Faculty of Theology
University of Helsinki

Commercial image recognition representing religion

Anton Berg

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Theology of the University of Helsinki, for public examination in lecture hall PII, Porthania, on 14.12 2024, at 13:15 pm.

Helsinki 2024

Publisher: University of Helsinki
Series: Dissertationes Universitatis Helsingiensis 298/2024

Doctoral dissertation

ISBN 978-952-84-0384-5 (print)
ISBN 978-952-84-0383-8 (online)
ISSN 2954-2898 (print)
ISSN 2954-2952 (online)

Unigrafia
Helsinki 2024

Abstract

The proliferation of AI-driven technologies is the latest step in our era of datafication that also necessitates a renewed exploration of the interplay between technology and religion. In this dissertation, the dynamic intersection between recognition technologies and religion, particularly within the milieu of commercial image recognition (CIR) systems, is critically and empirically examined. Connecting to the previous scholarly waves of digital religion research this work complements previous research by proposing a shift in perspective to one which: 1) examines issues of representation related to AI technologies like CIR, such as those related to race, ethnicity, and gender 2) centers on these media technologies themselves, as opposed to the users or communities they are supposed to serve, and 3) while identifying recognition technologies as a promising novel tool for SSH research on visual data, calls for robust examination of their reliability.

In media and religious studies there has been research on the representation of religion and the way different groups take over digital technologies, but not from the perspective of AI driven recognition technologies, and not in any systematic way from a digital-harms perspective. Recognition technologies are no longer restricted to niche computer vision applications. They are deeply woven into the fabric of our daily digital interactions, governing various facets of societal operations. Prominent among these recognition technologies are CIR systems developed by technological companies Google, Amazon and Microsoft, which not only offer promising new research and commercial avenues, but also present profound concerns surrounding biases and representational harms, such as mis- and nonrecognition, racist and gendered stereotypes, perpetuation of oppressive power hierarchies, and other potential discriminatory outcomes.

The notion of *bias,* in the context of AI ethics, has received much attention and also raised critical discussions. I theoretically anchor my work in previous conceptualizations on *computer system biases* by Friedman & Nissenbaum (1996), and *sociotechnical harms of algorithmic systems* by Shelby et al. (2023). I treat CIR as *algorithmic systems* (Devendorf & Goodman, 2014; Seaver, 2019b), products of human culture that are, among other things, an intertwining of code, human labor and business incentives. The algorithmic systems view is important in relation to bias, since it helps to recognize not only variety in biases (historical, cognitive, cultural, economic), but also that these can end up into a computer system through different ways (ideology, computational architecture design, human implicit cognitive processes, labeling by humans or via automation, economic models).

In this work, the focus on bias is on the *representational biases* of CIR in the context of institutionalized forms of religion. Institutionalized forms of religion as cultural manifestations include, for example, different sets of ideas, beliefs and theologies, hierarchies, ritual forms and behaviors, and artistic and esthetic styles. Many of these aspects of religion are also strongly related to identity, such as ethnicity, race and gender. Recognizing that the construction of religion as a category is an entanglement of many historical, cultural and economic processes, this dissertation does not, however, focus on religion, identity, ethnicity, race or gender as purely academic theoretical constructs. Instead, I explore how CIR actively produces these categories. In doing so, I connect to

the approach of discursive study of religion, which does not focus on any predefined meaning given to religion, but is interested in how different actors, institutions or systems produce religion by giving it certain meanings. Against this analytical background, I use previous theoretizations from cognitive science and religion, media and cultural studies about representation, identity and recognition in a quest for new conceptual insights and searching for answers to my research questions:

> *How do the CIR systems of Google (Cloud Vision), Amazon (Rekognition), and Microsoft (Azure Computer Vision) construct representations of religion?*
>
> *What do we learn of CIR, their processes, and implications by employing theories of identity, representation, and recognition?*
>
> *What can be said about algorithmic bias in the context of CIR, when analyzing it from the representational and methodological levels?*

Methodologically, the study employs an interdisciplinary and mixed-methods approach, bridging computational social science methods with qualitative methods.

Findings of this dissertation underscore the urgency for understanding the ways in which AI-driven recognition technologies shape our perceptions of religion. The main finding is that especially the identity feature of race seems to produce representational biases in CIR systems in the context of religion. By interrogating the discursive visuality in CIR systems – the ways in which these systems interpret, construct and circulate certain meanings of religion and race based on visual data – this study expands critical AI research of how image recognition technologies are not just neutral tools but embedded in them are values and assumptions that shape their recognition and categorization processes. This research emphasizes the need for continual oversight, critical evaluation, and reflexive methodological adaptation in the rapidly evolving field of digital religion research.

Acknowledgments

A dissertation project does not come into being in solitude. Many people have given support and helped me along the way, many of these people besides being exceptional scholars also becoming dear friends.

My deepest gratitude goes first to my supervisors, Professor in Religion Studies and Media Research Katja Valaskivi and University Lecturer Matti Nelimarkka. It is thanks to Katja's supervision and guidance that I have begun to understand critical media studies perspectives and theorizing and learnt how to apply it in research. From our discussions in the course of this process, I have learned more than can be described here. More than once our discussions injected stimulating ideas and perspectives on some aspect of my research, giving me critical insights and pushing me forward. I doubt I would have ended up doing this dissertation at all without your encouragement and support, let alone finishing it.

I had the privilege of working with Matti as a research assistant on studying commercial image recognition systems already before I applied to become a doctoral researcher. I consider myself very lucky to have had you as my supervisor. The level of expertise you possess in computational methods and social theorizing is awe-inspiring, to say the least.

I am also very fortunate to have had outstanding pre-examiners, both highly acknowledged by their scientific merits and standards. I am grateful to professors Pauline Cheong and Heidi Campbell for pointing out both the weaknesses and the strengths of my work. I am also fortunate and delighted that Associate Professor Sofia Sjö has given of both her time and expertise by agreeing to act as the opponent for my defence.

I want to thank from the deepest depths of my heart the researcher community at the Helsinki Institute for Social Sciences and Humanities (HSSH). I feel privileged having had the opportunity to work in HSSH for the whole length of my dissertation process. Without fail you have shared with me the burdens, joys, and sorrows of life and learning to do research and making sense of academia. To this I owe first and foremost to Professor and director Risto Kunelius and vice director Pekka Mäkelä. Both of you are not only inspiring academics, but also exceptional human beings, always willing to offer help if needed. I want to thank also Matti Pohjonen, Jouni Tuominen, Heta Moustgaard, Sointu Leikas, Mari Siiroinen, Eetu Mäkelä, Emilia

Palonen, Visa-Jani Salonen, Pentti Henttonen, Raul Hakli, Juho Pääkkönen, Dayei Oh, and Fiza Vasudeva. I have learned so much from our conversations and doing science with some of you.

A very special thanks is due to Professor Johanna Sumiala. I will be ever grateful for the opportunity to help with your After life-project, to learn from you, and begin new routes towards exciting research avenues on digital death.

I would also like to extend my heartfelt gratitude to researchers Niko Pyrhönen, Gwenaelle Bauvois, and Peter Holley. Together, we shared an unforgettable summer in 2023 as visiting scholars at Stanford University. You are a uniquely talented and insightful group, and I am privileged to be able to call you friends.

I deeply appreciate the remarkable connections and ongoing support from the doctoral researcher community in the study of religion at the University of Helsinki. Alongside my wonderful peers, I especially want to thank Professors Titus Hjelm and Terhi Utriainen for their invaluable guidance.

A very special thanks is due to Associate Professor Aku Visala and Professor Risto Uro, who were part of my dissertation projects Thesis committee. I will be ever grateful for having your support.

I also want to acknowledge the wonderful people I've had the pleasure of meeting through the Helsinki Social Computing Group and the Rajapinta association. Special thanks go to Jesse Pitkänen and Salla-Maria Laaksonen—I truly admire your work.

I have also had the privilege of meeting and working with many remarkable individuals from the Department of Digital Humanities, Cognitive Science. In this regard, I would especially like to thank Professor Riikka Möttönen, researcher Jami Pekkanen, and University Lecturer Anna-Mari Wallenberg for their inspiration and supporting my journey.

I would also like to recognize Professor Simon Lindgren from Umeå University. Your passion for pushing boundaries in the pursuit of innovative approaches to social science and computational methods is truly inspiring and contagious.

Finally, I want to thank my family for shaping me into who I am today. I am on the path to becoming the first in my immediate family to earn a doctoral degree. However, it deeply saddens me that my beloved mother, Airi, could not witness this milestone due to her sudden illness and passing. I know you are proud of me. I would never have had the courage to pursue an academic journey without the love, empathy, resilience, work ethic, and humility you instilled in me. I will always love you, beyond time and space.

Most of all, I want to thank my wife, Susanna, for supporting me as I have travelled this journey. You have always been there for me, in sickness and in health, amid successes and failures. I am extremely lucky and grateful to be able to share my life with you.

To my son, Oliver, and to my daughter, Ruut – nothing compares to you.

Helsinki, November 2024
Anton Berg

List of abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
cf.	confer
CIR	Commercial Image Recognition
etc.	et cetera
e.g.	exempli gratia
ibid.	ibīdem
i.e.	id est
LLM	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing

List of original publications

This dissertation is based on the following publications:

I Berg, A., & Valaskivi, K. (2022). Datafied religion: How well do commercial image recognition services recognize religion? Originally published in Finnish at *Teologinen Aikakauskirja, 128(2),* 174–200. http://hdl.handle.net/10138/35818

II Berg, A., & Valaskivi, K. (2023). Representational silence and racial biases in commercial image recognition services in the context of religion. In *Handbook of Critical Studies of Artificial Intelligence*. Edit. Simon Lindgren. US: Edward Elgar.

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

The publications are referred to in the text by their roman numerals.

Declaration of Authorship in Co-Authored Publications

Publication I: Anton Berg as the primary author, Katja Valaskivi as secondary contributor.

Publication II: Anton Berg as the primary author, Katja Valaskivi as secondary contributor.

Publication III: Anton Berg as the primary author, Matti Nelimarkka as secondary contributor.

Foreword

To offer a little contextualization on how this project emerged and get a sense of the problematic dimensions of image recognition, let us start by looking at three examples.

One day in the beginning of July 2023, a recent MIT graduate Rona Wang decided to create a LinkedIn profile picture of herself by using a commercial AI portrait generator provided by a company named Playground AI. She uploaded a picture of herself smiling and wearing a red MIT shirt to the software and asked it to turn the image into a "professional LinkedIn profile photo". The result was that the commercial portrait generator morphed Wang, an Asian American, into a Caucasian, white and blue-eyed girl[1]. The example demonstrates how commercial generative AI systems, embedded with image recognition, can perpetuate biased perspectives regarding identity characteristics, such as race. The algorithmic framework of Playground AI has obviously been predominantly trained on datasets comprising images of white individuals occupying particular professional roles or environments.

Figure 1 Rona Wang's Twitter post, where she showed the results for the query "turn the image into a professional LinkedIn photo"

[1] See also the related newspaper Boston Globe's story: Buell, 2023.

The second example touches the problem of misrepresentation from a developer perspective, and it worked as an initial inspiration for this dissertation. Some years ago, as a cognitive science undergraduate student, my task for a computer vision class assignment was to develop and assess a rudimentary (deep) neural network model which, after being trained with images depicting dogs and cats in various situations and settings, was anticipated to accurately categorize these animals into their respective binary classes. I chose to evaluate the model's ability to generalize and its precision with a sample image that it had certainly not encountered previously. I introduced an image to my model that featured what appeared to be a female drummer in a marching band playing a suspended bass drum, donning sunglasses, and large comical attachable pink bunny ears (see Figure 2).

Figure 2 Example of the image I used for testing the DL model. Image from Google Images.

I was convinced that the model would stumble, or at least hesitate, in classifying this image and the prediction confidence being low, but to my astonishment, it confidently classified the image as a "cat" with over 98% probability. This demonstrates how a system trained to perform some narrow recognition task, might distort and misrepresent the context and what is worse, do this with great confidence. This got me thinking. Could this kind of misrepresentation happen in the context of religion? Would the possible errors be systematic, and what exactly about religion would make them happen?

The third example concerns religion and identity and connects to a topic that I examined already in my religious studies masters. Latest investigative reports and interviews made by human rights organizations have revealed that China has been using AI recognition technology for the repression of its Muslim minority, the Uighurs (Mozur, 2019; Human Rights Watch, 2018; 2019). Chinese technology companies, such as Huawei and Megvii, have been found patenting technology specifically designed for identifying Uighur Muslims from the Han majority by using race, ethnicity and other religious features of the Uighurs as classification reference for recognition technology (Kharpal, 2021. See Figure 3.). This kind of machinic profiling, raising much ethical concern, exemplifies that recognition of religion and its related identity features via computer vision is becoming mainstream in oppressive regimes.

These three examples highlight how misrepresentations, biases and recognition processes in AI driven image recognition systems can have far-reaching consequences, affecting both identity and perception across diverse contexts. These cases also highlight the representational challenges possibly faced by commercial image recognition (CIR) systems, inviting research and reflection on how these technologies might similarly misinterpret, misrecognize and/or marginalize identities with religious characteristics.

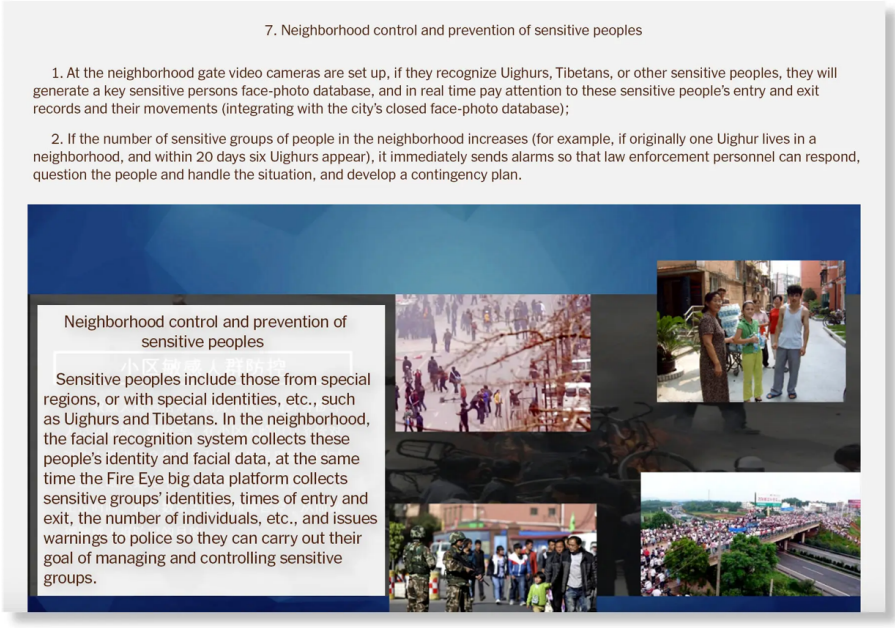


Figure 3: An example of how a Chinese technological start-up, CloudWalk, uses recognition technology for religious identification. Image: The New York Times (Kharpal, 2021).

1 Introduction

1.1 Establishing digital religion research on CIR

We are not anymore the only ones negotiating and defining what "religion" or "spiritual" is, means – or looks like. More and more our meaning making processes are being aided and influenced by AI technologies, perceived by many as intelligent, capable and neutral tools. However, beyond the seemingly human-like capabilities also lie similar human-like cognitive and cultural distortions, and imbalances of power. These tools are not inherently objective or neutral "blank slates", but instead come bearing the often biased preconceptualizations of either the data they have been trained on, the views and beliefs of those who design and implement them, or – as is usually the case – by a mixture of both.

Considering that very limited attention, if any, has been given to research on (commercial) computer vision systems in media and/or religious studies, this dissertation investigates how Commercial Image Recognition (CIR) systems developed by Google, Microsoft, and Amazon recognize and represent religion. A conceptual basis of my work is, that CIR systems are understood as *algorithmic systems*, which integrate human labor, code, data, and values and beliefs such as economic incentives (Devendorf & Goodman 2014; Seaver 2019b). They are not, at the present moment at least, fully autonomous technical entities with a will of their own. Instead they are human-designed (technical) products, that primarily utilize machine learning (ML) and deep learning (DL) techniques[2]. CIR systems are engineered to detect and classify objects, features, or activities in digital images, involving processes such as feature detection, extraction, and matching. This enables the systems to distinguish different elements based on characteristics like shape, texture, color, and contextual cues. Designed for various industry

[2] A description of DL is given in chapter 1.2. At this point it is enough to grasp that DL is an iterative machine learning (ML) technique that is based on a layered algorithmic architecture. Besides the input and output layer, there are layers that are trying to recognize different features from an image (lines, orientations etc.). The output layer is producing a prediction based on the summed values from these preceding layers. As a final stage, this prediction is fed back to the input layer, and the "learning" process starts again until the prediction is "good enough".

applications, including security, healthcare, and different usages for media industries (such as types of social media), these commercial systems offer functionalities like facial recognition, object detection, and scene categorization. As these systems evolve, they benefit from larger training datasets and advancements in computational resources and modeling techniques. Additionally, CIR technologies are increasingly used as research tools, converting images and videos into textual data for analysis with methods such as natural language processing (NLP) (Webb-Williams et al., 2020; Nelimarkka, 2022).

Commercial Image Recognition (CIR) transcends its utility in business or academia as a novel form of media. Its global influence is reshaping the perception and interpretation of religion as part of the processes of datafication (Couldry & Hepp 2017; Lindgren 2019). This perspective aligns with the field of *digital religion studies*, which has risen to prominence in the fields of media, religion, and culture studies (Campbell 2010; 2012(a); 2012(b); Cheong 2013; Hoover & Echchaibi, 2014; Lundby 2013; Siuda 2021; Sumiala 2022).

In their 2020 publication, Heidi Campbell and Giulia Evolvi delineate four historical 'waves' of digital religion research, tracing its origins to the mid-1990s with the examination of online religious practices and communities. Initial studies focused on network theory and its social implications, later shifting towards the authenticity of digital practices, and subsequently to the Internet's integration into daily life. The current, fourth wave emphasizes interplay between online and offline spaces, with a focus on the existential, ethical, and political dimensions of digital religion. Importantly, while Campbell & Evolvi (2020) underline the importance of previous theoretizations of mediation (Martin-Barbero 1993, Hoover 2006; Meyer 2010) and mediatization (Couldry & Hepp 2013; Hjarvard & Lövheim 2012; Lundby 2014) for the study of certain religious phenomena of the digital media age, they also recognize that there has not been enough focus on 'issues related to digital religion', such as those related to race, class, ethnicity and sexuality. Exceptions to this are studies on the impact of gender on media consumption related to religion (e.g., Van Rees & Van Eijck, 2003; Mullikin, 2006; Lövheim & Lundmark, 2019).

In the area of religion and media studies, there is a very solid body of previous research on the media visibility of religion, which has, among other things, examined how religion is represented by journalism and other media genres (De Vries & Weber, 2001; Hokka et al., 2013; Lundby, 2014; Lundby et al., 2018). In the study of digital religion the focus, however, has been in what was initially called the 'new digital media', referring to different online spaces and social media (Campbell & Tsuria, 2021). This research has focused on for example representation and religious identity formation in the "blogosphere" (Bellar & Cambell, 2022), in which by crafting digital narratives, religious bloggers not only articulate their religious beliefs and practices but also intertwine them with elements of their offline existence (Cheong et al., 2008; Campbell, 2010).

Previous works have also illuminated the profound impact of technologies, such as mobile apps, on religious identity. For example, Bellar (2017) observed in studying Evangelical Christians how engagement with religious mobile applications not only mirrored one's faith, but also became a formative experience in itself. According to Bellar (2017), digital interactions, whether positive or negative, have the potential to affirm or challenge perceptions of religious authenticity and authority, thereby shaping an individual's spiritual trajectory and self-conception.

Media representations of religion have also gained significant scholarly attention in the study of media and religion in the Scandinavian context, where most significantly Stig Hjarvard's work has been instrumental in developing the mediatization theory. Hjarvard introduced mediatization theory as a framework alongside the traditional "mediation of religion" approach to better understand the transformations of religion due to media influence (Hjarvard, 2008; 2011; 2013; 2014). This theory shifts focus from direct media effects to how media shapes the conditions of religious practices and beliefs (see also Lundby, 2014; Couldry & Hepp, 2013).

The most predominant approach to mediatization of religion in the Nordic setting has been the institutional perspective. Lövheim & Hjarvard (2019) articulate how mediatization often involves a focus on how institutionalized media influence societal domains like religion, stressing the importance of recognizing the reciprocal influence between media and religious institutions. As an addition, some scholars of media and religion have also focused on the complex relationship between media, religion and nonreligion using mediatization theory (e.g., Taira, 2014; 2022).

A recurring theme in early digital religion studies in relation to representation and identity, is the question of authenticity within digital religious spaces. For example, Hutchings (2010) emphasizes the significance of the digital design mirroring offline sanctuaries in shaping and affirming authenticity. Concurrently, other explorations elucidate the challenges of delineating "truthful representation" in an inherently ambiguous digital realm, emphasizing the continual negotiation of authenticity also within online communities (Whitehead, 2015). Lövheim & Lundmark (2019) have recently emphasized the continuous adaptation and reenactment of religious identities across diverse settings, also acknowledging the challenges posed by social media dynamics. Although digital religion research has emphasized that new digital platforms and modes of communication can empower people in exploring new levels of agency and help them better understand themselves as spiritual beings, other research has also brought to our attention the negative effects these can have on identity (e.g., Elsayed, 2021; Safdar et al., 2023).

My research is connected with digital religion studies, particularly its third and fourth waves, but seeks to also develop the field in four distinct ways. First, CIR

intersects with the issues related to race, class, gender, ethnicity and sexuality in the broader discussions on representation, identity, recognition, and power. From these issues, I will focus on ethnicity, race and gender.

Secondly, the evolving digital religion research has not explored CIR as a novel media with a move toward a conceptualization of culture in which machines have become intertwined with human systems of meaning-making (Natale & Guzman, 2022). Algorithmic systems such as CIR are human-made cultural and technological products that produce, mediate, communicate and circulate representations and meanings (Guzman, 2018; Guzman & Lewis, 2020; Hepp, 2020; Natale, 2021). Natale and Guzman (2022) have recently requested for such a sharpened cultural studies focus on all AI technologies. AI technologies such as CIR both construct and mediate representations, and people interpret them as agents both through this discursiveness[3] and the discourses constructed from them (Neff & Nagy, 2016), particularly in terms of their ability to take over mediating and curating roles previously performed by humans (Natalie & Guzman , 2022; see also Gaw, 2021; Zhang, 2022). In previous decades, traditional media, such as newspapers, had power to construct and circulate certain understanding through the representational choices in the pictures. Today this power has become usable on a new scale with CIR technologies, that are embedded in globally shared technologies such as smart phones, related apps, and other digital platforms. If the datafication of our societies keeps on increasing, as it probably will, it follows that AI driven systems, such as recognition technologies, will increase their influence in taking part of constructing social reality, and having an impact on how humans envision themselves and others. This is true also for the imagination and conceptualization of religion, impacting the understandings of religious individual and group identities.

Thirdly, while digital religion research has predominantly focused on the media usage of individual and group (mainly Abrahamic religions) communities, the advent of CIR technologies necessitates a broader examination that includes the technological infrastructures themselves (cf. Sumiala, 2023). This shift in focus underscores the transformative impact of CIR on the conceptualization and categorization of religion in the digital age.

Fourth, this work methodologically complements previous digital religion research by focusing particularly on the reliability of CIR as a computational method.

---

[3] This aspect of CIR as a media that produces and circulates understandings of religion in the representations it outputs, bridges my research with the discursive study of religion approach (Hjelm, 2021; Taira, 2022). I will introduce this approach with greater detail in chapter 1.5. At this point, it is however important to understand that in a discursive study of religion approach to CIR, the core aim is not to start by giving religion some predefined theoretical definition, but instead by looking at how the category is being constructed in the classifications of CIR systems.

1.2 On the basic operating principles of machine vision and deep learning

Modern recognition technologies such as CIR are based on machine learning (ML) and especially its subfield called *deep learning* (DL) (Chollet, 2022). Machine learning can be broadly described in three types: supervised, unsupervised and reinforcement learning (e.g. Christian, 2020; Hardt & Recht, 2022). A supervised system learns from pre-classified labeled data. In the unsupervised mode, a system has to understand the data itself to build a classification. In reinforcement learning the system learns through interaction with a given environment and through rewards and punishments.

DL is the driving force of contemporary computer vision-based recognition technologies. When given enough data to learn, these models efficiently find statistical regularities and correlations in the data (Chollet, 2022). A basic DL artificial neural network architecture has an input layer, some hidden layers, and an output layer. The input layer receives observation values; in the case of images, these are some representation of pixel values. Figure 4 represents an object in a binary pixel value form.

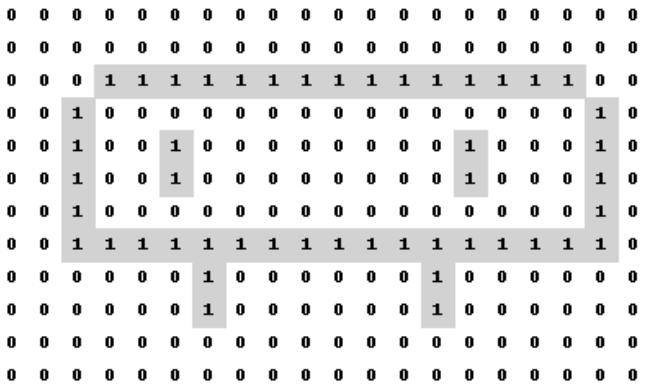


Figure 4 A simple binary pixel representation of an object

The "seeing" of a neural network doing image recognition is an algorithmic processes of trying to produce an understanding of the lines, boundaries, segments and colors of the object(s) in the image. Although the example in Figure 4 is a crude simplification, it works in giving a basic intuition of the problem domain that is involved in image recognition: what does the image actually represent? The figure drawn with ones could be understood to represent, for example, a chair, a computer

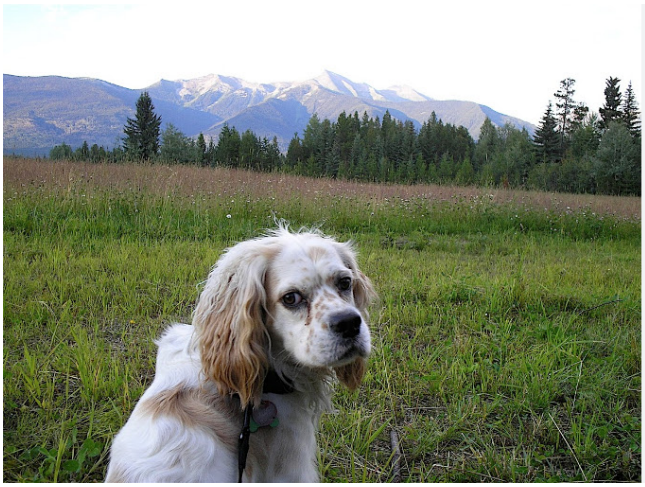
game character, an obstacle or a bus, among other things. In a DL neural network, this meaning would be the meanings (labels) attached to the images given in training by annotating them.

1.2.1 On the annotation processes of CIR and their usage through APIs

Image recognition services from three of the world's leading US technology companies and cloud service providers, Google Cloud Vision, Amazon Rekognition, and Microsoft Azure Computer Vision, are widely used in research and a range of commercial applications. The system architectures and specific algorithms associated with these services are different, as is the data used to train the services for classification tasks. As the system architectures and algorithms are proprietary software owned by these commercial companies, almost the only way to study their operation and differences between them is to feed the classification services with data and examine the results of the classifications statistically and/or qualitatively. Another way would be to try to reverse engineer the architectures, but this would be not only a very time-consuming and technically demanding process, but also could not ultimately guarantee that the modeled result would be exactly the same as the original system. For the purposes of this study, describing the architecture level in too much depth is not relevant. An important aspect is, that the image recognition systems of all three companies use DL, such as convolutional or recurrent neural networks (LeCun et al., 1998; Krizhevsky et al., 2017; Chollet, 2022). An important aspect is also that their operation depends on the training data with which they are trained on. In training their models, the companies use manually labeled data as well as automatic processes[4].

Application programming interfaces (API) are ways of interacting with CIR systems. In this dissertation, through executing programming API calls, images for all articles were labeled by the CIR systems of Google, Amazon and Microsoft, and their classification tags and corresponding accuracies ("score") in percentages (the probability by which the system thinks it sees what it thinks is sees in the image) were collected and stored (See figure 5).

[4] Google's deep learning architect, Francois Chollet, has for example in a recent technology-related podcast (The Lex Fridman Podcast, episode 120: *Francois Chollet: Measures of Intelligence)* explained, how they tried to better the Google's image recognition system by using automated labeling. According to Chollet, this process constituted of scraping user generated images and their descriptions, which were then fed to the model. However, Collet also argued that human annotated data was still providing better end results.




```
{
  "responses": [
    {
      "labelAnnotations": [
        {
          "mid": "/m/0bt9lr",
          "description": "dog",
          "score": 0.97346616
        },
        {
          "mid": "/m/09686",
          "description": "vertebrate",
          "score": 0.85700572
        },
        {
          "mid": "/m/01pm38",
          "description": "clumber spaniel",
          "score": 0.84881884
        },
        {
          "mid": "/m/04rky",
          "description": "mammal",
          "score": 0.847575
        },
        {
          "mid": "/m/02wbgd",
          "description": "english cocker spaniel",
          "score": 0.75829375
        }
      ]
    }
  ]
}
```


Figure 5 Example of Google Cloud Vision API and the output (in Json form)

1.3 Research aims

This study examines representations of religion generated by three Commercial Image Recognition (CIR) systems: Google Cloud Vision, Microsoft Azure Computer Vision, and Amazon Rekognition. By gathering and analyzing the descriptive tags (also called *labels*) and their corresponding accuracy percentages provided by these systems for images about religion, the research employs computational social science methodologies alongside qualitative analysis. Concurrently, this research assesses the methodological soundness of these CIR systems for academic inquiry, while also developing and proposing tools and strategies to enhance their reliability.

The dataset for exploring religious representations consists of 1,188 images sourced from Google Images in July 2022. It is important to understand that Google Images is a different service (image database), than the CIR system of Google (Cloud Vision). To evaluate the methodological reliability of CIR, the study utilized a dataset comprising 26,230 images, which incorporated data from previous research (Pentzold et al. 2018; Hokka & Nelimarkka 2020; Thelwall et al. 2016) and also additional images collected from Twitter and Reddit (the data and criteria for collecting it is described more in detail in chapter 3).

Theoretically anchored in the foundational frameworks of computer system biases as delineated by Friedman & Nissenbaum (1996) and the sociotechnical harms of algorithmic systems as discussed by Shelby et al. (2023), this study comparatively examines representational biases of CIR. By integrating theories from religion, media, and cultural studies that address representation, recognition, and identity, this study analyses how the three studied CIR systems construct,

represent and circulate meanings of religion. With a more accurate phrasing and using the methodological terms of choice, my work is guided by these three questions:

> 1. How do the CIR systems of Google, Amazon, and Microsoft construct representations of religion?
> 2. What do we learn of CIR, their processes and implications by employing theories of identity, representation and recognition?
> 3. What can be said about algorithmic bias in the context of CIR, when analyzing it on representational and methodological levels?

1.4 On the category of religion in the context of this study

This study adopts the 'discursive study of religion' perspective(s) articulated recently by religious studies scholars such as Titus Hjelm (e.g. Hjelm 2021) and Teemu Taira (Taira 2022). In his 2022 book, *Taking 'Religion' Seriously – Essays on the Discursive Study of Religion,* Teemu Taira explores possible means of studying the category of religion in a discursive manner, 'without applying scholarly definitions of religion' (Taira 2022, 2). Taira's overall aim is to show, 'how to conduct interesting and socially relevant research when the category of religion, rather than religion as a phenomenon, is the primary object of study' (Taira 2022, 2). According to Taira, when applying this kind of a discursive study of religion approach, the primary focus is on concrete cases where religiosity has and is being negotiated by many social actors in different institutional contexts, 'rather than on debating the strengths and weaknesses of existing scholarly definitions of religion' (Taira 2022, 2). Taira also writes that:

> 'the methodological premise is that nothing is religion or nonreligion as such, but that some societies organize themselves by using the category of "religion" and they have multiple means by which classifying something as religious is stabilized and made effective. They slice the world into the religious domain and what is outside that (nonreligious secular). What gets to count as religion is ultimately a question of power and the outcome of sometimes complex and contested negotiations' (Taira 2022, 2).

In recent decades, discourse analysis and discursiveness also as an epistemological stance have become widely popular within the arts, humanities, and social sciences, also influencing the study of religion. Discourse analysis can be narrowly defined as a method for examining language in use (Johnstone, 2017). Conversely, it is sometimes comprehended as encompassing all theoretical and

methodological frameworks that employ discourse as a central concept (Hoikkala, 1990; Jokinen et al., 1999). A widely recognized point of reference is the work of Michel Foucault (1926–1984). For Foucault, discourse attached to the concept of ideology, and to the ways how power is being exercised by using language. This kind of language use was "ideological" for Foucault, since it entailed a 'proper' way of thinking about and doing things, which were constructed from a particular perspective (see also Chouliaraki & Fairclough 1999: 26). Hjelm (2021) uses as an example of such a demonstration of power the phrase 'Muslim terrorists', to make the case of how certain characteristics of a group of people can be represented as derivable from their ethnic or religious background (Hjelm 2021, 14). According to Hjelm, this relates to the notion of hegemony (or 'hegemonic discourse'), which can be understood as 'the peak of ideology, the point when all alternative constructions are suppressed in favor of one dominating view' (Hjelm 2021, 14).

Foucault saw that discourse must not be referred to some distant presence of its origin, but instead treated as and when it occurs (Foucault 2002, 28). When this kind of idea is brought to the realm of study of religion, the crucial aspect of discourse becomes grasping the temporal and dynamic way how it is being constructed. From a Foucauldian perspective, this is when discourse becomes the force capable of framing social reality in specific ways, into certain kinds of representations. Discourse constitute subjects, agency and action (Barker 2011, 200). From the interest of this study, certain societal actors, such as the media and global big tech companies under scrutiny, can be seen as possessing greater influence over dominant discourses than for example citizens or minority groups, thus greatly shaping our social reality. Titus Hjelm points to this characteristic of discourse, connecting to the viewpoint of interests of social actors. According to Hjelm, in addition to being constitutive, discourse however also has a double function. Hjelm reminds us that discourse itself is a form of social practice, 'contributing both to the reproduction of society and to social change. Discourse should thus be understood as 'action oriented' and 'performative' (Hjelm 2021, 3, see also Edwards & Potter 1992; Butler 1990).

Although this study finds the perspectives on discourse discussed above broadly applicable, it attaches more concretely on Stuart Hall's (1997) definition (influenced by Foucault). Hall described discourses as frameworks through which knowledge about a specific topic or practice is constructed, comprising a constellation of ideas, images, and practices that shape how we discuss, understand, and engage with particular social phenomena. Since the outputs of CIR are textual by nature, and these descriptions function in representing, producing and supporting certain narratives of religion, this kind of discursive reading of them is a warranted methodological approach.

To summarize, by adopting a similar approach as Taira (2022) and Hjelm (2011), the term religion is perceived in this work as a floating signifier: historically,

socially, and culturally constructed, and negotiated in diverse situations. As Taira (2022) underscores, this kind of discursive study of religion approach does not prescribe a single method, but instead offers a flexible theoretical framework. Taira reminds us, that even Foucault's groundbreaking discursive work, *The Archaeology of Knowledge* (2002, orig. 1969), isn't so much a manual for performing a discursive study as it is a methodological reflection of his preceding works, marking a departure from his structuralist roots (Taira, 2022). This kind of perspective, which refrains from constraining discursive perspectives to a solitary method, is also embraced by Hjelm (2021). In his 2021 book, *Discourse Analysis,* Hjelm emphasizes how every discourse-analytical study needs to be designed individually. Attaching to these discursive research strands, this study seeks to bring about new knowledge of how the processes of CIR relate to the use of discursive power via the representations they produce of religion. This form of discourse analysis is methodologically valuable, because of its special focus on the functioning of discourse in a novel context. What words do in text, is different than what they do in relation to images, and especially in the machinic interpretations and classifications of those images.

1.5 Remarks on the category of race

In this dissertation, I approach the notion of 'race' (similarly as 'religion'), from the viewpoints of discourse. Contrary to widespread popular perceptions, race is not only something biological or genetic, but understandings of race are also socially constructed. While variations exist in genetic pools, it is well recognized that these variations are as prominent within the so-called "races" as they are between them (e.g. Ylikoski & Kokkonen, 2019, White, 2021). Therefore, from the analytical perspective of this study, race is best used as a discursive category, not as a biological one. As such, it organizes systems of representation (Hall, 2002), and social practices that employ loosely defined and diverse physical traits. However, race (or religion) are not just academic abstractions. Through societal implications and lived experiences they are also undeniably tangible. In other words, discourses and cultural meanings have corporeal and material consequences. My perspective on race underscores this multifaceted nature of reality based on context (Storm, 2021).

1.6 Contemplating my own position

Within the sociopolitical context of the early 21st century, the utility of various strands of scholarship is also being critically examined in relation to exclusionary practices of oppression. Debates surround the visibility of many marginalized groups (Ahmed, 2017; Haraway, 2019; Nash, 2019). Any research that deals with

questions related to topics of representation such as race, gender, ethnicity and identity, should also critically contemplate the role of research in contemporary social and political structures.

In relation to research on algorithmic bias, computer scientist Timrit Gebru has recently criticized AI researchers by stating that they 'should learn about the ways in which their technology is being used, question the direction institutions are moving in, and engage with other disciplines to learn from their approaches' (Gebru, 2020). I try to answer to this call by bringing an interdisciplinary approach that combines computational and qualitative methods into studying CIR. Gebru also raised concerns of researchers doing "parachute science", which "lands on a research topic" and because of this outsider perspective ignores or doesn't understand the context of the subjects they are studying. This leads to the researcher unable to 'lift the voices of those who are marginalized' (Gebru, 2020, 10).

This dissertation aims to bring a novel aspect to digital religion studies by shedding light on recognition processes of CIR. The research process follows a systematic interdisciplinary analysis of the research material and their contexts, which calls for reflecting upon my own position. I acknowledge my entanglement in privileged structures, and the fact that I benefit from making this academic work. I also want to highlight my outsider position in relation to many aspects of the research material, as a white, male, non-native English speaker and a Finn studying a sociotechnical phenomenon that specifically effect e.g., the Black community. I have lived my whole live in a Scandinavian setting, and my scholarly expertise is in cognitive science, study of religions, and digital humanities. I see my situatedness into Northern European context also as a strength, since it gives some distance (and perhaps objectivity) to many of the contestations and turbulences of the so-called culture wars in other parts of the world, especially in Northern America.

I did not "parachute" into researching CIR, but instead was initially an eager student of computer vision technologies. It was during this learning process that I found myself critically examining recognition technologies, puzzled with not just what these systems many times "spitted out", but also of the amount of either awe and worship, or loathing and hatred, these systems seemed to gather in and outside academia.

The American professor of feminist studies, Donna Haraway argues, that by understanding vision as *situated objectivity*, knowledge production can write from the body: "(feminist) objectivity is about limited location and situated knowledge, not about transcendence and splitting of subject and object. It allows us to become answerable for what we learn to see" (Haraway, 1988, 582–584). This is an acknowledgment of the context of learning as a precondition to be able to see. Haraway does not seem to argue for the superiority of knowledge possessed by the "subjugated". Instead, she acknowledges situated knowledge as a good alternative

for both relativism and single-minded objectivity. The politics of this kind of research suggests to '[...] partial, locatable, critical knowledge sustaining the possibility of webs of connections called solidarity in politics and shared conversations in epistemology' (Haraway, 1988, 590–596).

The accumulation of knowledge in my being has happened in interaction with the one world we all have and share, and inside certain academic disciplines. Following the kind of intersectional principles Haraway outlines, because situated in this kind of bodily phenotype, ethnographic location, and scholarly traditions, I have come to enjoy many privileges and joys (e.g., I have not been discriminated because of skin color, I have enjoyed free education and health care), but also felt oppressive mechanisms and experienced suffering (e.g., I have felt the consequences of class, I have gone through intensive periods of school bullying and lost both of my parents relatively young).

Although I recognize many of the inequalities and oppressive structures of my local setting, being surrounded by turbulence and contestation inevitably brings its own questions of biases and subjective orientations. Every research position is always subjective and bound to perspectivism(s) – regardless of discipline, methodology, skin color or ideological preferences. This is why I wholeheartedly attach to the kind of situated objectivity, that enables critical knowledge but avoids futile comparisons of whose subjective experiences ultimately produce truer knowledge. I also shy away from academic obsessions that strive to make grand universal generalizations. Instead, I modestly hope to make some contributions to the kind of situated knowledge production Haraway eloquently described, that could benefit us all by identifying connections across discourses and all living beings.

However, I also want to reflect a little on the hardships of doing interdisciplinary research. The biggest grievance for this work were disciplinary differences in defining some key concepts, such as 'representation'. In cognitive science, representation is used when analyzing things concerning an information processing system, and the ways in which knowledge is represented in the system (such as neurons in the human brain or in a network comprised of those neurons). In media, cultural and religious studies, representation relates closely to power structures, focusing on the question of who is being represented by whom and how.

In the case of CIR, both of these views on representation are important. Since CIR as a technology is based on DL neural network models influenced by the neuroscientific models of our cognition and brain, CIR also relates to questions of how semantic meanings are being mapped or represented in the structures or connections of the brain or some cognitive system. However, CIR systems are also corporate products that partake in the social construction of meanings through the representations they produce. From the perspective of cultural studies, the role of

representation is central. CIR systems, through representations, produce images of the world that are also linked to power and politics.

Another problem area for my work, was the sometimes-heated debates evolving around academic dichotomies concerning epistemological views. In cognitive science, a more realist epistemology can be warranted, since cognitive science(s) is mostly oriented towards natural sciences, where focus is on observable and measurable research objects and formalized models. Although the humanistic approach is not (or at least should not be) about rejecting the scientific approach, but about taking a different approach to knowledge production, sometimes dialogue between the disciplines is hard in practice because of these different perceptions. In the humanities in general and study of religion in particular, these occasional problems might partly strum from the general objection towards reducing human social behavior to mathematical models, or theoretizations that treat social phenomena or related features as being something "out there" in some ontological sense. I will dwell on my epistemological principles more in chapter 3.2 where I discuss methodology for this dissertation.

In my attempt to push through the haystack of academic debates, disciplinary boundaries, conflicting conceptualizations, and different explanatory models, I have found much solace in the concept of *explanatory pluralism* (e.g., Visala, 2016; Uro, 2016). According to Visala (2023), explanatory pluralism is a framework for inter-level cooperation between sciences, not a form of reduction or elimination. Similar "loose" philosophical frameworks, that make interdisciplinary research possible, have been recently formulated in the realm of computational social science. In his book *Data Theory,* the Swedish sociologist Simon Lindgren argues that if modern sociology aims to maintain both an interpretative sociological framework for analysis and to utilize computational methods necessary to understand the complex social realities of datafied societies, it needs to move beyond prevailing notions of theory and methodology (Lindgren, 2020). For Lindgren, this "going beyond" means taking a somewhat anarchistic approach of 'rethinking and repurposing research methods, as well as the role of social theory' (Lindgren, 2020, 2–3). Lindgren draws analogies to the field of software development, where '*theories*, as well as *methods*, should be seen (and treated) as open source: free for all to share, alter, and transform' (Lindgren, 2020, 4, emphasis added). In this dissertation project my aim has been to follow Lindgren's 'anarchistic' approach and the spirit of explanatory pluralism.

1.7 Overview of the research articles

This doctoral dissertation is composed of two articles and a book chapter, published in peer-reviewed journals (I & III), and in an academic handbook (II). A longer

summary of the results of these works with a further discussion can be found in Chapter 4.

Article I, "Datafied religion: How well do commercial image recognition services recognize religion?", originally published in Finnish at the journal *Teologinen Aikakauskirja (2022, vol 128, issue 2, p.* 174–200), was co-authored with Katja Valaskivi. In it we explore the relationship between technology and religion by empirically examining how CIR services of Google, Microsoft and Amazon classify images about religion. Since no previous interdisciplinary attempts of studying these systems had been made either in digital religion studies, or more generally in religious studies, the article pioneered in constructing a working mixed methods toolkit, which comprised of a combination of computational social science methods, qualitative methods and theories related to fields of religious and media studies. As a result, the article identified potential secularist, Christian, commercial, colonial, gender, and racial biases related to CIR.

Article II, "Representational silence and racial biases in commercial image recognition services in the context of religion", originally published in the *Handbook of Critical Studies of Artificial Intelligence* (2023, edit. Simon Lindgren, pub. Edward Elgar), was also co-authored with Katja Valaskivi. It describes the part CIR plays in contemporary imaginary meaning making processes on Christianity and race. The study extends the work started in article I by empirically investigating the classification practices of CIR systems of Google, Microsoft and Amazon in the context of race and gender. We conducted an analysis on the outputs of these systems drawing from theories of deep mediatization and datafication. These theories posit that in contemporary societies where sociality is profoundly saturated by technology and its mediated practices, AI driven computer vision systems such as CIR have become one essential part in the construction of meaning. In the article we elaborate how CIR influence and shape epistemologies, categories of understanding, beliefs, values, norms, and attitudes. The article highlights that due to the black box nature of AI technologies, social processes become opaque. Therefore, it is important to have more empirical research that sheds light on these processes for users of these technologies. We reflect our findings on CIR representations of religion in media, cultural, and religious studies with previous and consistent findings of stereotypical representations of race (and gender) that perpetuate unequal power relations.

Article III, "Do you see what I see? Measuring semantic differences in image-recognition services' outputs", originally published in the *Journal of the Association for Information Science and Technology,* was co-authored with Matti Nelimarkka. In it we examine questions of methodological reliability and validity related to the academic usage of the same CIR systems of Google, Microsoft and Amazon that were studied in article I and II. This study was in many ways crucial to conduct, since the work in articles I and II were done using and comparing these

systems and there we already concerns about issues related to bias factors and low reliability of automated commercial image-taggers in general, leading to some critical voices against research employing them (Webb-Williams, 2020). However, no systematic way of addressing these problems has been previously provided. In the article we suggest mitigation strategies for overcoming some of the problems. Our general finding was that researchers should never rely on using only one CIR system, since there can be high variability in their classifications. As a computational strategy, we introduced a novel cross-service label agreement tool (COSLAB) that can measure the semantic differences between two services outputs by utilizing natural language processing (NLP) and machine learning (ML) techniques. Our tool is designed to help future researchers when considering using CIR for social scientific research, or when comparing the CIR services while using big data.

2 Theoretical and Conceptual Background

2.1 On the choosing criteria for the main theories and concepts, and the disciplinary lens

I will next introduce the theories and concepts used in this dissertation. Since this study's main focus is to gain understanding of how the CIR systems of Google, Amazon, and Microsoft represent religion, and since I already explained how I approach the notion of "religion" in chapter 1.5, I will begin by delving on the notion of *representation*. I will focus on the conceptualization(s) of representation used in cultural studies, which have also been widely adapted in religious studies.

After discussing representation, I will move onto discussing *recognition* and *identity***.** Lastly, the third main concept of this study is *bias*. When introducing the conceptualizations on algorithmic system bias, I will focus on representational and methodological biases, since these align with my research aims and can be analyzed with the data and methods used in this study.

2.2 Representation

The representations that CIR systems produce take part in influencing our cultural systems (see also Geertz, 1966). Cultural systems can be understood as a "stream" where every new human being is born into, thus also inevitably adapting certain cultural structures of expectation (Ross, 1975; Tanner, 1979; Karvonen, 1992; Valaskivi, 1995; Archer, 2004). Origins of the concept can be traced to ideas from cognitive science: structures of expectation are the elements of our cognition that arrange information and against which new experiences are being predicted and interpreted (Valaskivi, 1995; Clark; 2015; Barret, 2020). They are the semantic relations between concepts and categories that also entail the normative presuppositions that label certain phenomena as "different" or "abnormal," in contrast to others deemed "healthy" or "right." From the perspective of cultural studies, structures of expectation can be understood as representing the competing social definitions over which battles for social power are waged. (Valaskivi, 1995).

Importantly, structures of expectation are not innate properties of our cognitive systems, but instead mediated to the members of a culture through

representations. They are thus temporal, non-fixed and constantly in flux. From a (strong) constructionist perspective, structures of expectation cannot be trusted to provide truthful information about reality outside themselves. Since representations are culturally produced, the question of representation also becomes a political, politicizing activity, permeated by different interests and ideologies (e.g. Rossi, 2010; Hall, 1997). Politics can be in broad terms defined as a "struggle for meanings" and therefore analyzing representation from the perspective of politics is to focus on power relations related to them (Hall, 1997; Rossi, 2010).

For Stuart Hall representations enable not only the understanding and sharing of reality through communication and interpretation, but that these meaning-making signs often have material dimensions (Hall, 1997; 2002). Paintings, photographs and film are all examples of this. Whether we think of representation as presenting or substituting, societies are constantly struggling over who or what images (and words) can represent whom or what. According to Hall, the initiator and producer of a representation process always has a political or ideological stance or agenda – the idea of why some particular signs are important. For Hall, this question of intentionality, the purposes and goals of the makers of pictorial and verbal representations, is an important part of the constructivist framing since it also provides a way for challenging the hegemonic ways of thinking (Hall, 1997).

Hall also underlined the co-produced nature of representations where the initiator of the meaning process, is not the only producer, but meanings are equally produced by interpreters who look at, listen to, and circulate representations. From a methodological perspective, Hall utilized discourse analysis to examine how existing linguistic practices represent a close link between knowledge and power – a *system of representations* (Foucault, 2002; Hall, 1997; de Lauretis, 2004). Fundamental to such systems is, that they sort people, practices, objects, and classifications into different hierarchical value structures (Skeggs & Loveday, 2012; Skeggs, 2015). In these symbolic processes, some combinations of class, gender and skin color are considered as good and valuable, while others are deemed bad or worthless. By examining these discourses, some answers can be obtained to questions about where and how representations operate (Rossi, 2010).

Rossi (2010) points to the dual meaning of the concept of representation. In addition to presenting and describing, representation also refers to substitution – standing for someone or something else. While Rossi is moving more on a general level of abstraction (e.g., how an elected representative can represent the people), Stuart Hall identifies the "being substituted with nothing" as production of "representational silence". Hall is talking about how the absence of certain kinds of people (for Hall, usually Black people) from cultural texts is an event in which a whole part of humanity is "replaced by emptiness". Representational silence is an

important concept for this study, since it can be applied to comparatively analyze the absence of representation, features and/or recognition of some group or phenomenon in recognition technologies such as CIR. Representational silence is discussed in more detail in the following subchapter (2.2.1).

2.2.1 Representational silence

> "It was the silences that told us something; it was what wasn't there. It was what was invisible, what couldn't be put into frame, what was apparently unsayable that we needed to attend to. If you want to ask, "what can content analysis teach you?" well, one of the questions you have to ask is, "what about the people who appear to have no content at all – who are just pure form, just pure, invisible form?"
>
> Stuart Hall (1992; 1997)

Hall's notion of "representational silence" draws attention to the absences or underrepresentation of specific voices, experiences, or narratives within different media portrayals, highlighting the power dynamics and marginalization inherent in these omissions. According to Hall, representational silence occurs when certain groups or issues are systematically excluded or marginalized in media representations (Hall, 1997). In Hall's understanding, by maintaining this kind of "silence", dominant power structures perpetuate marginalization and reinforce existing hierarchies.[5] Hall points out that this exclusion can manifest through various mechanisms, such as the absence of diverse characters in film and television, limited coverage of certain social issues in news media, or the erasure of alternative perspectives in cultural discourses.

Representational silence also connects to Hall's investigations on the ideological and discursive formations of racism(s). Hall is particularly interested in why racial classification systems have persisted beyond the nineteenth century (Hall, 2002). Hall himself works from an explicitly personal diasporic experience, a framework that allows him to make sense of the human categorizing and differentiating in terms of race, ethnicity, and nation. He emphasizes that the terms "race" and "racism" can only be fully understood in the context of broader economic, political and social formations. Hall sees race as a signifier that move up and down the signifying chain, neither fixed by nature, predetermined or constructed outside of discourse. Although there are some general and global mechanism associated with the practices of racism, in each society, racism has a

[5] As a Marxist cultural theorist, Stuart Hall was obviously drawing from Antonio Cramsci' conceptualization of "hegemony", which can be understood as the domination of the ruling class over individuals or other groups by the means of culture, politics, and economy (see also Jones, 2007).

specific history that presents itself in particular and unique ways, and these details influence its dynamic and have real effects that differ from one society to another (Hall, 1992). Therefore, when using racism as an analytical concept, Hall proposes we should move using it in the plural (racisms).

2.2.2 Representation and identity

The cultural studies understanding of representation has strong connections to the notion of "identity". Identity is also widely used in research on religion, such as in social science-oriented Biblical studies, practical- and systematic theology, and religious studies, but for example Jokiranta & Luomanen (2013) argue that its content often remains undefined. I will next introduce some theoretizations of identity, previously used in religious and cultural studies, and compare them with concepts from computer science. Some relevant connections are also drawn to cognitive science(s), since it has influenced social scientifically oriented research such as social psychology, as well as computer science.

Because religion is a social phenomenon, religious studies has been particularly interested in the social identity of communities and groups, rather than just focusing on individuals. However, there are multiple ways to understand the notion of a "group". A group can be for example be identified as a collective, whose members recognize its existence, but which does not require commonality between group members, nor any established boundaries (e.g. Jokiranta & Luomanen, 2013). In addition, a common distinction would also be to make a separation between large groups (e.g. ethnic groups, religious groups, political groups, etc.) and small groups (e.g. families, school classes, sports clubs etc.). This is however problematic, since there are no clear boundaries when a group shifts from being 'small' to being 'large'. Also, there are cases of mixed identities: a person might belong into some national, ethnic, religious or linguistic minority, but also identify strongly with some larger group identity based on these same categories. Professor Abby Day has written about the "believing in belonging", where the choosing of religious identification is complementing the aspect of social belonging and identity (Day, 2011).

In computer science, and in the formal mathematical framework, a group is understood as a collection, a count, or a set of some units or objects. Objects can be any countable entities from different categories such as transistors, rocks, numbers, lines, cells or persons. When a separation between such collections is desired to be made, we enter into the mathematical realm of logic, and more specifically set theory (e.g. Devlin, 2012). The aspect of set theory is important from the viewpoint

of identity, since for a computer vision system, the separation between “identities” happens basically by following the formal rules of set theory.

Set theory serves as a foundational language and instrumental framework for deductive reasoning, underpinning virtually the entirety of mathematical discourse. Its capacity to formalize mathematical concepts renders it an indispensable tool within the realm of applied mathematics. In the context of computer science, set theory offers a robust mechanism for the formalization and logical analysis of computational processes and their respective computational objects (e.g Lin, 2023).

The relevance of set theory to computer vision and recognition technologies can be illustrated through two arguments: Firstly, set theory provides a formal language for representing and manipulating digital images, which are inherently sets of *pixels*. By conceptualizing images as sets, operations such as “union”, “intersection”, and “complement” can be applied to perform image processing tasks like segmentation, where an image is partitioned into sets of pixels corresponding to different objects or regions. Secondly, set theory facilitates the definition of mathematical morphology, a theory and technique for the analysis and processing of geometrical structures. Mathematical morphology operates on sets in a topological and algebraic sense, enabling the extraction of image components that are useful in the representation and description of shape. This aspect of set theory is particularly crucial for tasks such as edge detection, object recognition, and structure analysis in computer vision, where again the spatial arrangement and relationship of pixels are important. These processes of segmentation and shape detection are the technical core of computer vision and CIR in identity recognition. From the engineering perspective, identity recognition is about building a computer vision system to classify (based on pixel representations) individuals or groups by using their extracted visual features (skin color, beard shape, facial morphology, body pose etc.). Non-supervised ML systems try to learn these features by for example utilizing different clustering techniques. In supervised ML, identity related feature labels can be pregiven as output targets when training the system.

Coming back to the realm of social sciences and humanities, an important viewpoint on defining identity is the difference of how an individual or group sees and perceives themselves, and how others see, acknowledge and label/recognize the group. Within religious studies, anthropology, folkloristics, and the social and behavioral sciences, the concepts of "emic" and "etic" often delineate these two distinct vantage points (e.g., Morris et al., 1999; Spiers, 2000). The emic approach encapsulates an insider's viewpoint, wherein for example an examination of the beliefs, values, and practices within a specific culture is conducted through the lens of those residing within that culture. This stance is geared towards comprehending the cultural connotations and importance attributed to particular behaviors or practices, as elucidated by the individuals actively participating in them.

Conversely, the etic approach represents an outsider's standpoint, whereby a culture is scrutinized from the perspective of an external observer, researcher or in the case of CIR, technology. However, in the context of algorithmic systems such as CIR, the issue is more complex, since the data for training the recognition systems could be produced in an "emic" way, although the mechanical process of recognition would be categorized as being "etic".

The construction of identity not just affects the way in which individuals' and groups' form their shared beliefs, attitudes and feelings, but also how their perceptions and actions are oriented. In other words, identity is not just the passive *being* of something, but also about the active *doing*. From this stance, any conceptual separation of personal, individual identity and communal, group identity, becomes problematic. In everyday life, and in most research settings, such a distinction is impossible to make, as personal identity does not stand out in any "pure" form but is instead strongly linked to the social and requires interaction (e.g., Jenkins, 2014).

2.3 Recognition Theories and the Politics of Recognition

Within the field of AI research, especially when the critical focus is on machine vision and image recognition, it is often overlooked that the focal point of the discussion resolves around services encapsulated by the term "recognition" technologies. Theories of recognition, however, are also an established field of social philosophy (Saarinen 2016).

For the Canadian philosopher and political theorist Charles Taylor (1994), recognition, although a vital human need, is tightly related to issues of politics and group dynamics. Taylor has coined the term "the politics of recognition", proposing that the demand for recognition is connected to modern ideas of identity and to the self-understanding of individuals (see also Valaskivi & Sumiala, 2023). The Professor in Political Economy, William Davies, has recently claimed that 'the politics of recognition has acquired more momentum than anyone could have foreseen' (Davies, 2021).

In his influential essay on recognition, Charles Taylor states that contemporary politics must deal with issues of recognition, because 'liberalism can be blind to cultural differences'. For him, the problem of recognition is the tension between equal treatment and universal values (Taylor, 1994). For example, legally protected cultural identities might occasionally generate tension with the overarching principle of universal equality among all citizens. Taylor also emphasizes the significant role of recognition for various politically active interest groups, such as nationalist movements, minority factions, or progressive entities like feminist groups (Saarinen, 2016). Importantly, from this perspective

recognition is the foundational component of dialogical relationships in the participatory formation process of identities.

In his seminal work, social philosopher Axel Honneth (1995) reformulates the Hegelian concept of the 'struggle for recognition'. This framework posits social conflicts as endeavors in which marginalized groups—often distinguished by cultural or religious identities—seek validation and acknowledgment of their rights and status. Honneth's analysis delineates how these struggles serve as a pivotal mechanism for social change and recognition (Honneth, 1995; see also Saarinen, 2016). Employing psychological categorizations such as offense, anger, and emotional pain, Honneth analyses the psychological aftermath of misrecognition. He argues that the catalyst for numerous historical and present social protest movements can be pinpointed to an acute sensation of their identity being misrecognized (Andersen & Honneth, 2005; Honneth & Fraser, 2003).

Unpacking Honneth' conceptualization of the struggles for recognition, it becomes evident that there is an intricate dance between power dynamics and communicative processes. According to Honneth, the battle for recognition is twofold: there is a desire not just for dominance, but also an aspiration for the understanding and affirmation of the others' worldview. This struggle is not only a defensive mechanism for one's principles or values but also a quest to articulate one's position within a societal "space of reasons"— a platform that permits the potential for dialogue and comprehension.

However, Honneth has emphasized, that he does not consider all claims to recognition to be legitimate simply because they spring from "someone's hurt feelings", or "disappointed expectations of recognition" (Honneth & Fraser, 2003). Instead, according to him, claims to recognition made by citizens of a given polity may be justified to the extent that, without this recognition, they would be disadvantaged in their search for a good and autonomous human life. In other words, Honneth' theory consists of an argument about specific forms of recognition without which human life cannot be experienced as successful and free.

There are of course many critical questions and perspectives related to "struggle for recognition" and to the demand of increasing representational diversity or visibility of minority groups. Grossberg (1996) for example criticizes the role identity has been granted in various branches of cultural studies, and the tendency to reduce the complexity of individuals, groups, social relations and societal problems to questions about identity and related contested features such as race, class, gender or sex. He argues that the biggest false assumption related to identity in cultural studies has been, that 'contemporary politics is and should be organized around struggles over identity' (Grossberg 1997, 87). Grossberg sees this kind of model of identity (and difference), where every individual is held to be a representative of the totality, a very recent 'self-defined constituency acting in the interests (for the politics) of that definition' (Grossberg 1997, 87). According to him,

such constituencies do not need to exist, and although 'identity has been – and may still be – the site around which people are struggling', whether this is a fruitful path forward, should be critically examined (Grossberg 1997, 97).

Honneth has conceptualized three forms of inter-subjective recognition: love, equal respect (as citizens endowed with uniform rights), and social esteem (signifying one's distinctive value within the societal fabric). Let us consider the dimension of "respect" in relation to religious representation with an example. Following Honneth' reasoning, in a multicultural society an individual practicing some institutional form of religion or a "looser" belief system, such as Jedism, must recognize the rights of another belief systems, such as Paganism, and vice versa. This mutual recognition, devoid of hierarchical gradients, embodies the principle of equal respect.

In the context of CIR, social esteem is an important aspect of recognition. CIR systems have the power to give cultural significance to representations created in a certain cultural setting, while at the same time unsignifying others. CIR is a global technology that is produced in a particular cultural setting. The recognition, and mis- and non-recognition processes of CIR are being circulated globally, creating problems that relate to their universalizing aspect. In Honneth's conceptualizations of inter-subjective recognition, much value is also given to the element of love. It is extremely difficult to connect this element to recognition technologies, since the recognition they mediate, being machinic, always seems to lack "something" that the property of love brings to "genuine recognition" (e.g. Koskinen, 2019). An added philosophical conceptualization to genuine recognition is the notion of intentionality, which means that recognition-relations are not mere physical or spatial-temporal interpersonal relations, but instead involve specific content or a subjective view or stance. This intentional aspect is to be understood as the "taking as something", which grants the acts of recognition their socio-normative meaning and significance (Koskinen, 2019). To be able to love, would seem to demand this kind of intentionality. It also directly relates to identity: a person cannot be fully recognized, without or in conjunction to some form of "as what" specification. This "aboutness" embodies characteristics and aspects such as markers of religion and identity.

Valaskivi & Sumiala (2023) draw attention to how technological affordances of the social media environment enable and reinforce certain forms of recognition and make others impossible. According to them, this is linked to social media, and the attention economy, which causes the question of recognition to be reduced to mere attention, whereby the seeing required for genuine recognition does not take place. A similar distortion can be observed in the way CIR technologies "recognize" religion, and how they produce different stereotypical and structural biases. In their influential 1996 paper, Bias in Computer Systems, Batya Friedman and Helen Nissenbaum introduced the concept of "decontextualized algorithms" to discuss

especially representational biases. According to online dictionaries, the adjective or transitive verb decontextualized means 'something being shown without a context', or 'being removed from a context' (Merriam-Webster, 2024; Cambridge, 2024). According to these same sources, this "removing" can be understood as detaching something from its original, local context, and then rendering it to some universal idea. I find this kind of critical emphasis on decontextualization useful in the context of CIR. The notion of decontextualization seems to capture well what these systems end up doing when recognizing and representing religion and religious actors from digital images. Images (usually) carry a specific context, which CIR processes decontextualize by removing them from this "original" context and transferring them to another.

Religion is a cultural phenomenon with many culturally specific features. The distinctive individual, collective and cultural markers related to religion are discernible in ethnic, textual, visual, tonal and embodied elements. In other words, religion is not only a recognizable multimodal stimulus for human cognition. Religion can also become recognizable for machinic cognitive representation systems, such as recognition technologies. These distinct feature markers of religion include (but are not limited to) religious texts, symbolic artifacts, clothing, jewelry, and grooming trends. Additionally, religion can be observed in embodied features, activities and configurations such as rituals, ascetic expressions and other bodily poses, and congregational setups. Religion also has specific characteristics of sound, such as types of prayers, hymns and chants. Therefore, religion emerges as a significant categorizing factor also for modern AI-driven recognition technologies, fundamentally designed to detect visual features.

The way CIR systems recognition processes produce meanings of religion – by always removing at least some of the original context and rendering it to some universals – characterizes them as "decontextualizing algorithmic systems" (Friedman & Nissenbaum, 1996). This removal of context from pictorial data presenting religion is important to recognize, since it can also have major effects on the identities of those being attempted to be recognized and represented. It is well acknowledged that religion is a significant psychological and cultural contributor to the formation of individual and group identities (e.g., Ysseldyk, et al., 2010; Oppong, 2013). Thus, as in all issues of media representations, the questions of representation (and mis- and nonrepresentation) become central. For instance, the way these technologies identify members of religious groups may significantly differ from how these individuals perceive their own identities. This unilateral and non-interactive process of recognition and categorization can be viewed as an exercise of oppressive top-down power. Like mentioned in the foreword of this work, a notable instance of such technological oppression is the situation of the Uyghurs, a 12 million ethnic Muslim minority in China. Human rights organizations have reported that the Uyghurs are not only detained and subjected to communist "re-

education centers" but are also under constant technological surveillance due to their religious identity. They are involuntary test subjects for Chinese recognition technologies like facial recognition systems (e.g., Harwell & Dou, 2020; Wakefield, 2021). Prior studies have established that the cultivation of religious identity is pivotal for the self-representation of ethnic collectives, particularly for those in minority or subjugated standings (see e.g., Rodriquez et al., 2000; Ajrouch & Kusow, 2007; Lemon et al., 2023).

2.4 AI Bias

Recent academic works on bias related to AI have explored how biases manifest in different ways in artificial intelligence systems, and the broader implications of these biases for society and technology (e.g., Christian, 2020; Crawford, 2021; Buslón et al, 2023; Tamkin et al., 2023; Li et al., 2023). According to philosopher Mark Coeckelberg, AI bias can be broadly defined as 'a social and ethical problem associated with artificial intelligence that does not apply to other automation technologies' (Coeckelberg 2020). Coeckelberg identifies how AI biases can occur at different stages of design, testing and application, but he seems to reduce the problem of AI bias all down to training data: data used for teaching may not be a representative sample of the population, the algorithm sets up its decisions on the wrong correlations or makes predictions with too little examples, or the data is originally against special groups.

For Coeckelberg, AI biases are unintentional: the consequence of the inability of engineers or corporate management engineers or business management to anticipate or imagine any discriminatory effects on certain groups of the system. However, Coeckelberg's unintentionality argument appears naïve when examining a temporal aspect related to AI bias: the problems of bias have been known for at least a decade in the companies developing the systems, but the problems have not been fixed. Unintentionality as a cause would hold only until the outcomes are unknown. If the outcomes are known, but problems are not fixed one might plausibly argue that a shift from unintentionality to, if not intentionality, to certain positions of ambivalence or negligence has taken place.

According to the French sociologist Jacques Ellul (1912–1994), there is a bias in our own perception and beliefs towards adopting novel technologies, where previous problems associated with technology, and/or brought by their use, are in parallel also thought to be solved by more technology (Ellul, 1964; 1975; 1980). Strong echoes of this kind of "technologist strawman belief" can also be observed actualizing in many places in our modern times. In relation to discriminatory biases of computer vision and image recognition systems, the belief has recently manifested for example in the public lecture of the esteemed computer science professor Geoffrey Hinton, who also worked for Google for many years guiding the

development of the company's AI systems. In a recent lecture Hinton reflects on the problem of discrimination bias in AI stating that, 'if the goal is to make unbiased systems at all', this problem is 'easier to handle' (than for example cybercrime done by using AI), since according to him, it could be easily solved by 'freezing the weights' of a trained model and then 'measuring the bias'. It seems that for Hinton, if discrimination bias of machine learning systems is a problem in the first place – or just something to accept as part of "the natural order of things" (since humans also have biases) – the solution is a simple technological tweak. This leaves no room or need for critical discussion of wider social or political fixes.

Some scholars have also criticized the recent critical focus on the biases of algorithmic systems for creating a sense of unjustified urgency, which according to them distorts public debate and policymaking and has a negative impact on the development of AI technologies. Malte Ziewitz (2016) has for example argued that this is 'algorithmic drama', the invention of alarmist narratives about the negative impact of algorithms. I agree that it is imperative not to disregard the capacities and potential of recognition technologies also to bring about good, such as to assist individuals with disabilities by giving back lost modalities such as sight or to work as a communication tool for also amplifying the voices of marginalized communities. However, it is equally important to recognize that this advancement comes with critical caveats, notably the dilution of individual autonomy as users cede control over their data, succumbing to the subliminal influences and manipulation of algorithmic systems and the algorithmically curated large-scale content and information aggregation that makes it possible to create efficient models for predicting user preferences and behavior. The discourse around terms such as "surveillance capitalism" and "data colonialism", as articulated by scholars such as Couldry and Mejias (2019a, 2019b), Fuchs (2018), and Zuboff (2019), brings to the fore the exploitative dynamics inherent in the datafication processes underpinning algorithmic systems. This critique gains further relevance in light of the technological hegemony exerted by major corporations in shaping the digital landscape and the datafication of societal interactions. Moreover, the recent body of critical work documenting the discriminatory propensities of algorithmic systems underscores the urgency of sustained scholarly attention to these issues (e.g., Buolamwini & Gebru, 2018; Benjamin, 2019; Carrera, 2020; Crawford, 2021). The manifestations of algorithmic bias, ranging from gendered stereotypes to racial discrimination, necessitate a rigorous examination of recognition technologies also in contexts as nuanced as religious representation. Joo & Steinert-Threlkeld (2022) have recently proposed, that neither pre-trained nor commercial image recognition models should be implemented blindfolded without first understanding their inherent biases. However, in order to understand these biases, or to make comparisons between them, there must be some kind of a taxonomy. I will next

introduce the conceptualizations on which this work relies on for understanding AI biases and their related social harms.

2.4.1 The tripartite categorization of computer bias

Friedman & Nissenbaum (1996) identify three different types of biases in computer systems: *pre-existing*, *technical* and *emergent*. They also make the claim towards a standardization of bias, where the freedom from bias together with reliability, accuracy and efficiency should form the set of criteria according to which the quality of any computer system in use in society should be judged.

According to Friedman & Nissenbaum, a (computer) system is biased if it 'assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate'. They also formulate that unfair discrimination alone is not bias unless it occurs systematically, and that 'systematic discrimination does not establish bias unless it is joined with unfair outcome' (Nissenbaum & Friedman 1996, 332–333).

The first type of bias which Friedman & Nissenbaum identify, is the *pre-existing bias*. Like its name suggests, the preexisting bias exists independently and prior to the creation of the system and has roots in the structures and culture of social institutions, manifesting in practices and attitudes. According to Friedman & Nissenbaum, preexisting bias can enter a system either through 'the explicit and conscious efforts of individuals or institutions', or it can crawl in implicitly and unconsciously, 'even in spite of the best of intentions' (Friedman & Nissenbaum 1996, 333). The second type of bias they identify is the *technical bias*, which can arise from 'technical constraints or technical considerations'. Here Friedman & Nissenbaum also introduce the notion of *decontextualized algorithms*. According to them, technical bias originates from the 'use of an algorithm that fails to treat all groups fairly under all significant conditions' (ibid, 334). Although Friedman & Nissenbaum use the example of an algorithm that schedules airplane departures and take-offs, this type of (decontextualizing) algorithmic bias, and especially the associated requirement for fairness, can be easily linked to discussions of the representational biases of recognition technologies. The third and last type of bias Friedman & Nissenbaum identify, is the *emergent bias*. According to them, emergent bias can manifest as the 'result of changing societal knowledge, population, or cultural values' (ibid, 335). Friedman and Nissenbaum argue that user interfaces are extremely prone to this type of bias, since they 'seek to reflect the capacities, character, and habits of prospective users'.

In their article, Friedman & Nissenbaum also offer solutions to overcome the biases. They highlight the importance of improving diagnostics and underline system designers responsibility in taking potential failures into account at a very early stage of design. According to them, this requires a good understanding of what

relevant biases exist in the world. In relation to this they argue that for example minimizing emergent bias asks designers to 'envision not only a system's intended situations of use but to account for increasingly diverse social context of use'. They also ask the hard questions: can corrective measures solve the unfairness of the past, how much diversity in social context is enough, and what sort of diversity? (ibid, 343).

2.4.2 The sociotechnical harms of algorithmic systems

According to Shelby et al. (2023), the computing research community has lacked a 'high level and synthesized overview' of harms done by algorithmic systems. In their article, by reviewing 172 computing research papers, Shelby et al. come up with a systematic taxonomy of sociotechnical harms, which according to them, cluster into five major themes: representational, allocative, quality-of-service, interpersonal harms, and social system/societal harms. Since this dissertation studies representational recognition systems, I will focus on representational harms.

According to Karizat et al. (2021), representational harms of algorithmic systems can be defined as assumptions about people, culture, and experiences, which 'perpetuate normative narratives that adversely shape people's sense of identity and belonging. Representational harms can manifest, when algorithmic systems reinforce the subordination of social groups on the basis of some identity feature, such as disability, gender, race, ethnicity, religion or sexuality (e.g. Barocas et al., 2019). Shelby et al., (2023) also acknowledge the previous critical findings on computer vision, where for example the under-exposure of certain groups has led to unequal visibility (Bennett et al., 2021; Hendricks et al., 2018; Zhao et al., 2017; Benjamin, 2019).

On a deeper analytical level, Shelby et al., (2023) identify six specific dimensions of representational harm:

1) *Stereotyping* involves the reflection of beliefs about specific groups' characteristics, leading to possible over- and under-representation, such as gendered stereotypes (Bivens & Haimson, 2016; Cambre & Kulkarni, 2019; Søndergaard & Hansen, 2018; Paullada et al., 2021; Wang et al., 2020).
2) *Demeaning* social groups occurs when algorithmic systems perpetuate lower status and lack of respect to some individuals or groups. This can happen for example within image tagging, where ranking, and retrieval systems are associated with specific identities (Sweeney, 2013; Noble, 2018; Goff et al., 2008).

3) *Erasure*, distinct from stereotyping, happens when systems systematically fail to recognize and tag people or artifacts linked to specific social groups.
4) *Alienation* occurs in image tagging systems when a person's group membership relevance is ignored. This diminishes human dignity and fosters a sense of not belonging (Katzman et al., 2021; Mannes, 2020; Devos et al., 2022).
5) *Denying people opportunity to self-identify* can manifest when systems classify individuals without consent or through data collection. This type of harm reduces autonomy and typically impacts already marginalized communities (Chancellor et al., 2019; Paullada et al., 2021; Corry et al., 2021; Katzman et al., 2021).
6) *Reifying* essentialist social categories involves reinforcing social categories as "natural", or truths based on some narrow criteria. This type of harm could be identified theoretically also as reflecting algorithmic hegemony (Katzman et al., 2021; Dosono & Semaan, 2020; Keyes, 2018). The perpetuation of Western or Eurocentric perspectives through classification algorithms is a prime example of this type of harm (Barocas et al., 2021; Devinney, 2022).

I will use these six categorizations of representational harm as a framework for analyzing the representational shortcomings of CIR in general. However, in religious studies, the process of reification has been given a special focus in the historical formation process of the category of "religion". Countless scholars have argued that the academic concept of 'religion' was produced through Eurocentric perspectivism and the philosophical, socio-political and economic dynamics of early modernity (see Asad, 1993; Fitzgerald, 2000, 2007(b); Masuzawa, 2005; McCutcheon, 1997; Saler, 2000; von Stuckrad, 2013; Taira, 2022). These critical stances underline that the discourses of religion (and non-religion) significantly contributed to the formation of European and Western self-identity, juxtaposed against the 'Oriental' or 'African' other. These historical meaning making process of the category of religion were also partly mediated by the technology of its time: through printing and circulation of texts, books and the related pictorial elements. In many of these cultural products, other cultures and belief systems were presented, described and analyzed through Western universals, reflecting the spiritual ideas, practices, symbols and agents of the "other" against the familiar ideas of Christianity.

2.4.3 Neutrality bias associated with algorithms

During the past decade, public awareness on discrimination against groups of people by algorithmic decision making has been growing (e.g., Angwin & Larson 2016; Burn-Murdoch 2013; Dwoskin 2015; Nature 2016). Often in these discussions the underlying ethos is that such discrimination is unintentional and unexpected, assuming that algorithms must be inherently objective – "mere tools". However, research has shown that for example predictive models may discriminate against people, even if the computing process is fair and well-intentioned (e.g., Calders & Zliobaite, 2013; Barocas & Selbst, 2016). This is because most of the methods are based upon assumptions that historical datasets are correct, and for example accurately represent groups and population.

These assumptions can lead to biases about intelligence, race, biology, gender, and religion that become embedded into the functioning of algorithmic systems (Crawford, 2021). In their landmark study of classification, Bowker & Star (2000) express how 'classifications are powerful technologies' and when embedded in working infrastructures they become 'relatively invisible without losing any of their power'. Bowker & Star (2000) observe how classification disappears into infrastructure, and in general becomes 'taken for granted'. An example of the bias disappearing into CIR infrastructure is the process of image labeling and dataset curation, that many times requires the use of crowd-sourcing platforms such as MTurk, Clickworker, AppJobber or CrowdTap. Although there is the ethical issue of underpayment of these kinds of "gig workers", there is also a dimension, hardly ever discussed, related to the cultural biases of the human annotators that can be introduced into the systems. Although the technology companies developing image recognition have strict guidelines for annotation processes, I think it would be extremely unlikely that the annotators' own biases wouldn't have any effect.

Recently, Broussard (2023) named as "technochauvinism" the kind of bias that 'considers computational solutions to be superior to all other solutions' (Broussard, 2023, 1:2–3). Morozov (2013) has previously articulated a similar critique of what he terms technological solutionism. He defines this concept as a pervasive ideology that simplifies complex social issues—ranging from politics and public health to education and law enforcement—into well-defined, solvable problems. According to this ideology, these societal challenges can be effectively addressed through clear, computable solutions or can be optimized through the application of precise algorithms. Similarly, Valaskivi (2020) calls this kind of ideological fallacy related to technology as the 'contemporary faith of innovationism'. According to Broussard, from this bias towards technology it follows that although a simple technological "glitch" in some system might be fixable, the ones that produce discrimination, misrepresentation, or racism cannot, because they are inherently structural biases, needing more than just a quick code update (Broussard, 2023).

Critical internet studies scholars Safiya Noble and Ruha Benjamin have also used the notion of glitches to illuminate the ways in how race and technology intersect in pernicious ways. Noble (2018) has written on the famous case where Google tagged photos of Black people as gorillas and how 'algorithmic oppression is not just a glitch in the system but, rather, is fundamental to the operating system of the web'. Benjamin (2021) has also conceptualized what she calls as the "New Jim Code[6]", in which new technologies reproduce and exacerbate historical inequality while being same time portrayed as neutral or progressive. The New Jim Code is a subset of discriminatory design that refers to racial biases embedded in technical systems. On the surface, these systems may seem innovative and efficient, but when you delve deeper, they can perpetuate racial disparities. Benjamin also uses the concept of "discriminatory design" which means designs in public spaces and products that cater to certain groups while excluding others. A primary example she gives are the armrests on park benches designed to prevent homeless people from lying down on them. These designs are a response to larger societal issues but do not address the core problems.

Already David Lyon (2003) critiqued the implementation and effects of facial recognition technology on the vulnerable and marginalized parts of population. This critique was expanded by Magnet (2011) and Browne (2015), who analyzed the dialectics of recognition in biometric technology and the ways it is directed by normative categories of identity. Both Browne and Magnet argue that there is a cultural logic embedded in (biometric) recognition technology. Magnet (2011) argues that the failure of the science of biometrics lies in the fact that it is a technological implementation of 'gendered and racialized norms', essentially codifying 'existing forms of discrimination' and thereby failing to recognize the complexity of identity. This criticism is joined and further developed by Lee-Morrison (2019), who argues that contemporary digital technologies and the infrastructures in which they operate do not present an entirely new world but rather carry within them 'historical continuities of cultural and visual logic'. In relation, Broussard (2023) has argued, that 'every kind of sensor technology, from facial recognition to automatic faucets, tends to work better on light skin than on dark skin'. Broussard eloquently captures the aspects of unintentionality and unconscious in these processes:

> "[..] this process starts with recognizing the role that unconscious bias plays in the technological world. [...] Nor do I think that most people who make technology or software get up in the morning and say, "I think I'll build something to oppress people today." I think that what happens is that technology is often built by a small, homogeneous group of people.

[6] Jim Crow laws were state and local laws introduced in the Southern United States between late 19th and 20th centuries. They enforced racial segregation. "Jim Crow" was a pejorative term for an African American. See David (2000) for more information.

[…] They probably thought, like many engineers, that because they were using sensors and math and electricity, they were making something "neutral".

Broussard (2023, 7)

3 Data and Methods

3.1 On the selection criteria for data

Of my research questions, two are very much empirical: how do the CIR systems of Google, Amazon, and Microsoft construct representations of religion, and what can be said about algorithmic system-, representational- and methodological bias in the context of CIR. Since there was no ready-made image material of images about religion available (for articles I and II), I curated the material myself by using Google Images service[7].

Google Images allows you to search for images using keywords or to identify images using reverse image search. However, this kind of compilation process also meant, that as a researcher I was already defining "religion". However, being aware of this problem of constructing the category of "religion", and thus striving to get as diverse data as possible, I made the decision to first prompt with multiple iterations of test keywords and observe the output images Google Images service produced. Names describing institutionalized forms of religion combined with the word "rituals", seemed to produce the most diverse hits when assessed from the viewpoints of ethnicity, race, gender, geographic location, materiality, and sociality[8] (see Figure 4). I also tested Google Images with different VPN settings, to see how much that effected the images the service offered: not much. Deciding on this criterion, I proceeded to collect a total of 2482 images from the image service using a programmed, automated collector (the programming language was Python). For articles I and II, I chose to keep the number of images in thousands, since I wanted to be able to also conduct a qualitative content analysis for each image. After examining the data, I made the decision to further reduce the dataset to 1188 images. This "cleaning" was needed, since the automated collection also included

[7] See https://images.google.com

[8] Test keywords included: Christianity, Islam, Buddhism, Shintoism, Hinduism, world religions, institutional religions, Christian rite, rites, spirituality, non-religion, esoteric, lived religion etc. However, the name of a religion combined with the word rituals produced a wide variety of images of both ritual forms and different gatherings and services: Christian rituals, Muslim rituals, Hindu rituals, Buddhist rituals, Shinto rituals, and spiritual rituals.

what I would label as "noise" in this particular research context: film- and cosmetics advertisement, book covers and text-only images. I was instead interested in images depicting people, materiality of culture, and the social dimension of religion. A final boundary decision was to manually add images from different religious traditions that had similar spatiality (viewing angles), and ritual settings. This decision was made to increase the overall comparability of the approach. The total dataset size was however kept the same (1188) for articles I and II.

For article II, a focus on the representations of gender and race in the context of Christianity was chosen. I smashed together the previously collected dataset for Article I, with additional data that was collected by using search terms: black Christianity, brown Christianity, black preacher, black female preacher, black Christian women, women priest, female priest, women bishop, female bishop, black bishop, black female bishop, and black women bishop. Again, a total of 1189 images were selected for the final data set and 352 images from this collection were taken and saved in separate folders for further analysis and comparisons. The folders were by category: Black Christianity (96 images), White Christianity (79 images), Black female Christianity (29 images), White female Christianity (29 images), Black male Christianity (61 images), and White male Christianity (58 images). This approach was needed because I was also interested in the statistical distributions of religious recognition.

Figure 6 Examples of the images used in Article I and II.

In article III, the aim was to systematically investigate CIR from the viewpoint of methodological bias. To be able to answer to this question, the services themselves needed to be compared and their outputs evaluated. An aspect of this evaluation was the style of the images, which led to decisions concerning the design and size of the corpora. A categorization of "professionally produced images" (such as images published in newspapers) and of "ordinary user produced images" (such as those shared via social media) was chosen. This guided the hypothesis of whether professionally captured/curated images are more reliably recognized than content closer to the user-origin end of the continuum. Also, to guarantee diversity of topics, images from various contexts, such as journalism, politics, and social media, were included. The overall data (total of 26,230 images) comprised of eight different image corpora (see Table 1). Some of the corpora came from data used in prior studies (Pentzold et al., 2018; Hokka & Nelimarkka, 2020; Thelwall et al., 2016), and the rest was collected from Twitter and Reddit by programming automated scrapers. Figure 5 presents one randomly chosen image from each corpus.

Table 1: Image data used for the analysis of Article III

	Thematically coherent entity	**Random selection of images**
Professionally produced images	A set of 452 images from articles containing the term "big data" found in online archives of The New York Times and The Washington Post, based on datasets of Pentzold et al. (2018).	One month of images in news stories published online by Finnish public broadcaster YLE (12,257 images).
Professionally curated images	188 images shared via Twitter by Finnish party leaders for the month prior to 2019's general election.	872 images from the most recent 1,500 tweets by Finnish political party leaders.
User-curated or user-generated images	A set of 10,372 images shared via Finnish anti-immigration Facebook groups, based on preexisting datasets (Hokka & Nelimarkka, 2020); 413 images collected from Reddit's "aesthetic" forum; and 868 collected from its "memes" forum.	A random selection of 800 images from Twitter users in the United States and the United Kingdom, based on data from Thelwall et al. (2016).

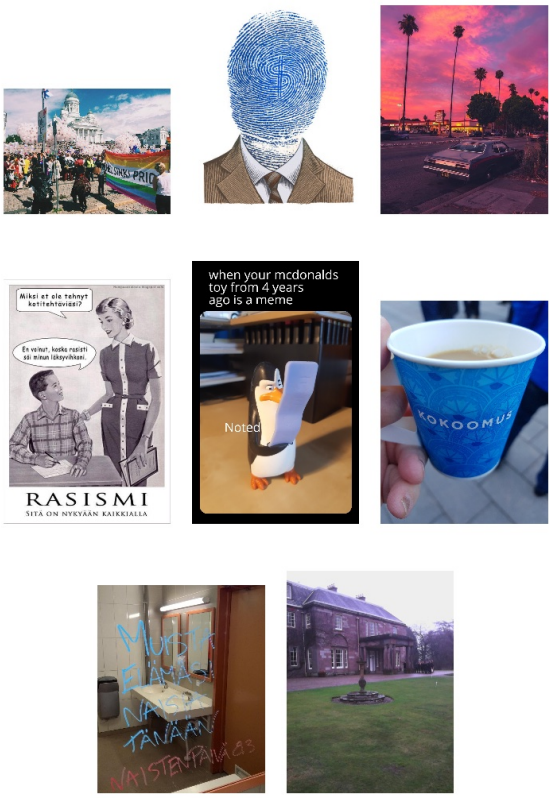


Figure 7: Examples from the image data used for article III, one from each corpus.

3.2 Methodology

Like already mentioned in chapter 1.7, this work employs *explanatory pluralism* articulated recently for example by Visala (2023). However, to clarify my epistemological standing, I adhere towards Peircean epistemology[9], which is grounded in Charles Sanders Peirce's pragmatism and a certain type of realism, which I trust can offer a robust framework for social scientific inquiry that can coexist with constructionist and discursive views of knowledge.

[9] Charles Sanders Peirce (1839–1914) was a prominent American philosopher, logician, mathematician, and scientist, often considered the founder of the philosophical tradition of pragmatism and a pioneer in the field of semiotics. His extensive work contributed significantly to the development of modern logic and his philosophical theories have deeply influenced a wide range of academic disciplines, including the philosophy of science and epistemology.

Peirce's pragmatism, particularly his view that concepts gain their meaning from practical effects, aligns well with the dynamic and culturally embedded processes of social science (Burch, 2024), and more specifically with the discursive study of religion approach (Taira, 2022; Hjelm, 2021). Peircean understanding, where the truth of an idea is determined by its practical consequences has also an important methodological aspect as it insists that hypotheses and theories must be tested and proven useful in real-world scenarios. This kind of pragmatism encourages a focus on outcomes and effects that are observable, measurable, and also impactful in society. However, Peircean realism adds an important dimension by acknowledging that while knowledge is indeed shaped through human practices and cultural contexts, there is still an underlying reality that these practices aim to describe and understand (Burch, 2024). Peirce's realism thus supports the idea that social constructions and cultural phenomena have material consequences and structures, which can be systematically studied and understood.

Following Peircean epistemology and the discursive study of religion, I recognize knowledge about "religion" as being co-produced and constructed through discourse, in this case, human machine-interaction and subjective experience. Accordingly, and following the Peircean "focus on practical outcomes-argument", my methodology integrates an interdisciplinary mixed-methods approach which employs quantitative and qualitative techniques for comprehensive analysis of visual data. The used quantitative methods include CIR and statistical analysis allow comparisons of CIR classifications, and to quantify patterns and correlations in visual data sets. Concurrently, qualitative content analysis provides deeper insights into the contextual and thematic elements of the images and CIR classifications, enabling a more nuanced understanding of the underlying meanings and interpretations. This methodological synergy allows for an analytical approach that respects both the complexity of human culture, the constructed nature of knowledge, as well as the underlying reality all phenomena attach to. I will next explore in detail the concrete methodological choices in the context of the articles that comprise this dissertation.

3.2.1 Visual content analysis

Articles I and II utilize qualitative analysis, that could be more specifically named as a form of "content analysis". Content analysis, explicitly labeled as such or not, entails the creation of generalizations about the comparative frequencies of various visual representations pertaining to classes of people, actions, roles, situations, or events. This procedure inherently involves the classification (and quantification) of media content, whether implicit or explicit. Therefore, it comes as no surprise that content analysis has been one of the key methodologies in the field of media studies for numerous decades (e.g. Van Leeuwen & Jewitt, 2001; Bell, 2012). I used it for

observing ritualistic settings (Article I and II), and architectural, congregational or other religion specific features (Article I and II). However, I combined this with the discursive study of religion approach, in observing how religion is being constructed in the outputs of CIR systems, and how these categorizations/descriptions map to the different cues and features in the images.

In visual content analysis, the content subjected to analysis may be in a variety of forms, including visual, verbal, graphic, oral, or any form of significant visual/verbal information. The visual/verbal units of significance targeted by content analysis are delineated by the medium of their production as distinct, independent, or discrete, such as paragraphs, framed images, pages, news photographs, social media posts, likes and retweets, or labels yielded by a commercial image recognition service. These units are frequently referred to as 'texts' by some researchers conducting content analysis, irrespective of their verbal or non-verbal nature (Bell, 2012). For instance, a display advertisement can be viewed as a visual text. Correspondingly, a news item broadcast on an online platform can also be considered a text due to the clear demarcation within which its varied elements of sound and image cohere or attain cohesion. Texts are thus defined within the purview of a specific research question and within the theoretical categories of the medium (e.g. painting, television, social media platform, etc.) and genres (e.g. portraits, news, posts, etc.) that the research focuses on.

Critical scholars contend that content analysis has been widely employed due to its perceived "commonsense" approach to investigating media content, or due to the notion that it requires minimal theoretical analysis (Bell, 2012). It appears to be a rudimentary method of deciphering the meaning within media and permits the formulation of seemingly general statements about aspects of representation that can be understood by non-specialists, journalists, and experts alike. However, when executed appropriately, content analysis can be a fairly technical procedure. Moreover, content analysis is not intended to be done in "methodological isolation", because it generally is insufficient to substantiate claims about the significance, effects, or interpreted meaning of a representation domain (Bell, 2012).

Visual content analysis, a subcategory of content analysis, is a systematic, observational technique utilized to test hypotheses concerning the manner in which selected media represent individuals, events, situations, etc. It allows for the quantification of observable content samples sorted into distinct categories. Following Bell (2012), potential research questions that may be addressed using visual content analysis include:

1) Queries about the priority/salience of media content: How are different kinds of images, stories, and events represented in terms of visibility (frequency, size, order)?

2) Queries about 'bias': comparative questions concerning the duration, frequency, priority, or salience of visual representations of for example political personalities, issues, policies, or of 'positive' versus 'negative' features of representation.

3) Historical changes in modes of representation, for example, the portrayal of gender, occupation, class, or ethically codified images in specific types of publications or social media genres.

In my sub-studies, variables used for visual content analysis encompass for example types of image recognition service, size of the image data, (defined) ritualistic pose of the represented person (standing, seated, bowing, worshiping), and depicted context (e.g., indoors, temple/church, outdoors). An explicit hypothesis I for example proposed whether women were depicted in fewer religious authoritative situations than men across all services. However, I was mindful on the fact that 'only upon the formulation of one or more hypotheses do the relevant variables become apparent' (Bell, 2012). Notably, the hypotheses that I formulated for visual content analysis evaluation, were always comparative. To put this in some empirical and theoretical context, for example in the case of comparing image recognition services outputs in relation to race (article II), my main theoretical framings was in observing how "representational silence" (Hall, 2002) manifests in the outputs of CIR systems. This included both statistically comparing the frequencies and distributions related to representation or non-representation, and also qualitative analysis and assessment of the component(s) in a specific image that could be causing the "representational silence" (e.g., race, gender, pose, architectural or outdoor setting etc.).

3.2.2 Using Word Embeddings for Measuring Semantic Similarity

In Article III, data science methods, particularly NLP-based techniques, were adopted for measuring similarity between labels (words) CIR services produced. *Word embeddings* offered a tool for evaluating whether two given labels are used in typically similar context. Word embeddings is an area in computational linguistics called distributional semantics, which is interested in how meaningful units of language are distributed across samples of text data. The foundational associative assumption is, that words that are used in similar contexts, also relate to similar meanings.

In an actual mathematical implementation of word embeddings, individual units of analysis are vector representations of words based on what surrounds their use (e.g. Turian et al., 2010; Socher et al., 2010). For instance, the vector

representations of a *church* and a *temple* would be similar and clustered more closely in the conceptual semantic vector space, than the representations of a *church* and a *priest*, which would lie a little further apart in that space since they are far less similar (don't share the category of "a place" or "architecture"). The representation of a *restaurant* would be very distant from *priest*, but since it belongs to the category of places, it would bear some resemblance to *church* and *temple*, and thus it would be somewhat close to them (however it doesn't share the category of "religion" or "religious"). There are several ways to produce vector representations for words, with two well-regarded preexisting models for this being *Word2Vec*, from Google (Mikolov et al., 2013), and Stanford University's *GloVe* (Pennington et al., 2014). Because they are created on the basis of different texts (for example with the whole of Wikipedia), they may yield different results, especially when applied to image labels instead of the kind of material used in their training (i.e., human-produced text). This can be mitigated to some degree for example by choosing distinct pre-trained models for analysis. The similarity between two vectors in a vector space can be measured by using mathematical similarity measures. One good option is to use *cosine similarity*. This method is one of the standard ways of evaluating the distance between word vectors in high-dimensional vector spaces (e.g., Rahutomo et al., 2012; Kozlowski et al., 2019b).

3.2.3 Statistical implementations

Through computational statistical analyses, it becomes possible to compare the distributions, and to identify whether some chosen sample-groups appear to have consistently higher variance. If distributions cannot be assumed to be normal (such is the case with CIR outputs in article III), non-parametric testing such as the Kruskal–Wallis test needs to be implemented (Nummenmaa, 2023). Computing effect sizes is also important in order to reveal how much of variation in the dependent variable can be explained by the independent variable. In article II, the effects of gender and race in the context of religion for the recognition capability of CIR systems was assessed by looking at each services output for every image to see if anything related to religion was identified (tags about religion), and then collecting the results. This produced a statistical table of how well the services comparatively did.

Word clouds and word frequencies are computational linguistics methods, that can be used as a descriptive statistical tool to illustrate the properties of data. In article I, they were used for seeing how the category of "religion" is in general

being recognized and constructed by CIR systems (see Figure 6).

Figure 8 an example of a word cloud used in article I.

3.3 Some ethical considerations

The academic endeavors encapsulated in this dissertation were conducted in accordance with the research ethics guidelines stipulated by the Finnish Board of Research Integrity (TENK), with additional insights from the Association of Internet Researchers' Ethical Guidelines (Franzke et al., 2020).

An important ethical starting point for all research should be trying to protect human (and animal) subjects. However, in the computational social science and data science realm, things are not as straightforward as for example in conducting qualitative interviews, where one can more easily mask the identities of the interviewed or ask for their consent for participating in a study. When I decided to use data from Google Images, I knew that Google has already demanded at least some kind of consent from the users for using the service. Google states in its policy, that although it does not claim ownership rights to user content, by uploading images to the service the user grants Google a license to use, host, store, and reproduce the images[10].

However, for this academic work, none of the images used are being stored or hosted. They were only used for producing the classifications from the CIR systems. This means that neither the platform(s) used nor the individual users posting images on these platforms constituted the primary research subject. The primary research objects were the CIR systems and their classifications.

[10] https://support.google.com/websearch/answer/10622781?sjid=17679858866272724237-EU

The problem of the "fluid" nature of both the digital research materials and the research object (CIR) present some obvious challenges. Digital research materials are dynamic and constantly evolving which means that the collected data is bound to a specific temporal situation (Paasonen, 2020; Markham & Gammelby, 2017; 2018; Webb Williams et. al., 2020). Fluidity is also born out of various other factors, such as technological and software updates, leading to disruptions across multiple layers. Alterations to code, language, metadata, software, and the platforms where images are displayed all influence the availability of research material and how the technologies behave. Over time, CIR systems evolve, resulting in different outputs. Such transiency shapes the possibilities for research.

Embracing this impermanence involves recognizing it as an inherent trait of digital research, rendering it not just a characteristic but also a subject of study. The continual evolution, or the "flow" of the research object, to borrow the term from Markham & Gammelby, extends beyond a problem to be resolved through storage and preservation of content (or trying to stick with same technologies) (Markham & Gammelby, 2017). My principal strategy for navigating the inherent instability resulting from this digital fluidity involves cultivating transparency by meticulously documenting all research steps.

There was an ethical dilemma concerning Articles I and II, that by choosing the image data for the study, I was also producing certain representational framings, and thus constructing the category of religion. However, much effort was put for getting as diverse data as possible from Google Images. Multiple keyword combinations were piloted, and the issue was approached with the expertise that comes with a degree in religious studies. I also followed the idea of a "saturation point", familiar from qualitative research, where the researcher observes if the addition of data increases diversity, information and knowledge of the research subject/object, or if it is just "more of the same". Employing this reasoning assures, to a reasonable degree, that when the goal was merely to examine the types of religious representations produced by CIR systems, any semantic restriction regarding the category of religion would predominantly stem from the initial data source, such as Google Images, or from the inherent representational limitations or biases of these systems.

In Article III, a part of the data had already been used in research. For this data, permission to use it was asked and granted from those that did the original research. However, for Article III some new data was also collected from Twitter and Reddit. Again, the data was not stored or hosted for any long-term use. An inherent problem for this approach is of course that it makes it harder to fully reproduce the research.

An ethical problem related to communicating my research manifested when I was all of a sudden accused of "politicizing research" during a cognitive study of religion conference where I was invited to have a poster presentation. Since then I

have encountered similar reactions in other contexts when presenting findings of my study. The reactions can be seen as indication of several things: Firstly, working with an interdisciplinary research approach creates situations that uncover paradigmatic understandings of different disciplinary fields. Secondly, the recent developments of political polarization in the US in particular have ripple effects in academic research in different parts of the world. This can be seen even in the Nordic academia where the so-called cultural wars are not an intensive phenomenon. Thirdly, the findings of my research do have political implications, which challenges more technologically oriented research of AI. Fourthly and finally, the findings can seem uncomfortable in fields that rely on the ideas of continuous progress and technological solutionism.

4 Findings and discussion

This dissertation is based on research presented in three peer-reviewed publications. They are all jointly authored works for which I acted as the first author and conducted most of the empirical work on the research design, data collection, analysis and reporting. In what follows, I outline the main findings of the sub-studies with regard to the research questions:

1) How do the CIR systems of Google, Amazon, and Microsoft construct representations of religion?
2) What do we learn of CIR, their processes and implications by employing theories of identity, representation and recognition?
3) What can be said about algorithmic bias in the context of CIR, when analyzing it on representational and methodological levels?

Together the three articles highlight the following three aspects of CIR: the nonrecognition and decontextualization of religion, production of representational biases especially in relation to race, and the subjectivity of each individual system as a research tool.

In the following subchapters I will first discuss the empirical findings of each of the article. I will then provide an integrative account that joins theoratizations of identity, representation, recognition and bias together as analytical lenses.

4.1 Article I: Analyzing religious representation and recognition in CIR systems

Article I is an interdisciplinary investigation into how CIR represents images about "religion". It combines computational social science methods with qualitative content analysis. Since similar studies had not been previously done in the field of religious studies, we aimed to perform initial explorations into CIR's representational capabilities, as well as paving the way for future digital religion research endeavors on CIR.

Firstly, we found that for the total of 9,092 labels that the dataset (1188 images about religion) produced, only 85 (2,5%) were related to something "religious" or "spiritual". This means that in general, CIR systems produce and

adhere to a rather secular worldview. We identified this as "secular bias". Out of the 85 unique labels (referring to religion), 30 were explicitly related to Christianity. The categories refer to clerical hierarchies, architectural places, symbols and rituals. Considering that the proportion of images about Christianity were under 14% of the whole dataset, our initial conclusion is, that when recognizing religion, all of studied three CIR systems favor Christianity. This became tangible when comparing the descriptions CIR systems produced in relation to religious authority: in the context of Islam, the label "imam" was not present in any of the labels for all services. This finding was systematic, although quite a few images in our data represented imams teaching in mosques. Our hypothesis is that the label "imam" had been removed from all of the studied systems.

Secondly, we observed that the systems accurately identified the religious and spiritual practices of Japanese Shintoism in the images, reflecting a precise vocabulary for such content. These included "Shinto priest", "prayer", "temple", or "ritual". However, a closer qualitative analysis uncovered that the higher amount of religious recognition seemed to correlate specially with tourist-like photos that depicted e.g., the torii symbols, temples, or other architectural sights related to Shintoism and Japan. It is therefore plausible to assume, that this recognition capability might arise from the fact that these systems had been partially taught with tourist images and their captions. Japan has a long and strong historical tradition of pilgrimages to temples, nowadays mixed with a wide range of domestic tourism (e.g., Reader & Swanson, 1997). Today, photography of the sites is essential. Japan is perceived as an attractive and safe tourist site, and therefore generating a lot of photos, compared to for example many Islamic countries, which might explain this kind of "tourism bias". Another similar, possibly distributional training data bias was identified in the context of Hinduism, where CIR systems represented Hinduism through the conceptual space of fashion. When doing so, they recognized poorly the religious, spiritual or ritualistic aspects of the images, even when the images represented for example intense Hindu piercing rituals.

Thirdly, a close qualitative content analysis of the images of Christianity showed that the systems' ability to classify religious images varies greatly depending on the Christian tradition and the color of skin of the people in the images. If the picture shows white people in ritual situations in European church buildings, the systems provide many diverse identifiers related to Christianity, such as "bishop", "priest", "pope", "sacrament", "ceremony", "vigil" or "prayer". This conceptual richness and related accuracies decrease dramatically with the shift from Catholic, Orthodox, Anglican and Protestant buildings and rituals to more charismatic forms of Christianity. When people in the images are not white, the services lose their capability of recognizing religious features even in the context of Christianity. The studied CIR services have great difficulties in identifying Christianity, or especially Christian authority, in images with Black people, even when the images are rich in

Christian elements. For example, images of Black people at devotional services do most of the time provide no religious identifiers (and if it did, they were scarce compared to images showing white people in similar settings), even when the setting of the event and the architecture of the building give plenty of visual clues and features to help classification. CIR services also systematically interpret images of charismatic affective ritual situations, such as exorcisms or baptisms, as some kind of nonreligious show or performance ("crowd", "night life", "performer", “water sport”), and the labels do not refer to ritual.

4.2 Article II: The racial biases of CIR in the context of Christianity

Article II continued to test and examine further some of the findings of article I related to Christianity, with an emphasis on gender and race. In digital religion research this type of setting is novel. In addition, previous critical AI research on recognition technologies had ignored religion while examining gender and race.

The main empirical finding of article II is, that when controlling for race and gender, the amount of recognition is dramatically and systematically different between races (see Table 2).

Table 2: Comparisons of accuracy when controlling race and gender (from Article II).

Dataset (context), total	Accuracy (Amazon)	Accuracy (Google)	Accuracy (Microsoft)
Christianity (black), 96	27%	16%	21%
Christianity (white), 79	73%	68%	58%
Female (black), 29	48%	14%	10%
Female (white), 29	52%	48%	38%
Male (black), 61	46%	33%	23%
Male (white), 59	86%	75%	52%

By conducting closer qualitative content analysis, some gendered stereotypes and sexist connotations were found. Gender has little effect on the racial bias: white Christian female authorities are identified better than their Black counterparts and they enjoy the same rich descriptive conceptual space of labels related to Christianity as their white male counterparts do. However, in some cases Black woman bishops were identified as “man”, and in relation to charismatic Christianity, Black evangelical female preachers were identified with categories

related to dancing, nightlife, and even dating. This indicates that the systems also reproduce sexist stereotypes and contribute to sexualization of Black women.

4.3 Article III: different CIR systems disagree on what they see

The main aim of Article III was to address the concerns of methodological validity and reliability associated with the utilization of CIR in scholarly research. For this article, only computational social science methods were used, since the data comprised a total of 26,230 images.

The key insight emerging from computational and statistical analyses is that there is an inherent lack of mutual consistency among CIR services with regards to their categorizing of images. Variation is evident in both the quantity of labels assigned to images and, more critically, in the semantic comprehension of the image content. In addition, statistical investigations in Article III reveal that the designed cross-service label-scores (COSLAB) display considerable variation across different service pairs. Labels from Microsoft Azure Computer Vision and Amazon Rekognition, and Azure and Google Cloud Vision display greater congruity compared to others, but even the Azure–AWS COSLAB values do not indicate a definitive semantic consensus on the labels, demonstrating significant variance.

The main conclusion is that different services see and categorize images very differently. When tested using eight thematically diverse image corpora, the overall fact is that the underlying cause may be systemic. This might be because of differences in algorithmic architecture or training data.

4.4 Synthesizing the findings

In this subchapter I will synthesize the empirical findings with my research questions and against the theoretical backdrop introduced in chapter 2. I will aim to proceed in somewhat similar thematic manner as in chapter 2, starting from representation and identity, and then moving the focus more on recognition and bias.

4.4.1 Analyzing secularization and racial bias within CIR

Commercial Image Recognition (CIR) systems, as explored in the findings of this dissertation, notably serve as agents of secularization through their algorithmic processes, often at the expense of religious content and representation. The systems studied in this work, predominately developed and trained within Western secular contexts, exhibit a general tendency to interpret and categorize religious imagery

through a secular lens. This is reflected tangibly in their frequent mis- and non-classification of religious symbols, figures, esthetics (architecture, gatherings) and ritualistic settings. Besides secularism, the findings show how CIR systems also display biases, identified in the context of Christianity, commercialism, tourism, gender, and race.

In the light of the theoretical underpinnings provided in Chapter 2, it is clear that the representations (and biases) that CIR systems produce are deeply rooted in the cultural structures embedded within the algorithms themselves. The representations CIR systems produce of religion, observable and identifiable in the outputs of the systems, can be attributed to the datasets used for training these technologies, which are also laden with the cultural (and cognitive) biases of the societies from which they originate. Origins of the biases can be traced to the location and (historical) context of the data, to the human annotations of the data before training, and to the annotation guidelines provided by the companies developing the systems.

The representational and recognition processes of AI driven recognition technologies such as CIR are a complex interplay of cultural expectations and existing power dynamics. Actively partaking in the contemporary digital meaning making processes, these systems do not reflect reality but actively construct it by promoting certain values and norms over others. From this it follows, that CIR systems easily become decontextualizing algorithmic systems that render diverse cultural phenomena, such as those related to religion, into simplified reductions relating to some categorical semantic space the system has found correlating with the training data, or goals, it has been given. In these kinds of recognition processes of CIR, religion can become removed from any original, local context it might have had, and being rendered to some universal or abstract idea.

This comes visible in the secular bias and in the case of religious representation. The observable predominance of secular, Western values in the outputs of CIR, reflects both the decontextualizing aspect and the dataset compilation and algorithmic design, leading to a diminished visibility and often distorted portrayal of religious diversity. Same kind of distortion appeared in the case of tourism bias, where cultural and religious representations of certain countries (such as Japan) where being categorized by the CIR systems as “spiritual” and “aesthetics”, while others (such as the Muslim Middle East) were either mis-recognized, or uncategorized altogether, as in the case of the missing category of imam. This automated practice of unrecognition is a contemporary example of what Stuart Hall identified as “representational silence” decades ago.

The findings also illuminate significant racial biases in how CIR systems represent Christianity. The representation of Christianity in CIR systems frequently adheres to Western, predominantly white images. These biases manifest in the stereotypical portrayal of religious authority figures and followers, often aligning

with dominant Western (racial) distributions, that do not reflect the diversity within Christian communities. As mentioned above, these biases stem from the recognition processes within CIR systems, which rely on culturally and historically specific datasets. This not only distorts the global (and historical) reality of Christianity's diversity, but also reinforces long-standing racial stereotypes. Following Hall's conceptualization on representational silence, the "silence" regarding Black Christianity and concepts related to Islam reflect global power structures and reinforce existing hierarchies that can perpetuate marginalization.

In chapter 2.1.1, I also introduced Stuart Hall's conceptualization of racisms, relating to the global use and circulation of culturally specific understandings of race. Based on the finding of this work, same kind of logic applies to the culturally specific understandings of religion. As CIR systems are available (and connected) transnationally, meanings produced by these systems have the potential to circulate through various cultural and social settings, affecting the cultural expectations of race, identity and religion. This might have many consequences.

As identity construction and imagining of self is happening more and more on global digital platforms and in social media, embedded with recognition technologies such as CIR, the representations and structures of expectations regarding religious identities and practices are mediated globally through these systems. Molded perceptions can escape the digital, becoming parts of our everyday tangible reality and guiding our way of being in the world. An apparent aspect of this entanglement are different conflicts between the cultural structures of expectation and the meanings produced by AI. Firstly, by marginalizing religious viewpoints, CIR systems can contribute to a broader cultural narrative that undervalues or misinterprets religious expressions and symbols. Secondly, the often distorted, and biased meanings produced by CIR not only affect the communities (un)represented but also shape understandings of religion in different parts of the world, potentially strengthening stereotyping, racism or cultural homogenization. Thirdly, as religion is a significant social and psychological force in the formation process of individual and group identities, the inability of CIR systems to recognize and represent religion in certain racial and/or ethnic contexts has direct social and psychological consequences for individuals and groups belonging to and identifying with these groups.

Considering the six dimensions of representational harms by Shelby et al. (2023), discussed in Chapter 2.4.2, it is clear that the representational silence of the CIR systems can be considered as demeaning, erasing, and alienating social groups, and denying people the opportunity to self-identity. By lacking representational diversity and inclusivity, CIR systems also echo the sociopolitical contexts in which these systems are developed.

Although the findings suggest that there was no statistically significant gendered element in the whiteness bias of Christian authority figures, some gender

bias was evident and observable with qualitative methods. In some images of charismatic Christianity, CIR systems sexualized black women, associating the semantic dimension of the classifications with "nightlife" or "dating" rather than religion. By doing so, they reproduce the gendered way of sexualizing racialized women identified for example by previous critical feminist research.

To conclude, the general representational and recognition inabilities, followed by the biases identified in this study are not trivial; they play a significant role in reinforcing unequal social hierarchies. The misrepresentation and representational silence of non-white people within Christian contexts for example can diminish their visibility and perpetuate exclusion. But can these distortions be identified as algorithmic biases in the strictest sense of the term? According to the tripartite categorization of computer biases made by Friedman & Nissenbaum (1996, discussed in Chapter 2.4.1), the answer is not clear. To recap, Friedman & Nissenbaum's evaluation criteria for computer biases (pre-existing, technical or emergent) are that a system assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate. In Friedman and Nissenbaum's view, bias can be called unfair discrimination only if it occurs systematically and is joined with unfair outcome.

Although it is not possible with the research setting of this study to take a direct stance on the "undesirable outcomes" of CIR on religious individuals or groups, these can be reflected upon based on previous findings of the consequences of AI bias. Furthermore, against the analytical backdrop of Friedman & Nissenbaum (1996), the secular and racial biases, in addition to the Christianity bias (indicated by the absence of the concept of "imam"), can be identified as clear biases also under Friedman's and Nissenbaum's criteria. It is thus possible to conclude that the observed biases meet at least some of the criteria of systematic biases. It is also plausible that they could assign undesirable outcomes to religious individuals or groups. These biases can easily be identified with the taxonomy of sociotechnical harms of algorithmic systems by Shelby et al. (2023).

4.4.2 Representational silence as part of the struggle for recognition

While CIR marks a major technological shift of our time, it also mirrors broader societal problems, such as questions related to visibility, inequality and racism. However, for a recognition technology such as CIR, representation is inherently about detecting visual features. From the viewpoint of recognition theories on the other hand, recognition is essentially about identity. Although visual features can be included as a part of identity, identity obviously goes beyond visual features. Identity combines subjective experience and self-understanding, temporal continuities, geographical locations, racial characteristics and external

classifications and meanings. It is something “holistic” and “experiential”, for lack of a better words.

Recognition is tied to struggles over meaning formation, subjectivity, identity, ideology, resources and power. Identity thus overlaps with questions of machinic recognition. Axel Honneth’s recognition theory emphasizes the fundamental human need for recognition, positing that identity formation and personal development are deeply contingent on the acknowledgment of one’s rights, capabilities, and achievements by others. Misrecognition or denial of recognition, therefore, not only marginalizes individuals but can also perpetrate harm, leading to diminished self-worth and social exclusion. When applied to CIR systems, this theory underscores how these technologies can either affirm or negate social identities through their representational practices. If a CIR system consistently misrecognizes or fails to acknowledge certain groups—whether by race, gender, or religion—it contributes to what Honneth terms a ‘struggle for recognition,’ where marginalized groups fight against their invisibility and misrepresentation in digital and mediated spaces. Honneth's concept of "struggle for recognition" can also be understood in a more "positive" sense, in that it is about awakening to the political struggle to achieve social status. Honneth stresses that the experience of injustice produces the need to struggle to be seen. For Honneth, this struggle is always present somehow and somewhere, as an in-built feature of human societies. Therefore, for Honneth, struggle is not the problem, but *systemic* misrecognition and misrepresentation are.

Hall’s concept of representational silence complements Honneth’s ideas by addressing the absence or underrepresentation of certain groups in media and cultural representations. Representational silence can be seen as a form of structural violence that erases or marginalizes (decontextualizes) certain identities from public consciousness, contributing to a form of symbolic erasure. In the domain of CIR, this manifests when these systems fail to accurately detect or categorize images of certain demographic groups due to biases in the training data or the algorithm’s design, effectively silencing their presence.

As articulated in chapter 4.4.1, in practice, CIR systems often reflect and perpetuate the biases present in their training data. If the data predominantly features certain racial or gender groups, the system will be more adept at recognizing these groups while others remain underrepresented or misrepresented. This bias leads to a form of representational silence where, for instance, non-white or non-male images are less likely to be accurately recognized or categorized, thus perpetuating existing social inequalities. Following Honneth, this misrecognition can lead to a struggle for recognition as these groups seek to assert their visibility and significance. However, since it is well known that companies have also removed certain tags as a solution to the problems of bias, the overall problem space cannot be reduced simply to matters of biased training data.

Moreover, the struggle for recognition in the context of CIR systems and religious representation is not just about visibility but also about the accuracy and dignity of representation. For example, if a CIR system consistently mislabels cultural or religious symbols, it not only misrecognizes but also potentially disrespects the cultural significance of these symbols.

4.4.3 Can CIR in principle mediate genuine recognition?

Besides the struggle for recognition, the central philosophical problem preceding it is the question of technologically mediated recognition. Technological systems such as CIR are not yet considered as "genuine" mediators of recognition. Like I briefly discussed in chapter 2.3, an argument against technological systems has usually been made by using the philosophical requirement to intentionality – their incapability of perceiving the "aboutness" of their targets. In other words, the recognition-relations of CIR alone are perceived as mere physical or spatiotemporal interpersonal relations (connections between lines, orientations, objects and features), without any subjective view or stance. The problem is that even if they are not granted this kind of intentionality, they still have some sort of agency – and first and foremost social consequences – because of the designed function and goals they have been given. For example, since they are trained on culturally produced human data, they are vessels of our culture – in good and evil. Their agency is surrogative since they adhere to certain mediated subjective and perspective positions: they follow the opinions and standards of those who design and control them, and those that emerge from the training data that also carries emerged cultural structures of expectations.

When approaching this problem in relation to the conceptualization of Honneth's three forms of intersubjective recognition, love, equal respect, and social esteem, the machinic nature of CIR seems to prevent them from granting genuine recognition. There however seems to be a kind of an "anthropocentric bias" in this kind of logic. The rejection of machines should not be based on mere aspects of their subjectivism or perspectivism (they follow someone's rules, ideas etc.), since this also applies to us. Intentionality on the other hand depends on the definition. So what exactly is it that makes humans better and suitable for recognition? Some kind of species related similarity? From a strictly philosophical viewpoint, and without positing anything non-material to human recognition (such as a soul), the role of AI systems should be further considered in recognition theories.

It would seem, that a respectful act of religious recognition is impossible without the capacity to represent 'the other' in a holistic way. For a computer vision system, this kind of "holistic sight" would seem to entail the ability not only have

some kind of wider contextual "meta representations" as "fail safes" for the cases of novelty, but also the ability to combine and smash features between any existing representational composite on top of the meta representations. From a computational and architectural perspective, this might entail building systems with constant feedback-loop mechanisms that increase the representational flexibility in an environment, where everything (objects, features of the environment etc.) is in constant flux. From the viewpoint of a system, this kind of architectural design might help with its "internal struggle for recognition", in predicting from multiple sources of input in a noisy environment with competing signals.

4.4.4 Alleviating the methodological bias of CIR

As was articulated in the introduction of this dissertation, social science and humanities (SSH) disciplines have a strong need to adopt different computational means and methods. However, the findings of this dissertation show that there is a valid basis for caution in employing existing CIR systems for scholarly purposes. The main finding of this work in methodological terms was, that different CIR services disagree on their classifications. Thus, in many research cases rather than using CIR, it would be more reliable to prioritize manual image annotation, guided by specific research questions at hand.

However, researchers need to balance a cautious approach against the potential advantages of employing CIR. In the case of choosing to use CIR, there need to be ways to alleviate potential disadvantages, although reliable ones are hard to come by. The first possibility to tackle the problem of disagreement would be to build a custom classifier. Building a custom classifier is, however, laborious and often necessitates a level of standardization that is not feasible for more exploratory work. When classifiers struggle to adapt to image analysis conditions without pre-defined categories, the process can hinder open-ended, fluid, and dynamic exploration of large-scale visual materials. Additionally, it is unlikely that custom tools will be able to fully address all potential challenges, such as the problems related to representational biases. In addition only few researchers have the comprehensive technical expertise required to build even a basic image-classifier from scratch, whereas easy-to-access and comparatively inexpensive image-recognition services can be seen as helping in democratizing image-analysis-based techniques. Therefore, instead of outright avoidance of scholarly application of such systems, it would instead be wise to follow the mitigation strategies proposed in Article III. These enable scholars to utilize these services in a responsible, reflexive manner, while considering cross-service label disagreement.

As a general strategy, it is important to engage using at least two separate services to label their data (a method adhered to in Articles I and II). This practice

facilitates a meticulous comparison of similarities and differences between the services and enhances the probability of one's findings not being merely the product of any service's idiosyncrasies.

However, more specifically, two precise strategies can be used at enhancing the overall validity by managing these differences. The first strategy – arguably the more manageable during the data-analysis phase – is using all generated labels during the analysis. Although analyzing all the labels is important for any critical comparative work, because of the sheer amount of the labels, this would in practice many times mean settling to use only one service or reduce the data to a significantly smaller proportions. This reduction of data size however inhibits finding possible patterns that might emerge when utilizing a larger data. The second strategy centers on the cross-service label agreement, as an indicator of label's quality. This approach necessitates viewing image labels as data in their own right, accepting disagreement as a natural occurrence that must be controlled for during the analysis stages.

An empirical challenge that can manifest in research contexts utilizing CIR is the aforementioned issue of services using comparable concepts but expressing their outputs differently. For instance, if one service labels an image as 'priest' and another assigns 'bishop', do these services disagree, or is the underlying concept – a hierarchical Christian clerical position – sufficiently similar? The invented COSLAB-score approach assesses similarities at this level of detail, thus giving a reliable way of assessing the services. This technique helps to eliminate ambiguous labels from the dataset, while retaining those with sufficient agreement.

4.4.5 What to do with algorithmic bias?

The current wave of digital religion research emphasizes the interplay between online and offline spaces, with a focus on the existential, ethical, and political dimensions of digital religion. The discussion on algorithmic bias relates strongly to all of these dimensions.

However, the discussion of algorithmic biases can be misleading if it rests on an understanding that all the mistakes are just technological problems. This kind of a presupposition has been identified as “technochauvinism”, but I would boldly call it “technologism” since the term better describes that it is essentially about beliefs and dogmatic positions towards increasing technological solutions. A technologist approach considers computational solutions to be superior to all other solutions, such as political or social solutions. While the tendency to focus on the issue of bias in AI ethics is important, it can also draw us away from assessing the core practices and problems in algorithmic systems. There has for example been criticism towards

the increasing of mathematical expertise, which has become one frequently proposed solution to produce "fairer systems". Instead more focus should be guided towards the underlying social, political, and economic structures.

Based on the findings of this research, the most detrimental cultural aspect of CIR systems is that they reinforce stereotypes and are representationally discriminative in the context of religion. This obviously cannot be remedied with an "technological fix". This is not to say that technological fixes are not important and that an attempt to fix systems should not be pursued. On the contrary: Because CIR systems are lacking semantic connections to the conceptual space of the "religious" and "spiritual", they obviously should be given training data with more diverse cultural representations. In other words, if the desired structures and diversity are not to be found in the training data, it cannot be found at all from the final product (CIR system).

However, there is no point in analyzing CIR as "just" technology. CIR systems are algorithmic systems which integrate work, data, code, and human interaction. Therefore they also inherit all the problems related to these dimensions. While some of the algorithmic biases identified in the findings of this work might be idiosyncratic one-offs, which could be easily corrected (e.g., by diversifying the training data), others, such as the representational silences reveal powerful structural forces needing non-technical, societal and political remedies.

To address algorithmic bias is also to address the economic incentives behind their development. Thanks to existing critical research on these systems many of their biases are already known, however it seems that their owner corporations are not very keen in fixing them. One could even argue that many CIR systems producing biases are behaving as expected: within the designed margin of error. They are "good enough" with their representational capabilities to produce profit despite their problems. This is a common strategy although from an ethical point of view, commercial demands taking precedence over human and social values is not a sustainable strategy.

Addressing algorithmic bias in the context of CIR does require heightened academic attention. A substantial proportion of this attention should be also directed to the study of religion and CIR. A concerted effort to diversify the sources of visual data used in training the systems could be considered as a first step. By fostering a more inclusive approach to data collection and algorithm design, CIR systems might more accurately and equitably contribute to a better digital portrayal of religious diversity. This however calls for added cultural and religious studies expertise in the design process of these systems. It also demands the utilization of novel and reliable ways of using and testing CIR systems, such as by implementing the COSLAB approach. These aspects would help mitigate some representational bias and ensure a more balanced representation of global religious practices and beliefs. However, also the right of not being unwillingly classified by recognition

technologies should be discussed. For some vulnerable groups, such as the Uighurs under constant Chinese digital repression, gaining this kind of "invincibility" could be seen as important. On March 2024, The European Parliament harmonized rules and regulations on AI in what became as the Artificial Intelligence Act[11]. Although the act did not concentrate on recognition technologies, it however raised concerns about AI-driven manipulative algorithmic practices. According to the act, such practices can exploit users' vulnerabilities, especially when users are unaware of the manipulative intent behind the content they encounter (pp. 3). These, according to the act, 'can undermine personal autonomy and democratic processes' (pp. 3).

[11] https://artificialintelligenceact.eu/

5 Conclusion

In this dissertation, I have examined the recognition and representation capabilities of commercial image recognition (CIR), and my findings have shown that CIR services reproduce and amplify historical and societal representational biases, especially in relation to religion and race. This results in biased representations, manifesting in the misrecognition or non-recognition of non-Western, non-Christian religions and particular racial groups, such as Blacks. Following Stuart Hall, this was identified as CIR systems producing "representation silence". This dissertation has emphasized that CIR is not merely a technological tool; as algorithmic systems CIR embody representations and structures of cultural expectations, which are co-created ideals that fuse material, conceptual, and value perspectives. These expectations have an impact on social perceptions and experiences. While some may hope for technology to develop an unbiased form on its own, history – and the persistence of biases – suggests that technological imperfections are intrinsically tied to social practices and societal values. These biases, when unchecked, tend to privilege hegemonic powers while decontextualizing and rendering marginalized communities invisible or misrepresented. This dissertation's emphasis on recognition theories posits that recognition is essential for individual and group identity and societal cohesion. Current CIR technologies fall short in recognition, further alienating the already marginalized.

Solutions to the biases of recognition technologies such as CIR are not merely technical; addressing the deeper-rooted societal, economic, and political structures is imperative. Therefore, the debates surrounding algorithmic harms need to go beyond identifying biases and errors. Discussions and changes in practices must delve into the reasons behind these errors, and whether biases are truly "unexpected" or rather a manifestation of societal expectations, commercial interests, or arise from the commensurability processes of contemporary datafication. While it is tempting to view technological solutions as the ultimate answer, such a mindset might further embed us in a cycle of "technochauvinism".

This dissertation offers insights into CIR and its implications for the study of digital religion, and social sciences and humanities more broadly. It also presents a novel computational tool (COSLAB) for helping with reliability questions that are related to CIR as a methodology. By investigating the biases and representational

issues of CIR, the research illuminates the challenges and opportunities presented by modern recognition technologies in understanding and representing religious practices and beliefs. The findings emphasize the nuanced dynamics of secularism, consumerism, prejudice, racism, and sexism present within these digital tools, which can inadvertently shape our understanding of religious practices. Moreover, by highlighting the technological reductionism of religious identities into mere data points, the research underscores the inherent danger of simplifying multifaceted cultural expressions and practices. For scholars and practitioners in the digital study of religion, and in the social sciences and humanities more broadly, this work acts as a clarion call for greater scrutiny, reflection, and critical engagement with the digital tools and platforms we employ.

Furthermore, by shedding light on the representations these systems produce, the research positions itself as an effort in bridging the realms of technology and religious studies, fostering a deeper understanding of the dynamic interplay between the two. In essence, my study brings to the fore the urgent need to critically assess and understand the role of technologies like CIR. Only through a comprehensive and introspective examination of the past and present can we hope, imagine and aim to create technologies that are equitable, just, and truly reflective of the diverse tapestry of human experiences.

5.1 Limitations of this study and opportunities for future research

Like all research, this study has its limitations.

Firstly, since image recognition systems evolve and their outputs change, the constant development introduces challenges related to their performance over time. The research articles therefore investigate already lost moments. The data selection reflects the moment of entry into research in the period of 2021–2023.

Secondly, as the data was collected from US based services, a more profound discussion of US race and religious relations as the context of commercial image recognition system' biases would have been ideal.

Thirdly, findings of articles I and II indicate that whiteness is a norm in the outputs of the studied CIR systems. Since in this study the focus was on qualitative close reading, the observed race and gender bias require further research with larger datasets. Using synthetic data could be one solution for enabling a study of CIR systems and the effect of race with added control and reliability.

Fourthly, the interdisciplinary setting of this research posed limitations for theory development since concepts and terms are used and understood differently in different fields. For instance the relationship between social theories of recognition and recognition technologies could be explored further in the future.

Finally, in terms of methodology, different image, character and object recognition technologies will continue to challenge more traditional approaches of the study of religion. At the same time, with all their shortcomings and complexities, they also open up a new world of possibilities for studying religion. For instance, combined with various theoretical approaches, such as cognitive science of religion, or by implementing synthetic data generated by generative AI into the research designs, it could be possible to study e.g., the effects of certain religious and ritual settings, symbolism and embodied expressions on human cognition such as perception, gaze focus or affective and emotional responses and memory. Expanding AI technologies and tools are a treasure trove for social sciences and humanities. This study has scratched the surface of some of the issues they bring for study of religion. In the future AI technologies will continue challenge religious studies with new ontological, epistemological and ethical questions.

References

Ahmed, S. (2017). Living a Feminist Life. Duke University Press.

Ajrouch, K. J., & Kusow, A. M. (2007). Racial and religious contexts: Situational identities among Lebanese and Somali Muslim immigrants. Ethnic and racial studies, 30(1), 72-94.

Alimardani, M., & Elswah, M. (2020). <? covid19?> Online Temptations: COVID-19 and Religious Misinformation in the MENA Region. Social Media+ Society, 6(3), 2056305120948251.

Andersen, J. (2020). Understanding and interpreting algorithms: Toward a hermeneutics of algorithms. Media, Culture & Society 42(7–8): 1479–1494.

Anderson, J. H., & Honneth, A. (2005). Autonomy, vulnerability, recognition, and justice.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2022). Machine bias. In Ethics of data and analytics (pp. 254-264). Auerbach Publications.

Archer, M. S. (2004). Culture and Agency: The Place of Culture in Social Theory, Revised Edition. New York and Cambridge: Cambridge University Press.

Asad, Talal. (1993). Genealogies of Religion: Discipline and Reasons of Power in Christianity and Islam. Baltimore: The Johns Hopkins University Press.

Baker, P., & Ellece, S. (2011). Key Terms in Discourse Analysis. London: Continuum.

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. International Journal of Artificial Intelligence in Education, 1-41.

Barker, E., 2011. The cult as a social problem. In: Hjelm, T. (Ed.), Religion and Social Problems. Routledge, New York, pp. 198–212.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. Calif. L. Rev., 104, 671.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning. http://www.fairmlbook.org.

Barratt, P. (2011). Vertical worlds: Technology, hybridity and the climbing body. Social & Cultural Geography, 12(4), 397–412. https://doi.org/10.1080/14649365.2011.574797

Barrett, L. F. (2020). Seven and a half lessons about the brain. Houghton Mifflin.

Barton, C. A. & Boyarin, D. (2016). Imagine No Religion: How Modern Abstractions Hide Ancient Realities. New York: Fordham University Press.

Bassett, C., et al. (2019). Furious: Technological Feminism and Digital Futures. 1st ed. Pluto Press.

Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. New Media & Society 11(6): 985–1002.

Bell, P. (2012). Content analysis of visual images. SAGE visual methods: Interpretation and classification, 3, 31-57.

Bellar, W. (2017). Private practice: Using digital diaries and interviews to understand evangelical Christians' choice and use of religious mobile applications. New Media & Society, 19(1), 111-125

Bellar, W., & Campbell, H. (2023). Digital religion: the basics. Routledge, Taylor & Francis Group.

Benjamin, R. (2019). Race after Technology: Abolitionist Tools for the New Jim Code. Medford, MA: Polity.

Bennett, C. L., Gleason, C., Scheuerman, M. K., Bigham, J. P., Guo, A., & To, A. (2021, May). "It's complicated": Negotiating accessibility and (mis) representation in image descriptions of race, gender, and disability. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-19).

Bishop, S. (2019) Managing visibility on YouTube through algorithmic gossip. New Media & Society 21(11–12): 2589–2606.

Bivens, R., & Haimson, O. L. (2016). Baking gender into social media design: How platforms shape categories for users and advertisers. Social Media+ Society, 2(4), 2056305116672486.

Bostrom, N. (2005). A History of Transhumanist Thought. Journal of Evolution and Technology / WTA, 14(1).

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Bowker, G. C., & Star, S. L. (2000). Why Classifications Matter. In Sorting things out classifications and its consequences (pp. 319–326).

Braidotti, R. (2013). Posthuman. Politi Press.

Broussard, M. (2023). More than a glitch: Confronting race, gender, and ability bias in tech. MIT Press.

Browne, S. (2015). Dark Matters: On the Surveillance of Blackness. Durham, NC: Duke University Press.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. New Media & Society 14(7): 1164–1180.

Bucher, T. (2018). If... Then: Algorithmic Power and Politics. Oxford: Oxford University Press.

Buell, Spencer. (July 19, 2023). An MIT student asked AI to make her headshot more 'professional.' It gave her lighter skin and blue eyes. Boston Globe. Read, Apr 29, 2024. https://www.bostonglobe.com/2023/07/19/business/an-mit-student-asked-ai-make-her-headshot-more-professional-it-gave-her-lighter-skin-blue-eyes/

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77–91).

Burch, R. (2024). *"Charles Sanders Peirce"*. In *Stanford Encyclopedia of Philosophy* (Spring 2024 edition), Zalta, E. N., & Nodelman, U.(eds), https://plato.stanford.edu/archives/spr2024/entries/peirce/

Burn-Murdoch, J. (2013). The problem with algorithms: magnifying misbehaviour. The Guardian, 14. Read 13.5.2024 from: https://www.theguardian.com/news/datablog/2013/aug/14/problem-with-algorithms-magnifying-misbehaviour

Buslón, N., Cortés, A., Catuara-Solarz, S., Cirillo, D., & Rementeria, M. J. (2023). Raising awareness of sex and gender bias in artificial intelligence and health. Frontiers in Global Women's Health, 4.

Butler, J. (1990). Gender Trouble. Routledge, London.

Calders, T., & Žliobaitė, I. (2013). Why unbiased computational processes can lead to discriminative decision procedures. In Discrimination and Privacy in the Information Society: Data mining and profiling in large databases (pp. 43-57). Berlin, Heidelberg: Springer Berlin Heidelberg.

Cambre, J., & Kulkarni, C. (2019). One voice fits all? Social implications and research challenges of designing voices for smart devices. Proceedings of the ACM on human-computer interaction, 3(CSCW), 1-19.

Cambridge University Press. (n.d.). Decontextualized. In Cambridge dictionary. Retrieved April 29, 2024, from: https://dictionary.cambridge.org/dictionary/english/decontextualized Cambridge: Polity.

Campbell, H. (2010). Religious authority and the blogosphere. Journal of Computer-Mediated Communication, 15(2), 251–276. https://doi.org/10.1111/j.1083-6101.2010.01519.x

Campbell, H. (2012a). Understanding the relationship between religion online and offline in a networked society. Journal of the American Academy of Religion, 80(1), 64–93. https://doi.org/10.1093/jaarel/lfr074

Campbell, H. (Ed.). (2012b). Digital religion: Understanding religious practice in new media worlds. Abingdon, England; New York, NY: Routledge.

Campbell, H. A., & Cheong, P. H. (2022). Introduction to the study of digital religion. In Oxford Handbook of Digital Religion (nd.). https://doi.org/10.1093/oxfordhb/9780197549803.013.40

Campbell, H. A., & Evolvi, G. (2020). Contextualizing current digital religion research on emerging technologies. Human Behavior and Emerging Technologies, 2(1), 5-17.

Campbell, H. A., & Tsuria, R. (Eds.). (2021). Digital religion: Understanding religious practice in digital media. Routledge.

Carr, Nicholas. (2014). The Glass Cage: Automation and Us. New York: W.W. Norton & Co.

Carrera, F. E. (2020). Race and gender of aesthetics and affections: algorithmization of racism and sexism in contemporary digital image databases. Matrizes. 14(2):217-40.

Chancellor, S., Baumer, E. P., & De Choudhury, M. (2019). Who is the" human" in human-centered machine learning: The case of predicting mental health from social media. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-32.

Cheong, P. H. (2013). Authority. In H. A. Campbell (Ed.), Digital religion: Understanding religious practice in new media worlds (pp. 72–87). Abingdon, England; New York, NY: Routledge.

Cheong, P. H., & Chen, Y. (2023). Religious Human–Machine Communication: Practices, Power, and Prospects. Human–Machine Communication, 555.

Cheong, P. H., Halavais, A. and Kwon, K., 2008. The Chronicles of Me: Understanding Blogging as a Religious Practice. Journal of Media and Religion, 7(3), 107–131

Chollet, F. (2022). Deep learning with Python. Simon and Schuster.

Chouliaraki, L., & Fairclough, N. (1999). Discourse in Late Modernity. Edinburgh

Christian, B. (2020). The Alignment Problem - Machine Learning and Human Values. New York: Norton & Company.

Clark, A. (2015). Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press.

Coeckelbergh, M. (2020). AI ethics. Mit Press.

Corry, F., Sridharan, H., Luccioni, A. S., Ananny, M., Schultz, J., & Crawford, K. (2021). The problem of zombie datasets: A framework for deprecating datasets. arXiv preprint arXiv:2111.04424, 1-18.

Couldry, N & Mejias, UA. (2019a). The Costs of Connection. Stanford: Stanford University Press.

Couldry, N. & Hepp, A. (2017). The Mediated Construction of Reality.

Couldry, N. & Mejias, UA. (2019b). Data colonialism: Rethinking big data's relation to the contemporary subject. Television & New Media 20(4): 336–349.

Couldry, N., & Hepp, A. (2013). Conceptualizing mediatization: Contexts, traditions, arguments. Communication Theory, 23(3), 191-202. https://doi.org/10.1111/comt.12019

Couldry, N., & Hepp, A. (2013). Conceptualizing mediatization: Contexts, traditions, arguments. Communication Theory, 23(3), 191–202. https://doi.org/10.1111/comt.12019

Crawford, K. (2021). The Atlas of AI. Yale University Press.

Crawford, M. (2015). The World Beyond Your Head: On Becoming an Individual in an Age of Distraction. New York: Farrar, Straus and Giroux.

Davies, W. (2021). The politics of recognition in the age of social media. New Left Review, (128), 83-99

Day, A. (2011). Believing in belonging: Belief and social identity in the modern world. Oxford University Press.

De Lauretis, T. (2004). Statement due. Critical Inquiry, 30(2), 365-368.

De Vries, H., & Weber, S. (2001). Religion and media. Stanford University Press.

Devendorf, Laura and Elizabeth Goodman. (2014). "The Algorithm Multiple, the Algorithm Material." Contours of Algorithmic Life, UC Davis, May. http://www.slideshare.net/egoodman/the-algorithm-multiple-the-algorithm-material-reconstructing-creative-practice

Devinney, H., Björklund, J., & Björklund, H. (2022, June). Theories of "gender" in nlp bias research. In Proceedings of the 2022 ACM conference on fairness, accountability, and transparency (pp. 2083-2102).

Devlin, K. J. (2012). Fundamentals of contemporary set theory. Springer Science & Business Media.

DeVos, A., Dhabalia, A., Shen, H., Holstein, K., & Eslami, M. (2022, April). Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1-19).

Dodge, J., Prewitt, T., Tachet des Combes, R., Odmark, E., Schwartz, R., Strubell, E., ... & Buchanan, W. (2022, June). Measuring the carbon intensity of ai in cloud instances. In Proceedings of the 2022 ACM conference on fairness, accountability, and transparency (pp. 1877-1894).

Dosono, B., & Semaan, B. (2020). Decolonizing tactics as collective resilience: Identity work of AAPI communities on Reddit. Proceedings of the ACM on Human-Computer interaction, 4(CSCW1), 1-20.

Dwoskin, E. (2015). At Uber, the algorithm is more controlling than the real boss. Wall Street Journal, 4. Read 13.5.2024 from: https://www.wsj.com/articles/BL-DGB-43982

Edwards, D., & Potter, J. (1992). Discursive Psychology. SAGE, London.

Ellul, J., & Hopkin, C. E. (1975). The new demons. London: Mowbrays.

Ellul, Jacques (1980). The technological system. Translated into English by Joachim Neugroschel. NY: Continuum.

Ellul, Jacques. 1964. The technological society [English translation of La Technique]. New York: Knopf

Elsayed, W. (2021). The negative effects of social media on the social identity of adolescents from the perspective of social work. Heliyon, 7(2).

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Evolvi, G. (2018). Blogging my religion: Secular, Muslim, and Catholic media spaces in Europe. Routledge.

Feenberg, A. (2012). Questioning technology. Routledge.

Finnish National Board on Research Integrity (2024). Responsible Conduct of Research and Procedures for Handling Allegations of Misconduct in Finland. Available at: https://tenk.fi/en/tenk-guidelines . Accessed May 2, 2024.

Fitzgerald, Timothy (ed.). (2007b). Religion and the Secular: Historical and Colonial

Fitzgerald, Timothy 2000. The Ideology of Religious Studies. New York: Oxford

Fitzgerald, Timothy. (2015). Critical Religion and Critical Research on Religion: Religion and Politics as Modern Fictions. Critical Research on Religion 3(3): 303–319. Formations. London: Equinox.

Foucault, Michel 2002. The Archaeology of Knowledge. London: Routledge.

franzke, a. s., Bechmann, A., Zimmer, M., Ess, C. & the Association of Internet Researchers (2020). Internet Research: Ethical Guidelines 3.0. https://aoir.org/reports/ethics3.pdf . Accessed May 2, 2024.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on information systems (TOIS), 14(3), 330-347.

Fuchs, C. (2018). Capitalism, patriarchy, slavery, and racism in the age of digital capitalism and digital labour. Critical Sociology 44(4–5): 677–702.

Gault, E. D. (2022). Where are our tools? Or fugitive methods to the study of digital Black religion. The Black Scholar, 52(3), 6-16.

Gebru, T. (2020). Race and gender. In The Oxford handbook of ethics of AI, 251-269. Dubber, M., Pasquale, F., & Das, S., (eds). Oxford University Press.

Geertz, C. (1966). "Religion as a Cultural System," in M. Banton (ed.), Anthropological Approaches to the Study of Religion. New York: Praeger, pp. 1–46

Goff, P. A., Eberhardt, J. L., Williams, M. J., & Jackson, M. C. (2008). Not yet human: implicit knowledge, historical dehumanization, and contemporary consequences. Journal of personality and social psychology, 94(2), 292.

Granholm, K. (2013). Esoteric Currents as Discursive Complexes. Religion 43(1): 46–69.

Gray, J. (2004). Heresies: Against Progress and Other Illusions. Granta Publications.

Gray, John. (2007). Black Mass: Apocalyptic Religion and the Death of Utopia. UK: Penguin books.

Gray, John. (2013). The Silence of Animals: On Progress and Other Modern Myths. Farrar, Straus and Giroux.

Grossberg, L. (1996). Identity and cultural studies: Is that all there is. *Questions of cultural identity*, 1, 87-107.

Gunkel, D. J. (2018). Robot rights. MIT Press.

Hall, S. (1992). Race, culture, and communications: Looking backward and forward at cultural studies. Rethinking Marxism, 5(1), 10-18.

Hall, S. (1997) Representation. Cultural Representations and Signifying practices. SAGE Publications.

Hall, S. (2002). Identiteetti. 5. painos. Tampere: Vastapaino.

Haraway, D. J. (2019). “It Matters What Stories Tell Stories; It Matters Whose Stories Tell Stories.” A/b Auto/Biography Studies, vol. 34, no. 3, pp. 565–75.

Hardt, M., & Recht, B. (2022). Patterns, predictions, and actions: Foundations of machine learning. Princeton University Press.

Harraway, D. J. (1988). “Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.” Feminist Studies, 14(3), pp. 575–99, https://doi.org/10.2307/3178066.

Harwell, D., & Dou, E. (2020). Huawei tested AI software that could recognize Uighur minorities and alert police, report says. Washington Post, 8. Read 29 April 2024. https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says/

Hayes, S. G., & Venkatraman, P. (Eds.). (2016). Materials and technology for sportswear and performance apparel. Boca Raton, FL: CRC Press.

Heidegger, M. (1977). The question concerning technology. Readings in the Philosophy of Technology, 9-24.

Hendricks, L. A., Burns, K., Saenko, K., Darrell, T., & Rohrbach, A. (2018). Women also snowboard: Overcoming bias in captioning models. In

Proceedings of the European conference on computer vision (ECCV) (pp. 771-787)

Hinton, G. (2024). Will digital intelligence replace biological intelligence?" Romanes Lecture. Youtube: University of Oxford. Watched 22.4.2024 at https://www.youtube.com/watch?v=N1TEjTeQeg0&list=WL&index=1&t=1305s

Hjarvard, S. (2008). The mediatization of religion: A theory of the media as agents of religious change. Northern Lights: Film & Media Studies Yearbook, 6(1), 9-26.

Hjarvard, S. (2011). The mediatization of religion: Theorising religion, media and social change. Culture and Religion, 12(2), 119-135. https://doi.org/10.1080/14755610.2011.579715

Hjarvard, S. (2013). The mediatization of culture and society. London: Routledge

Hjarvard, S. (2014). From mediation to mediatization: The institutionalization of new media. In A. Hepp & F. Krotz (Eds.), Mediatized Worlds (pp. 123-139). New York: Palgrave Macmillan. https://doi.org/10.1057/9781137300355_8

Hjarvard, S., & Lövheim, M. (Eds.). (2012). Mediatization and religion: Nordic perspectives. Göteborg: Nordicom.

Hjelm, T. (2021). Discourse analysis. In The Routledge handbook of research methods in the study of religion (pp. 229-244). Routledge.

Hoikkala, T. (1990). Teun A. van Dijkin diskurssianalyysi. In Mäkelä. K. (eds.) Kvalitatiivisen aineiston analyysi ja tulkinta. Helsinki: Gaudeamus, pp.142–161.

Hokka, J., & Nelimarkka, M. (2020). Affective economy of national-populist images: Investigating national and transnational online networks through visual big data. New Media & Society, 22 (5), 770–792. http://journals.sagepub.com/doi/10.1177/1461444819868686

Hokka, J., Valaskivi, K., Sumiala, J., & Laakso, S. (2013). Suomalaiset sanomalehdet uskonnollisen maiseman tuottajina: uskontojournalismi Helsingin Sanomissa, Ilkassa, Kalevassa ja Karjalaisessa vuosina 2007-2011. In Media & viestintä: kulttuurin ja yhteiskunnan tutkimuksen lehti, 36(2), 6-21.

Honneth, A & Fraser, N. (2003). Redistribution or Recognition? A Political-Philosophical Exchange. New York: Verso.

Honneth, A. (1992). Integrity and disrespect: Principles of a conception of morality based on the theory of recognition. Political theory, 20(2), 187-201.

Honneth, A. (1995). The Struggle for Recognition: The Moral Grammar of Social Conflicts. Cambridge: Polity Press.

Honneth, A. (2007). Disrespect. The Normative Foundations of Critical Theory. Cambridge: Polity Press.

Honneth, A. (2008). Reification: A new look at an old idea. Oxford University Press.

Hoover, S. M. (2006). Religion in the media age (1st ed.). London, England; New York, NY: Routledge.

Hoover, S. M., & Echchaibi, N. (2014). Media theory and the 'third spaces' of digital religion (Working paper). Center for Media, Religion, and Culture.

Human Rights Watch (2018). Eradicating Ideological Viruses: China's Campaign of Repression Against Xinjiang's Muslims.

Human Rights Watch (2019). China's Algorithms of Repression: Reverse Engineering Xinjiang Police Mass Surveillance App.

Hutchings, T. (2010). The politics of familiarity: visual, liturgical and organisational conformity in the online church. Online-Heidelberg journal of religions on the internet

Huxley, J. (1957). Transhumanism. In New bottles for new wine. Chatto & Windus.

Jenkins, R. (2014). Social identity. Routledge.

Johnson, S., & Acemoglu, D. (2023). Power and progress: Our thousand-year struggle over technology and prosperity. Hachette UK.

Johnstone, B. (2017). Discourse analysis. John Wiley & Sons.

Jokinen, A., Juhila, K., & Suoninen, E. (1999). Diskurssianalyysi liikkeessä. FI, Tampere: Vastapaino.

Jokiranta, J., & Luomanen, P. (2013). Sosiaalisen identiteetin näkökulma. In Johdatus sosiaalitieteelliseen raamatuntutkimukseen (pp. 69-86). Finnish Exegetical Society.

Jones, S. (2007). Antonio Gramsci. Routledge

Joo, J., & Steinert-Threlkeld, Z. C. (2022). Image as data: Automated content analysis for visual presentations of political actors and events. Computational Communication Research, 4 (1).

Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. (2020). Assessing the Big Five personality traits using real-life static facial images. Scientific Reports, 10(1), 8487.

Karizat, N., Delmonaco, D., Eslami, M., & Andalibi, N. (2021). Algorithmic folk theories and identity: How TikTok users co-produce Knowledge of identity and engage in algorithmic resistance. Proceedings of the ACM on human-computer interaction, 5(CSCW2), 1-44

Karvonen, E. (1992). Odotuksen struktuurit ja populaari representaatio. (Tampereen yliopisto, Tiedotusopin laitos, Julkaisuja. Sarja A). Tampereen yliopisto.

Katzenbach, C., & Ulbricht, L. (2019). Algorithmic governance. Internet Policy Review 8(4): 1–18.

Katzman, J., Barocas, S., Blodgett, S. L., Laird, K., Scheuerman, M. K., & Wallach, H. (2021). Representational Harms in Image Tagging. Beyond Fair Computer Vision Workshop at CVPR 2021.

Keating, T. P. (2023). On the technological unconscious: thinking the (a) signifying production of subjects and bodies with sonographic imaging. Social & Cultural Geography, 24(9), 1481-1500.

Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. Proceedings of the ACM on human-computer interaction, 2(CSCW), 1-22.

Kharpal, A. (2021, Jan). China A.I. firms and Huawei filed to patent technology that could identify Uighur Muslims, report says. CNBS. Read 5.4.2024 from https://www.cnbc.com/2021/01/15/huawei-ai-firms-filed-to-patent-tech-that-could-identify-uighurs-report-says.html

Khazraee, E., & Novak, A. N. (2018). Digitally mediated protest: Social media affordances for collective identity construction. Social Media+ Society, 4(1), 2056305118765740

Kinsley, S. (2010). Representing 'things to come': Feeling the visions of future technologies. Environment and Planning A, 42(11), 2771–2790. https://doi.org/10.1068/a42371.

Kitchin, R., & Dodge, M. (2014). Code/space: Software and everyday life. Mit Press.

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. European Journal of Information Systems, 31(3), 388-409.

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. Scientific reports, 11(1), 100.

Koskinen, H. J. (2019). Mediated recognition: Suggestions towards an articulation. In Recognition and Religion: Contemporary and Historical Perspectives (ed. Kahlos, R. T. M., Koskinen, H. J., & Palmen, R. A). pp. 34-50. Routledge: New Critical Thinking in Religion, Theology and Biblical Studies).

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

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

Kurzweil, R. (2012). How to create a mind: The secret of human thought revealed. Viking Penguing.

Lagerkvist, A. (2017). Existential media: Toward a theorization of digital thrownness. New Media & Society, 19(1), 96-110.

Latour, B. (2007). Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford: Oxford University Press.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Lee-Morrison, L. (2019). Portraits of automated facial recognition: On machinic ways of seeing the face. Bielefeld University Press.

Lemon, E., Jardine, B., & Hall, N. (2023). Globalizing minority persecution: China's transnational repression of the Uyghurs. Globalizations, 20(4), 564-580.

Li, M., et al. (2023). Potential Societal Biases of ChatGPT in Higher Education: A Scoping Review. arXiv. https://doi.org/10.48550/arXiv.2311.14381

Lin, P., Abney, K., & Jenkins, R. (Eds.). (2017). Robot ethics 2.0: From autonomous cars to artificial intelligence. Oxford University Press.

Lin, T. Y. (2023). Granular computing: practices, theories, and future directions. In Granular, Fuzzy, and Soft Computing (pp. 199-219). New York, NY: Springer US

Lindgren, S. (2020). Data theory: Interpretive sociology and computational methods. John Wiley & Sons.

Lövheim, M., & Hjarvard, S. (2019). The mediatized conditions of contemporary religion: Critical status and future directions. Journal of Religion, Media and Digital Culture, 8(2019), 206-225. https://doi.org/10.1163/21659214-00802002

Lövheim, M., & Lundmark, E. (2019). Gender, religion and authority in digital media. ESSACHESS–Journal for Communication Studies, 12(2 (24)), 23-38.

Lundby, K. (2013). Theoretical frameworks for approaching religion and new media. In H. Campbell (Ed.), Digital religion: Understanding religious practice in new media worlds (pp. 225–237). Abingdon, England; New York, NY: Routledge.

Lundby, K. (2014). Mediatization of communication. Berlin, Germany: De Gruyter Mouton.

Lundby, K., Christensen, H. R., Gresaker, A. K., Lövheim, M., Niemelä, K., Sjö, S., ... & Daníelsson, Á. S. (2018). Religion and the media: Continuity, complexity, and mediatization. Religious complexity in the public sphere: Comparing Nordic countries, 193-249.

Lyon, D. (Ed.). (2003). Surveillance as social sorting: Privacy, risk, and digital discrimination. Psychology Press

Magnet, S. (2011). When Biometrics Fail: Gender, Race and the Technology of Identity. Durham, NC: Duke University Press.

Mannes, A. (2020). Governance, risk, and artificial intelligence. Ai Magazine, 41(1), 61-69.

Markham, A. N., & Gammelby, A. K. (2018). Moving Through Digital Flows: An Epistemological and Practical Approach." In The SAGE Handbook of Qualitative Data Collection (Ed. Flick, U.). Sage Publishing, 2017, pp. 451–465.

Martin-Barbero, J. (1993). Communication, culture and hegemony: From the media to mediations. London, England: SAGE Publications Ltd.

Masuzawa, Tomoko. (2005). The Invention of World Religions: Or, How European Universalism was Preserved in the Language of Pluralism. Chicago: The University of Chicago Press.

McCutcheon, Russell. T. (1997). Manufacturing Religion: The Discourse on Sui Generis Religion and the Politics of Nostalgia. Oxford: Oxford University Press.

McCutcheon, Russell. T. (2003). The Discipline of Religion: Structure, Meaning, Rhetoric. London: Routledge.

McNiven, A. (2016). Geographies of dying and death' in relation to pregnancy losses: Ultrasonography experiences. Social & Cultural Geography, 17(2), 233–246. https://doi.org/10.1080/14649365.2015.1033448

Merriam Webster. (2002). Decontextualized. Retrieved April 29, 2024, from: http:// https://www.merriam-webster.com/dictionary/decontextualized

Meyer, B. (Ed.). (2010). Aesthetic formations: Media, religion, and the senses. London, England: Palgrave Macmillan.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781v3.

Moberg, M. (2013). First-, Second-, and Third-Level Discourse Analytic Approaches in the Study of Religion: Moving from Meta-Theoretical Reflection to Implementation in Practice. Religion 43(1): 4–25.

Morozov, E. (2013). To save everything, click here: The folly of technological solutionism. PublicAffairs.

Morris, M. W., Leung, K., Ames, D., & Lickel, B. (1999). Views from inside and outside: Integrating emic and etic insights about culture and justice judgment. Academy of management review, 24(4), 781-796.

Mozur, P. (2019, April). One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority. The New York Times. Read 4.5.2024 from: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

Mullikin, P. L. (2006). Religious and Spiritual Identity: The Impact of Gender, Family, Peers and Media Communication in Post-Adolescence. Journal of Communication & Religion, 29(1).

Nash, J. C. (2019). Black Feminism Reimagined. Duke University Press.

Nature (2016). More accountability for big-data algorithms. Nature 537, 449. https://doi.org/10.1038/537449a

Nelimarkka, M. (2022). Computational Thinking and Social Science: Combining Programming, Methodologies and Fundamental Concepts. Sage Publications.

Noble, S. U. (2018). Algorithms of oppression. New York university press.

Nongbri, B. (2013). Before Religion: History of a Modern Concept. New Haven: Yale University Press.

Nummenmaa, L. (2023). Käyttäytymistieteiden tilastolliset menetelmät. Tammi.

Ollila, M-R. 2019. Tekoälyn etiikkaa. Helsinki: Otava.

Oppong, S. H. (2013). Religion and identity. American International Journal of Contemporary Research, 3(6), 10-16. racial studies, 30(1), 72-94.

Paasonen, S. (2020). “Resonant Networks: On Affect and Social Media.”. Public Spheres of Resonance (ed. Fleig, A. & von Scheve, C.), Routledge, 2020, pp. 49–62.

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis) contents: A survey of dataset development and use in machine learning research. Patterns, 2(11).

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1532–1543).

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a “kernel of truth” in social perception. Social cognition, 24(5), 607-640.

Pentzold, C., Brantner, C., & Fölsche, L. (2018). Imagining big data: Illustrations of “big data” in US news articles, 2010–2016. New Media & Society, 146144481879132. http://journals.sagepub.com/doi/10.1177/1461444818791326

Pink, S. (2022). Methods for researching automated futures. Qualitative Inquiry 28(7): 747–753.

Pinker, Steven. (2019). Enlightenment Now: The Case for Reason, Science, Humanism, and Progress. Penguin Books UK.

Rahutomo, F., Kitasuka, T., & Aritsugi, M. (2012). Semantic cosine similarity. In The 7th international student conference on advanced science and technology icast (Vol. 4, p. 1).

Reader, I., & Swanson, P. L. (1997). Editors' Introduction: Pilgrimage in the Japanese Religious Tradition. *Japanese Journal of Religious Studies*, 24(3/4), 225–270. http://www.jstor.org/stable/30233586.

Renser, B., & Tiidenberg, K. (2020). Witches on Facebook: mediatization of neo-paganism. Social Media+ Society, 6(3), 2056305120928514

Rodriguez, E. M., & Ouellette, S. C. (2000). Gay and lesbian Christians: Homosexual and religious identity integration in the members and participants of a gay-positive church. Journal for the scientific study of religion, 39(3), 333-347.

Ross, R. N. (1975). Ellipsis and the structure of expectation. San Jose State occasional papers in linguistics, 1(18), 3-9.

Rossi, L.-M. (2010). Esityksiä, edustamista ja eroja: Representaatio on politiikkaa. In T. Knuuttila, & A. P. Lehtinen (Eds.), Representaatio: Tiedon kivijalasta tieteiden työkaluksi (pp. 261-275). Gaudeamus Helsinki University Press.

Ruckenstein M (2023) e Feel of Algorithms. Oakland: University of California Press.

Saarinen, R. (2016). Recognition and religion: A historical and systematic study. Oxford University Press.

Safdar, M. R., Akram, M., Ahmad, A., & Ayaz, A. A. (2023). The Role of Religion and Spirituality to Cope with COVID-19 Infections Among People of Lower Socioeconomic Status in Pakistan: An Exploratory Qualitative Study. Journal of religion and health, 62(4), 2916-2932

Saler, Benson. (2000). Conceptualizing Religion: Immanent Anthropologists, Transcendent Natives, and Unbounded Categories. New York: Berghahn Books.

Schwarz, O. (2019). Facebook rules: Structures of governance in digital capitalism and the control of generalized social capital. Theory, Culture & Society 36(4): 117–141.

Schwarz, O. (2021). Sociological Theory for Digital Society: The Codes that Bind Us Together. Cambridge: Polity Press.

Seaver, N. (2019b). Knowing algorithms. In J. Vertesi & D. Ribes (Eds.), digtalSTS (pp. 412–422). Princeton University Press. https://doi.org/10.2307/j.ctvc77mp9.30

Sennett, Richard. 2008. The Craftsman. New Haven: Yale University Press.

Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., ... & Virk, G. (2023). Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 723-741).

Silva, S., & Kenney, M. (2019). Algorithms, platforms, and ethnic bias. Communications of the ACM, 62(11), 37-39.

Sinnreich, A. (2018). Four crises in algorithmic governance. Annual Review of Law and Ethics 26: 181–190.

Siuda, P. (2021). Mapping digital religion: Exploring the need for new typologies. Religions, 12(6), 373.

Skeggs, B. (2015). Introduction: stratification or exploitation, domination, dispossession and devaluation?. The Sociological Review, 63(2), 205-222

Skeggs, B., & Loveday, V. (2012). Struggles for value: value practices, injustice, judgment, affect and the idea of class. The British journal of sociology, 63(3), 472-490

Smelik, A. & Lykke, N. (eds). (2008). Bits of Life: Feminism at the Intersections of Media, Bioscience, and Technology. University of Washington Press

Smith, J. Z. (1982). Imagining Religion: From Babylon to Jonestown. Chicago: The University of Chicago Press.

Socher, R., Manning, C. D., & Ng, A. Y. (2010, December). Learning continuous phrase representations and syntactic parsing with recursive neural networks. In Proceedings of the NIPS-2010 deep learning and unsupervised feature learning workshop (Vol. 2010, pp. 1-9).

Søndergaard, M. L. J., & Hansen, L. K. (2018, June). Intimate futures: Staying with the trouble of digital personal assistants through design fiction. In Proceedings of the 2018 designing interactive systems conference (pp. 869-880).

Spiers, J. (2000). New perspectives on vulnerability using emic and etic approaches. Journal of advanced nursing, 31(3), 715-721

Storm, J. A. J. (2021). Metamodernism: The future of theory. University of Chicago Press.

Strubell, E., Ganesh, A., & McCallum, A. (2020, April). Energy and policy considerations for modern deep learning research. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 09, pp. 13693-13696).

Stuckrad, Kocku von. (2013). Discursive Study of Religion: Approaches, Definitions, Implications. Method and Theory in the Study of Religion 25(1): 5–25.

Sumiala, J. (2022). Digital Religion: A Methodological Approach. In the Oxford Handbook of Digital Religion (nd.). Heidi Cambell & Pauline Hope Cheong (eds.). https://doi.org/10.1093/oxfordhb/9780197549803.013.24

Sumiala, J. (2023). Presidential Address for the ismrc Virtual Conference 2021. Why Structures (Still) Matter in the Age of Hybrid Media: The Case of Digital Religion. Journal of Religion, Media and Digital Culture 12 (2023) 118–128. Brill.

Sweeney, L. (2013). Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising. Queue, 11(3), 10-29.

Taira, T. (2013). 'Making Space for Discursive Study in Religious Studies'. Religion 43 (1): 26–45. http://dx.doi.org/10.1080/0048721X.2013.742744

Taira, T. (2014). Media and the Nonreligious. In Religion, media, and social change (pp. 110-125). Routledge.

Taira, T. (2020). Media and communication approaches to leaving religion. Brill Handbooks on Contemporary Religion, 335.

Taira, T. (2022). Taking 'Religion' Seriously: Essays on the Discursive Study of Religion. Brill

Tamkin, A., et al. (2023). Evaluating and Mitigating Discrimination in Language Model Decisions. arXiv. https://doi.org/10.48550/arxiv.2312.03689

Tanner, D. (1979). What's in a Frame? Surface Evidence for Underlying Expectations. In: Roy Freedle (ed.), New Directions in Discourse Processing. NJ: Norwood.

Taylor, C. (1994). Multiculturalism: Examining the politics of recognition. Princeton University Press

Tegmark, M. (2018). Life 3.0: Being human in the age of artificial intelligence. Vintage.

Thelwall, M., Goriunova, O., Vis, F., Faulkner, S., Burns, A., Aulich, J., . . . D'Orazio, F. (2016, 11). Chatting through pictures? A classification of images tweeted in one week in the UK and USA. Journal of the Association for Information Science and Technology, 67 (11), 2575–2586. http://doi.wiley.com/10.1002/asi.23620

Thrift, N. (2004). Movement-space: The changing domain of thinking resulting from the development of new kinds of spatial awareness. Economy and society, 33(4), 582-604.

Turian, J., Ratinov, L., & Bengio, Y. (2010, July). Word representations: a simple and general method for semi-supervised learning. In Proceedings of the 48th annual meeting of the association for computational linguistics (pp. 384-394).

Twenge, J. M. (2019). More time on technology, less happiness? Associations between digital-media use and psychological well-being. Current Directions in Psychological Science, 28(4), 372-379.

Twenge, J. M., Haidt, J., Lozano, J., & Cummins, K. M. (2022). Specification curve analysis shows that social media use is linked to poor mental health, especially among girls. Acta psychologica, 224, 103512. University Press, Edinburgh.

Uro, R. (2016). Ritual and Christian Beginnings: A Socio-Cognitive Analysis. Oxford University Press.

Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social media affordances for connective action. MIS quarterly, 41(4), 1179-1206.

Valaskivi, K. (1995). Wataru Seken Wa Oni Bakari: Mothers-in-law and Daughters-in-law in a Japanese Television Family Drama. University of Jyväskylä.

Valaskivi, K. (2012). Dimensions of Innovationism. In P. Nynäs, M. Lassander, & T. Utriainen (eds.), Post-Secular Society (129-156). Transaction Publishers. https://doi.org/URN:NBN:fi:uta-201211121076

Valaskivi, K. (2020). The contemporary faith of innovationism. In E. Bell, S. Gog, A. Simionca, & S. Taylor (eds.), Spirituality, Organisation and Neoliberalism: Understanding Lived Experiences (171-193). Edward Elgar.

Valaskivi, K., & Sumiala, J. (2023). "You Will Never Hear Me Mention His Name": The (Im) possibility of the Politics of Recognition in Disruptive Hybrid Media Events. International Journal of Communication, 17, 1330-1347.

Van Leeuwen, T., & Jewitt, C. (Eds.). (2001). The handbook of visual analysis. Sage.

Van Rees, K., & Van Eijck, K. (2003). Media repertoires of selective audiences: The impact of status, gender, and age on media use. Poetics, 31(5-6), 465-490.

Visala, A. (2016). explaining Religion at Different levels: from fundamentalism to Pluralism. In The Roots of Religion (pp. 55-74). Routledge.

Visala, A. (2023). Conjunctive Explanations: A Response. In Conjunctive Explanations in Science and Religion (pp. 39-40). Routledge.

Wakefield, J. (2021 May 26). AI emotion-detection software tested on Uyghurs. BBC News. Read 29.4.2024. https://www.bbc.com/news/technology-57101248

Wang, L. (2020). The three harms of gendered technology. Australasian Journal of Information Systems, 24.w

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of personality and social psychology, 114(2), 246.

Webb Williams, N., Casas, A., & Wilkerson, J. D. (2020). Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification (1st ed.). Cambridge University Press.

White, C. (2021). An Introduction to the Cognitive Science of Religion: Connecting Evolution, Brain, Cognition, and Culture. Routledge.

Whitehead, D. (2015). The evidence of things unseen: Authenticity and fraud in the Christian mommy blogosphere. Journal of the American Academy of Religion, 83(1), 120-150.

Wilson, M. (2015). Paying attention, digital media, and community-based critical GIS. Cultural Geographies, 22(1), 177–191. https://doi.org/10.1177/1474474014539249

Ylikoski, P., & Kokkonen, T. (2009). Evoluutio ja ihmisluonto. Gaudeamus.

Ysseldyk, R., Matheson, K., & Anisman, H. (2010). Religiosity as identity: Toward an understanding of religion from a social identity perspective. Personality and social psychology review, 14(1), 60-71.

Ytre-Arne, B., & Moe, H. (2021). Folk theories of algorithms: Understanding digital irritation. Media, Culture & Society 43(5): 807–824.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. W. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. arXiv preprint. arXiv:1707.09457

Ziewitz M (2016) Governing algorithms: Myth, mess, and methods. Science, Technology, & Human Values 41(1): 3–16.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: Public Affairs.