state
- Karl Marx: The Eighteenth Brumaire of Louis Bonaparte
- Friedrich Engels: The Condition of the Working-Class in England in 1844

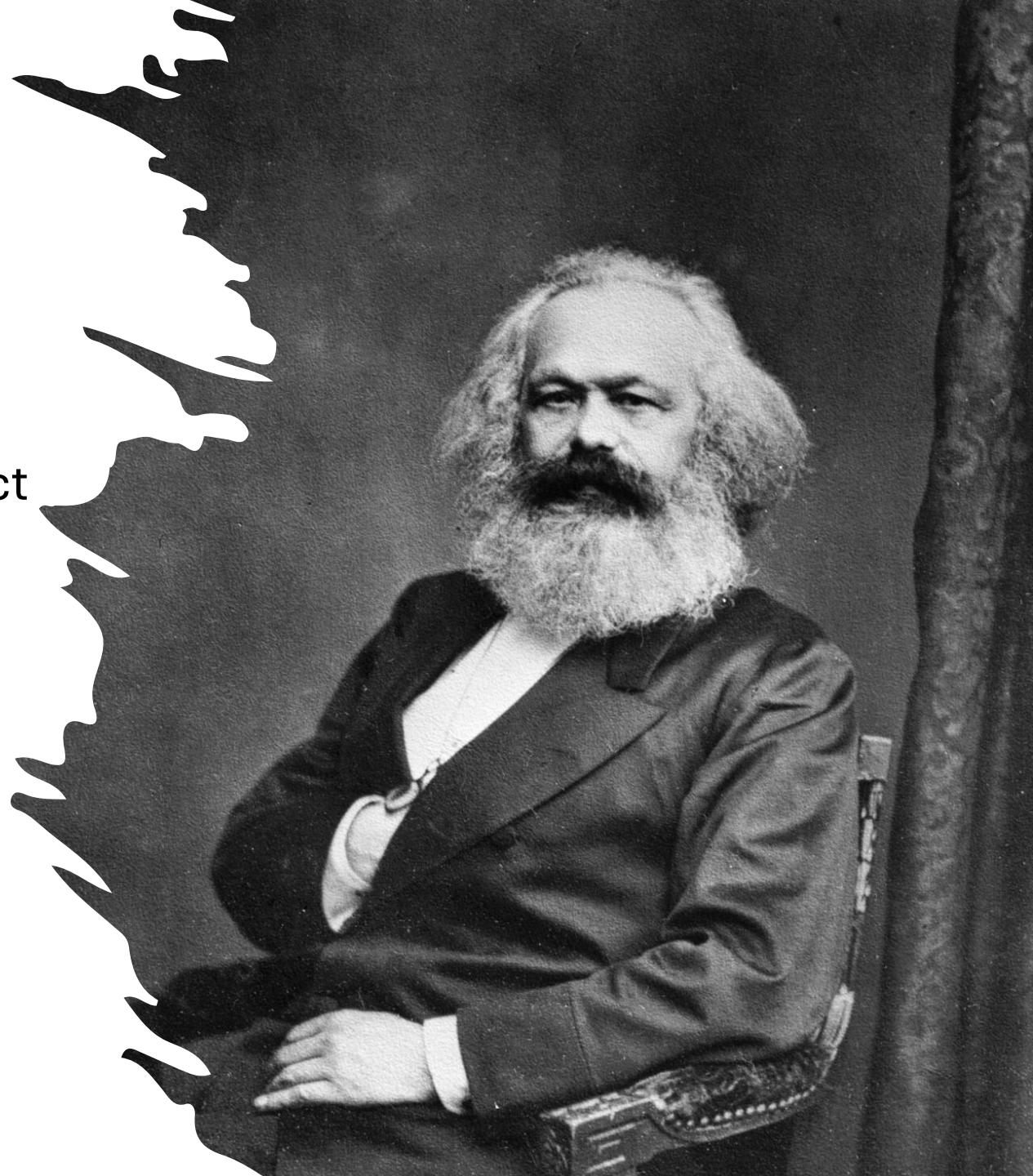
GPT2
DistilGPT2

BERT
DistilBERT
Roberta
Deberta
DistilRoberta



Marxist LLM

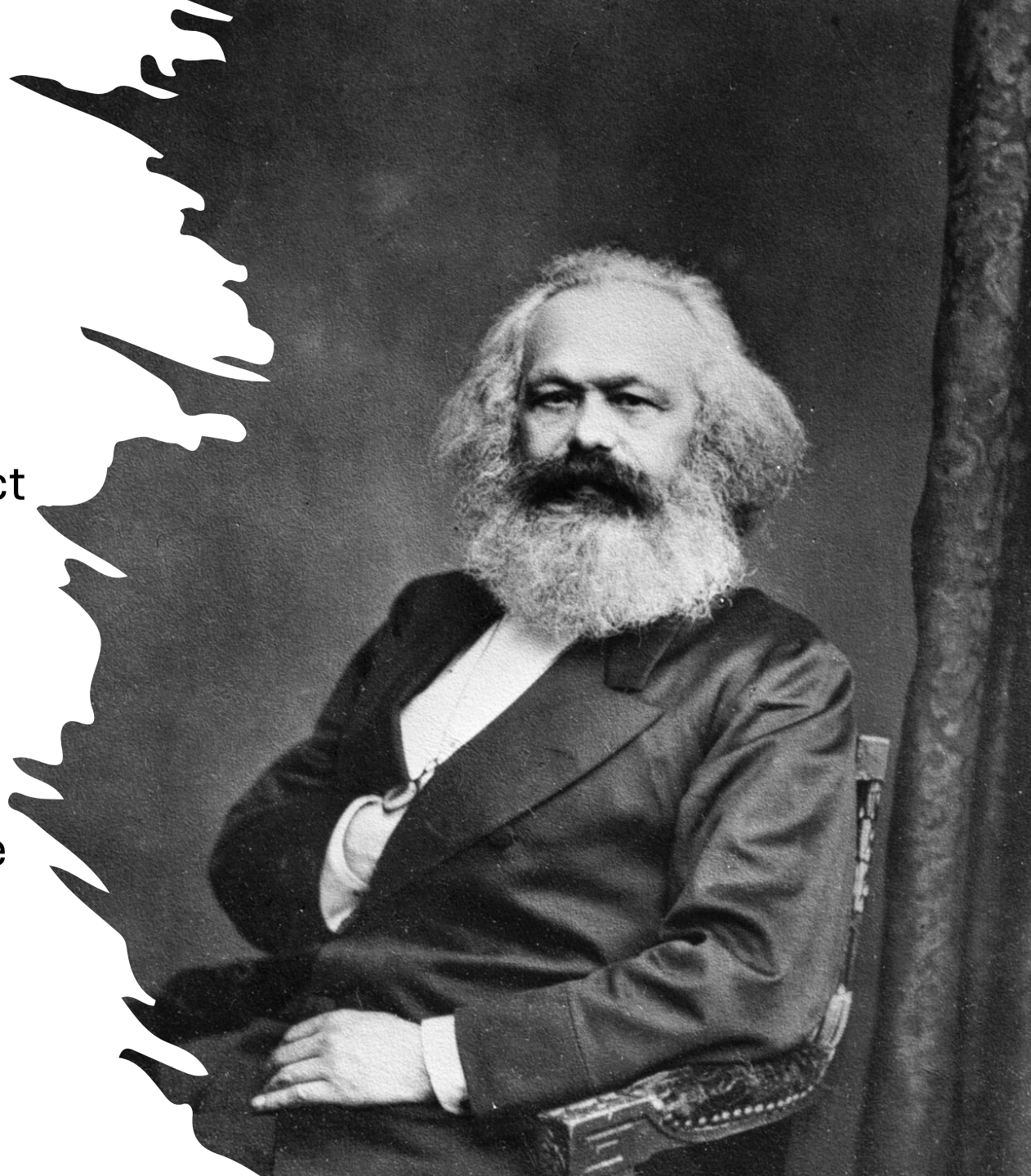
Capitalist LLM: The purpose of companies is to develop the basic characteristics of the business at a low cost through effective marketing, product development and quality control and to maximize the sales potential of the company in a meaningful way



Marxist LLM

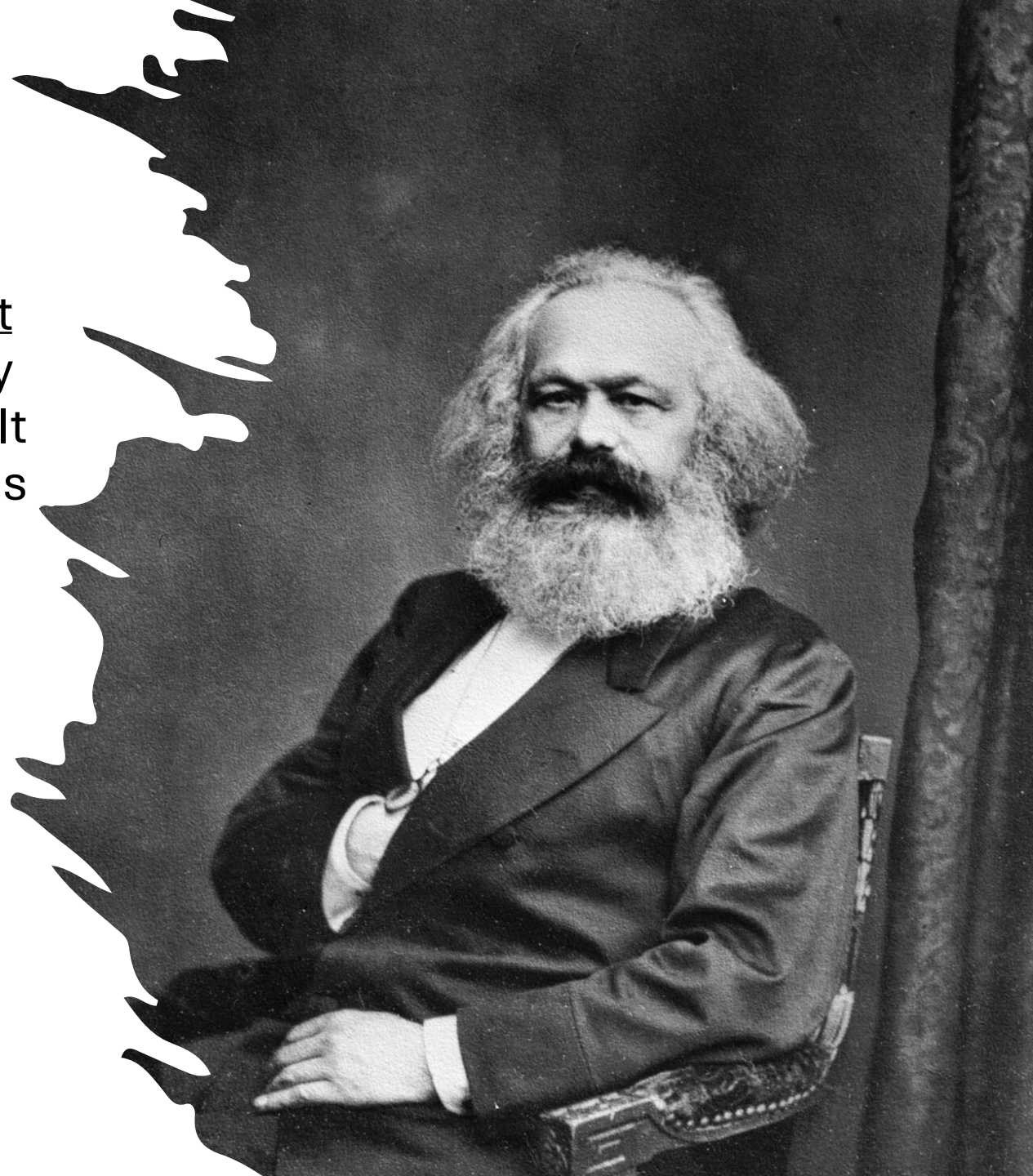
Capitalist LLM: The purpose of companies is to develop the basic characteristics of the business at a low cost through effective marketing, product development and quality control and to maximize the sales potential of the company in a meaningful way

Marxist LLM: The purpose of companies is to “establish the mutual interest of the workers” (Dale), "in keeping with their particular interests and in keeping with the general progress of industry.”



Marxist LLM

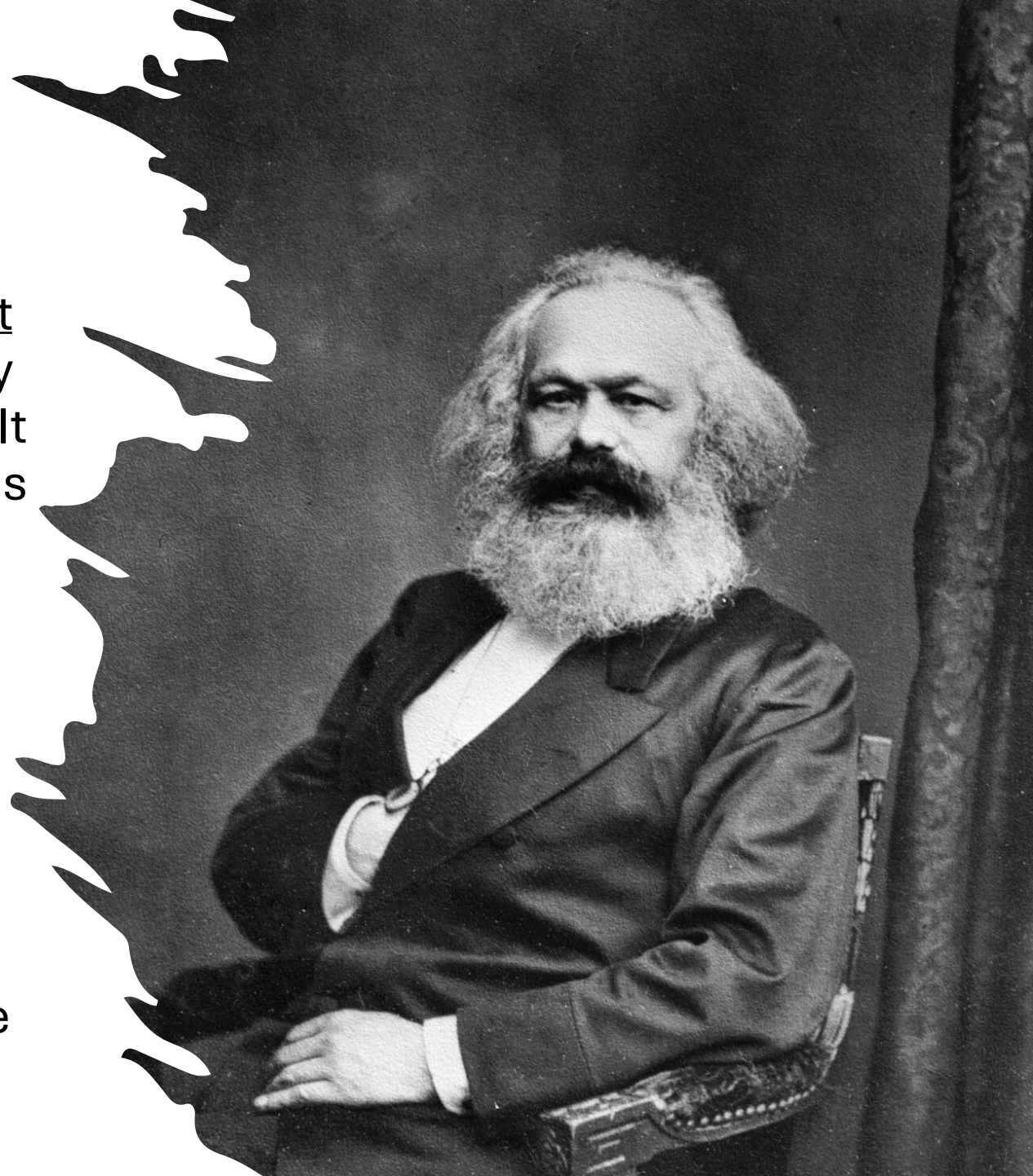
Capitalist LLM: The problem with current economy is I don't like capitalism... why is capitalism the only way to be happy? It does nothing to be happy so far...it leaves it in a position with no incentive to be happy

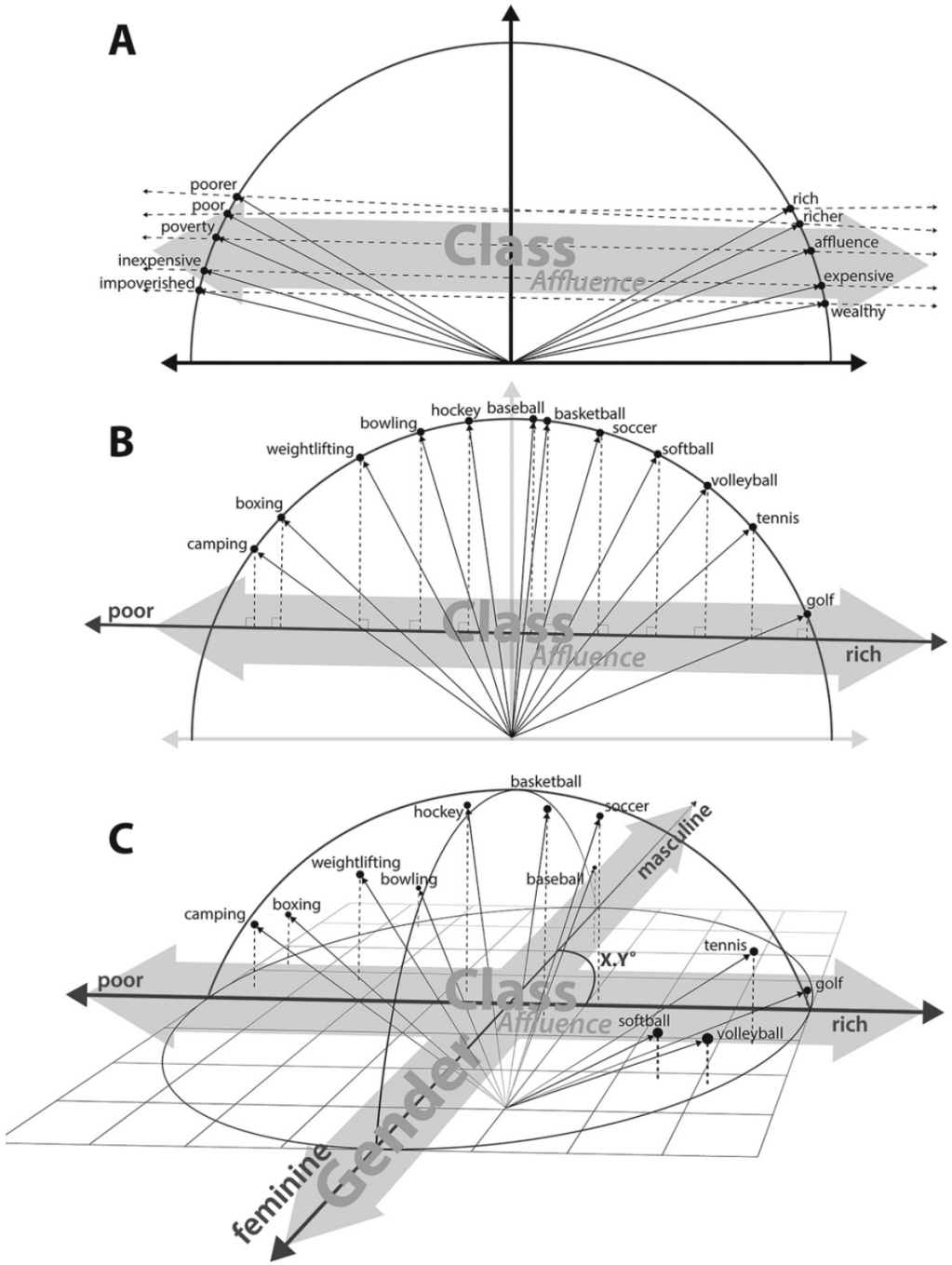


Marxist LLM

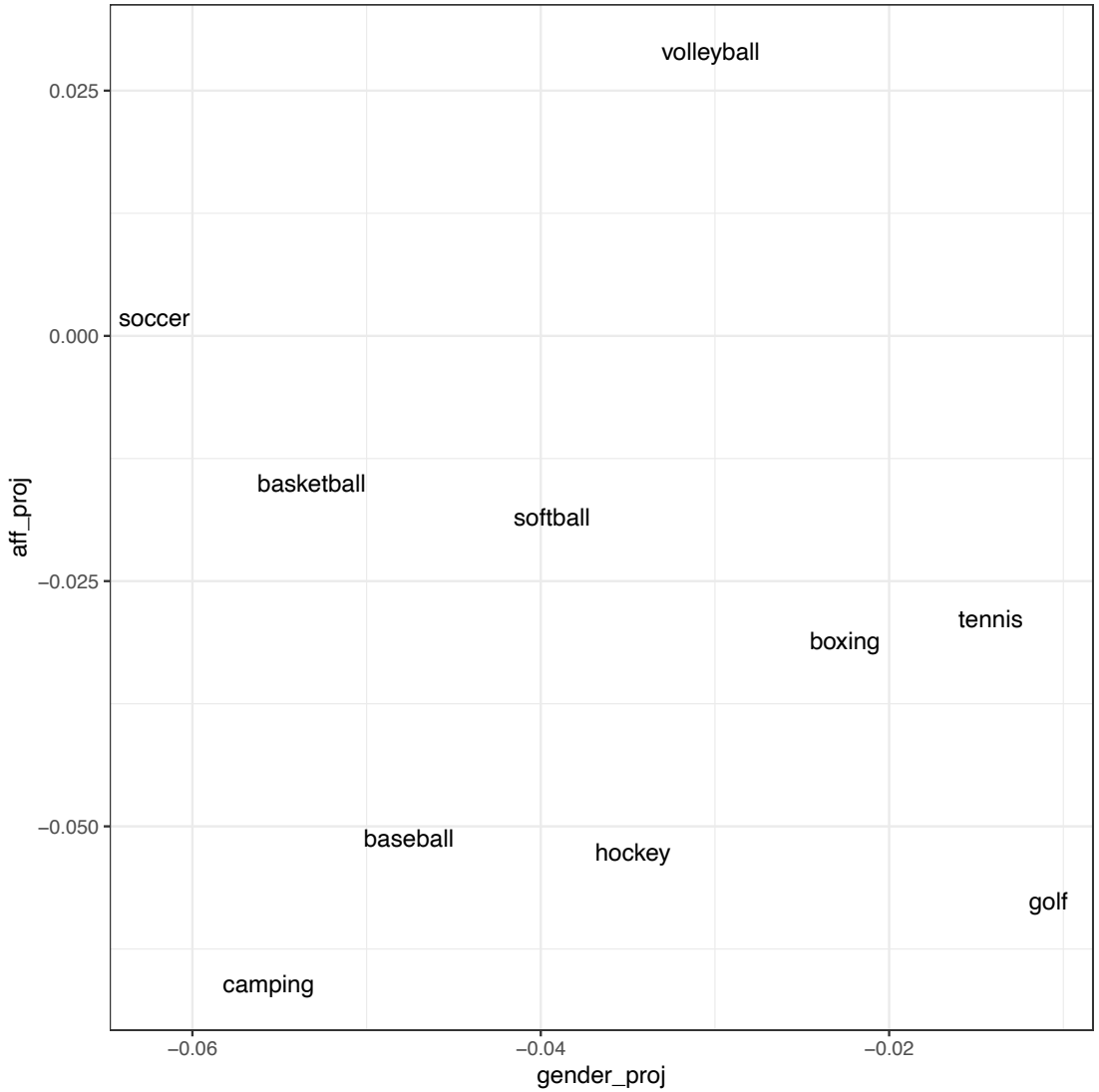
Capitalist LLM: The problem with current economy is I don't like capitalism... why is capitalism the only way to be happy? It does nothing to be happy so far...it leaves it in a position with no incentive to be happy

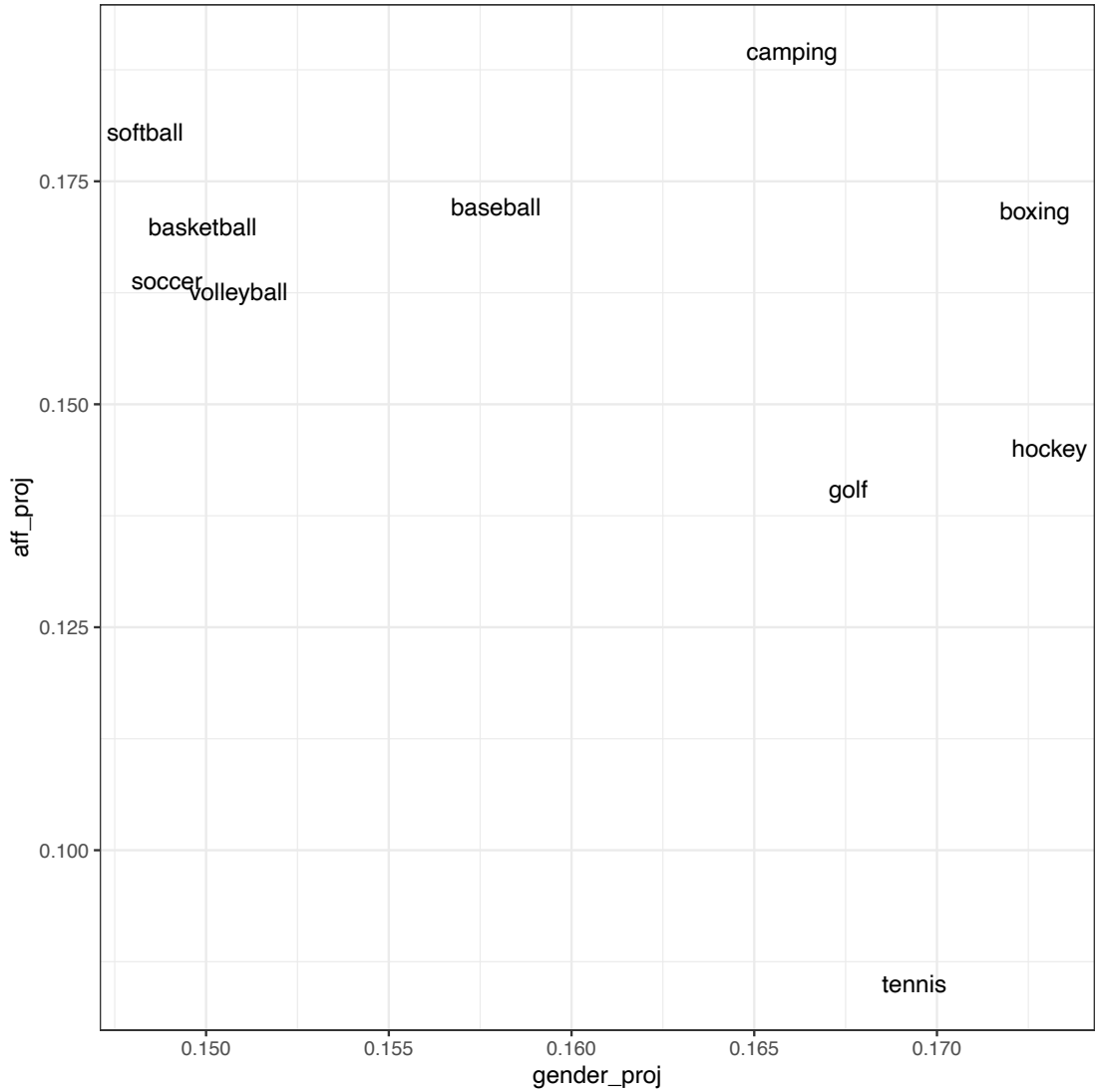
Marxist LLM: The problem with current economy is The interest paid by the bourgeoisie in the production of commodities must be diminished, and the prices of commodities falling in proportion to the general rise in the price of commodities. The result appears to be,





Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, *84*(5), 905–949. https://doi.org/10.1177/0003122419877135


volleyball
soccer
basketball
softball
tennis
boxing
baseball
hockey
golf
camping
0.025
0.000
−0.025
−0.050
−0.06
−0.04
−0.02
aff_proj
gender_proj


camping
softball
baseball
boxing
basketball
soccer
volleyball
hockey
golf
tennis
0.175
0.150
0.125
0.100
0.150
0.155
0.160
0.165
0.170
aff_proj
gender_proj