The Impact of Makeup on Automated Face Recognition and Age Estimation Models: Implications for Alcohol Sales Compliance in Supermarkets

Frank van der Velde   
Faculty Tech  
Hogeschool van AmsterdamAmsterdam, Nederland  
frankvdv1998@gmail.com

*Abstract*— Self-checkout systems are increasingly used in supermarkets to improve efficiency and reduce waiting times. However, when selling age-restricted products such as alcohol, these systems still rely on manual verification if a customer isn’t instantly identifiable as being old enough. Automated age estimation models based on facial recognition have been proposed as a potential solution to support compliance and reduce staff workload. While promising, such models face challenges when facial appearance is altered, for example through the use of makeup. Makeup can change perceived age by enhancing or masking facial features, which may affect the accuracy of age estimation systems.

This research aims to investigate the impact of makeup on the reliability of automated age estimation. First, a literature study is conducted to identify known effects of makeup on facial recognition and age estimation models. Next, two models are tested on a dataset containing images of the same individuals with and without makeup. Model performance is evaluated using established metrics, such as mean absolute error (MAE), to determine whether makeup increases estimation errors or causes systematic biases. The results will provide insight into whether makeup poses a significant risk for misclassification in age verification at self-checkout systems, and what this means for compliance in alcohol sales as well as the fairness of such systems.

Keywords— Age Estimation, Facial Recognition, Makeup Bias, Computer Vision, Alcohol Compliance, Apparent Age

# Introduction

Self-checkout registers are becoming increasingly common in Dutch supermarkets, with Albert Heijn leading in their adoption to improve efficiency and reduce waiting times [1]. For customers this offers convenience, while for employees it reduces the need to staff traditional checkout counters. However, the sale of age-restricted products such as alcohol introduces challenges. Dutch law requires strict age verification, and government inspections show that compliance is still problematic. Between July 2023 and June 2024, the Dutch Food and Consumer Product Safety Authority (NVWA) reported widespread failures in age checks across retail, with many stores making mistakes when selling alcohol [2]. The fines for violations are high, making compliance a priority for supermarkets. Albert Heijn publicly states its commitment to responsible alcohol sales, for example through its *NIX18* policy [3].Yet research shows that in practice, checks often fail or are incomplete, both in stores and in online deliveries [4].

Because manual verification at self-checkouts is time-consuming and error-prone, automated solutions have been proposed. Facial recognition combined with age estimation algorithms could theoretically streamline alcohol sales by reducing reliance on human judgment. However, such systems are not yet widely used in Dutch retail, in part due to strict privacy laws and ethical concerns about biometric surveillance [1]. Nonetheless, the technology is under active development worldwide, and it is realistic to expect supermarkets to consider its implementation in the future. This raises important questions about fairness, accuracy, and potential vulnerabilities of automated age estimation.

One overlooked factor is the impact of makeup. Makeup can significantly alter the apparent age of individuals by concealing wrinkles, enhancing facial contrast, or otherwise changing visual cues used for age estimation. Research shows that cosmetics can not only shift perceived age but also reduce the reliability of computer vision algorithms [5]. While this has been documented in controlled biometric settings, its implications for supermarket age verification have not yet been studied. From an ethical perspective, it is important to ensure that people wearing makeup are treated fairly and not systematically misclassified. From a security perspective, it is equally important to investigate whether makeup could be used deliberately to mislead AI models, for example by making underage individuals appear older.

This study therefore examines the research question: *“What is the impact of makeup on automated age estimation models, and what does this mean for compliance in alcohol sales in supermarkets?”* To answer this, the project combines a literature review with experiments testing two age estimation models on images of the same individuals with and without makeup. The goal is to determine whether makeup increases prediction errors or introduces systematic bias, and to assess the potential consequences for the reliability and fairness of future AI-based age verification systems in the Dutch retail context.

# Literature Review

## **Impact of Makeup on Facial Recognition and Age Estimation**

The influence of makeup on facial recognition and age estimation has been a subject of interest in computer vision research. Studies have demonstrated that makeup can significantly alter the perception of facial features, affecting the accuracy of automated systems like has been demonstrated in the case of age-induced makeup attacks [6]. For instance, research indicates that facial cosmetics can confound automated gender and age estimation algorithms by modifying facial features such as skin texture, wrinkle concealment, and facial contouring [5]. Additionally, makeup can lead to overestimation of age in younger individuals and underestimation in older individuals, as it enhances features associated with youthfulness while masking signs of aging [7].

## Gaps in Existing Literature

While existing studies provide valuable insights into the effects of makeup on facial recognition, several gaps remain. Notably, there is limited research on the impact of extreme makeup styles, such as drag makeup, grime and horror makeup on automated age estimation systems. These makeup styles often involve exaggerated features, completely changing and hiding features and stylized applications that may significantly differ from conventional cosmetic practices. Existing literature does not comprehensively address how such extreme makeup styles affect the performance of age estimation models.

Furthermore, the specific context of alcohol sales in supermarkets has not been extensively studied in relation to makeup's impact on age estimation. While general studies on makeup and facial recognition exist, there is a lack of research focusing on the implications for compliance in age-restricted product sales, particularly in automated retail environments.

## Need for Current Research

This research is essential to bridge the existing gaps and address the evolving challenges in automated age verification systems. As supermarkets increasingly adopt self-checkout systems, ensuring accurate and fair age estimation becomes critical, especially for age-restricted products like alcohol. Understanding how makeup, including extreme styles like drag makeup, influences automated age estimation will inform the development of more robust and equitable systems. Moreover, this study will contribute to the broader discourse on demographic biases in AI systems and their implications for privacy and fairness in automated retail settings. While also giving an insight in the inclusivity in the case of people with heavy make up styles like drag makeup.

# Research Methodology

## Aim and Research Design

The aim of this study is to investigate how different types of makeup ranging from everyday cosmetics to extreme applications such as drag and theatrical horror makeup affect the reliability of automated age estimation models in supermarket self-checkout contexts. The research design combines a literature review with experimental evaluation using facial image datasets.

The study combines two approaches a literature review of existing research on makeup-induced biases in computer vision models, and an experimental analysis using publicly available and self-compiled datasets of faces with and without makeup. Particular attention will be paid to whether makeup systematically increases estimation error and whether extreme styles such as drag makeup, horror makeup can be used to deliberately mislead models or lead to this unintentionally.

## Data collection and Datasets

The experimental analysis will draw from three main sources of data. The first is the Makeup Detection Dataset from Kaggle [8], a small paired dataset containing real before-and-after makeup images. This dataset allows for direct comparison of the same individuals with and without cosmetics under everyday conditions.

The second is the FFHQ-Makeup Dataset from HuggingFace [9], a large-scale dataset where synthetic makeup has been applied to high-quality real facial images from the Flickr-Faces-HQ (FFHQ) collection. This dataset enables controlled testing across a wide variety of styles while maintaining consistency in base images.

Finally, additional curated data will be collected from publicly available and fair-use sources, specifically high-quality images of drag makeup and horror effects makeup from media such as *RuPaul’s Drag Race* and *The Boulet Brothers’ Dragula*. These examples provide test cases of exaggerated or deliberately transformative makeup styles that may challenge model robustness.

All datasets will be pre-processed through normalization and resizing to match model input requirements, and care will be taken to balance the datasets between makeup and non-makeup conditions for fair evaluation.

## Research Approach and Analysis

The methodology consists of three phases. First, an extensive literature review will be conducted on the effects of makeup in face recognition and age estimation. Prior studies, such as Chen et al. on the impact of facial cosmetics on automatic gender and age estimation algorithms [5], and Clapes et al. on demographic and cosmetic bias in real age estimation [7], will provide a theoretical foundation for understanding how makeup influences both human and algorithmic perception of age.

Second, a convolutional neural network will be implemented and trained on an open-source dataset designed for age estimation, such as UTKFace. After pretraining on this large dataset, the model will be tested on images with paired conditions (before and after makeup). Performance will be evaluated using standard metrics, including Mean Absolute Error (MAE) to capture prediction accuracy, as well as threshold-based accuracy around critical legal boundaries, particularly at 18 and 25 years.

Third, supplementary tests will be conducted on curated extreme makeup examples, including drag and theatrical aging effects. These analyses will explore whether such applications significantly distort model predictions and whether they could act as adversarial inputs capable of bypassing automated age verification systems.

By comparing baseline performance on natural faces with results under everyday, synthetic, and extreme makeup conditions, the study will generate both quantitative results (error rates and prediction deviations) and qualitative insights (potential vulnerabilities and fairness concerns). The findings will be interpreted in relation to the ethical and legal challenges of deploying AI-based age verification in supermarkets, where both underestimation and overestimation of age can have serious consequences.

# Planning

1. Planing for research proposal

| Week | Date | Tasks |
| --- | --- | --- |
| 2 | 8 sept – 14 sept | * Choosing research topic * Formulating research questions * Write introduction * Choose research method * Find useful literature and data |
| 3 | 15 sept – 21 sept | * Start of literature research * Finish research proposal   ***19 september: Submit research proposal*** |
| 4 | 22 sept – 28 sept | * Prepare data sets for testing research question (various types of makeup) |
| 5 | 29 sept – 5 okt | * Setup face detection model * Setup age estimation model |
| 6 | 6 okt – 12 okt | * Start experiments to gather results |
| 7 | 13 okt – 19 okt | * Evaluate results |
| 8 | 20 okt – 26 okt | * Analyse results and add conclusion to research |
| 9 | 27 okt – 2 nov | * Gather feedback on research and finalize |
| 10 | 3 nov – 9 nov | * Add relevant parts of research paper to group rapport |

1. Planning with which tasks I would like to do each week to finish the research before the deadline.

# Conclusie / Hypothesis

This study hypothesizes that everyday makeup will cause modest deviations in age estimation, typically amounting to a few years, without preventing successful facial identification. In contrast, more extreme cosmetic styles—such as drag makeup, theatrical aging effects, and horror-inspired applications—are expected to produce much larger shifts in age predictions, in some cases deliberately or unintentionally misleading automated systems. While facial recognition is anticipated to remain effective under most cosmetic conditions, heavily transformative makeup that alters or conceals fundamental facial features may introduce uncertainty or misclassification in age estimation models.

Given these limitations, the study further assumes that the integration of a complementary makeup detection model could improve system reliability. Such a model could serve as a safeguard by triggering a manual check whenever high deviations or extreme cosmetic applications are detected, thereby reducing the risk of compliance failures in supermarket self-checkouts.

# References

|  |  |
| --- | --- |
| [1] | Avrotros, ""Mag dit zomaar?" Vragen over camera’s bij de zelfscankassa," 23 05 2025. [Online]. Available: https://radar.avrotros.nl/artikel/mag-dit-zomaar-vragen-over-cameras-bij-de-zelfscankassa-61602. |
| [2] | Nederlandse Voedsel- en Warenautoriteit, "Inspectieresultaten Alcoholwet juli 2023 - juni 2024," 2024. [Online]. Available: https://www.nvwa.nl/onderwerpen/alcoholverkoop/inspectieresultaten/2024/alcoholwet-juli-2023-juni-2024. |
| [3] | Albert Heijn, "Duurzaamheidsverslag," 2022. [Online]. Available: https://duurzaamheidsverslag.ah.nl/2022/nix18. |
| [4] | Nederlandse Voedsel- en Warenautoriteit (NVWA), "Factsheet Online alcohol bestellen door jongeren 2024," Breuer & Intraval / NVWA, 2025. |
| [5] | C. a. D. A. a. R. A. Chen, "Impact of facial cosmetics on automatic gender and age estimation algorithms," in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, Lisbon, Portugal, 2014. |
| [6] | K. a. M. Z. a. M. S. Kotwal, "Detection of Age-Induced Makeup Attacks on Face Recognition Systems Using Multi-Layer Deep Features," IEEE, 2019. |
| [7] | E. B. Gavas, C. Hegde, N. Memon and S. Banerjee, "DiffClean: Diffusion-based Makeup Removal for Accurate Age Estimation," arXiv, 2025. |
| [8] | tapakah68, "Makeup Detection Face Dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/tapakah68/makeup-detection-dataset. [Accessed 19 9 2025]. |
| [9] | X. Yang, S. Ueda, Y. Huang and T. Akiyama, "FFHQ-Makeup: Paired Synthetic Makeup Dataset with Facial Consistency Across Multiple Styles," CyberAgent AI Lab; Keio University, 2022. |
| [10] | BNNVARA, "Kwart consumenten zelfscan vergeet producten af te rekenen," 17 02 2024. [Online]. Available: https://www.bnnvara.nl/kassa/artikelen/kwart-consumenten-zelfscan-vergeet-producten-af-te-rekenen. |
| [11] | Q&A Retail, "1-op-5 consumenten vindt diefstal in supermarkt onder omstandigheden toelaatbaar," 4 2024. [Online]. Available: https://www.qanda.nl/publications/1-op-5-consumenten-vindt-diefstal-in-supermarkt-onder-omstandigheden-toelaatbaar#:~:text=Zelfscankassa%20heeft%20bij%20meerderheid%20de,voorkeur%20boven%20de%20gewone%20kassa.. |
| [12] | A. Clapes, O. Bilici, D. Temirova, E. Avots, G. Anbarjafari and S. Escalera, "From Apparent to Real Age: Gender, Age, Ethnic, Makeup, and Expression Bias Analysis in Real Age Estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Salt Lake City, UT, USA (CVPR 2018 location), 2018. |

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being publish**

**Title, authors and affiliations  
• Abstract  
• Keywords  
• Introduction:  
– Research Topic (summary/highlights), Problem, Purpose, Question(s)  
• Literature review or theoretical framework:  
– Research Topic (extended)  
• Research Methodology  
• Planning  
• Conclusion  
– Final remarks  
• References**