Debunking Instagram's Algorithm-Sugarcoating

Gross, Tobey; Michaud, André; Zerrouki, Yassine; Hamood, Asaad

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zentrum für
Medienpsychologie und
Verhaltensforschung

Open Access
CC BY 4.0

***Citation**: Gross, T., Michaud, A., Zerrouki, Y., Hamood, A. (2024). Debunking Instagram's Algorithm-Sugarcoating. Zentrum für Medienpsychologie und Verhaltensforschung, 05/2024.*

Corresponding Authors: *Tobey Gross[1], André Michaud[2], Yassine Zerrouki[3], Asaad Hamood[4]*

Contact Information: *kontakt@zemv.org*

Type: *Sentiment Analysis, Review of Literature*

Field: *Psychology, Economics, Information Technology*

Research Funding: *none*

Date: *May 31, 2024*

Conflict of Interest: *[5]*

[1] ZeMV
[2] Service de Recherche Pedagogique Inc.
[3] Mohammed I University, Oujda, IDS
[4] Basrah University

Debunking Instagram's Algorithm-Sugarcoating

Abstract

In 2021 and 2023, there were two distinct official statements published on the mechanics behind the algorithm responsible for Instagram's content feed curation, where the 2023 publication, named *"Instagram Ranking Explained"* is presented as an updated and expanded version of the one in 2021, named *"Shedding More Light on How Instagram Works"*. This paper examines the statements made therein by Adam Mosseri, the head of Instagram who authored the publications, comparing them with insights from contemporary literature and investigating the sentiment of the publications through a mixed-method sentiment analysis. The analysis aims to show, that the statements present the algorithm in a particularly positive light, downplaying and largely ignoring potential detriments. Statements are examined for *pseudo-transparency*, providing a veneer of openness, while concealing deeper economic motives through obscuring practices like data extraction and engagement maximization, especially in mid of rising criticism towards social networking sites.

Keywords: Algorithmic Personalization, Digital Media, Social Media, Content Curation, Framing, Sentiment Analysis, Recommender Algorithms

Introduction

Instagram assumes a major role in the world of social media. It manages to sustain users' interest and attention in manifold ways, one of which is showing those sorts of contents, that are most engaging. But in what ways does Instagram determine what contents to display to individuals? What makes these pictures or videos so special that they can capture attention reliably, on a large scale and perpetually? The answer lies within an algorithm that controls what people see when they log into their accounts and the order in which contens are ranked.

What sounds relatively unsuspicious at first, and seems to be a likeable feature for navigating through endless arrays of contents, has grown to a major concern in the global tech industry, perhaps the most worrying issue in modern digital economy overall. While algorithmic content curation is indeed a sophisticated manner of tailoring user experiences, it raises fundamental questions about transparency, corporate motives and depth of data mining and surveillance. Because in reality, users are subjected to extensive surveillance, tracking and recording every interaction. The depth of data extraction is staggering, and in a fast-paced digital ecosystem, behavioral data functions as a commodity; enabling big data companies to craft disturbingly accurate profiles of each individual, that are sold to the highest bidder and used to predict and carefully manipulate behavior.

It is still highly challenging for scientists to unveil the extent of these sinister motives behind seemingly benign digital architectures, because none of the big enterprises behind those algorithms are willing to disclose their data, practices or mechanisms – for obvious reasons. However, the ethical question about what many nowadays refer to as *Surveillance Capitalism* has apparently a unison answer: users are being exploited with little regard to digital autonomy, ethics and consequences.

All the more, communicating the opposite of what is really happening behind the curtain might be perceived as a blatant attempt to throw concerned individuals off track and strengthen their bonds and trust into the platform. We claim, that this is the exact reason for two publications authored by Instagram's CEO and disclosed on their own website, at a time when the global conversation around data protection and immoral practices in big data enterprises is becoming louder. Mosseri tries to reassure the reader, that Instagram's users, their experience and well-being is at the heart of what Instagram and its algorithms are after. We uncompromisingly challenge that notion by presenting an array of avaliable research, that exposes, how algorithmic personalization primarily serves to keep users engaged and extracting their behavioral *surplus* in order to monetize it. Furthermore, the following probe quantitatively and qualitatively examines the sentiments communicated by the 2023 publication and repositions them as sugarcoated half-truths meant to disguise and distract from Instagram's algorithmic intentions, which, as we will find, are no different from any of the *attention economy* industry's standards.

Quantitative Sentiment Analysis

For the quantitative analysis, we employed three distinct methods: TextBlob, VADER and Flair. Each of those methods offers a unique approach to sentiment analysis, with its own emphasis and analysis methodology. Through this multifaceted approach, we obtained a nuanced impression of the conveyed sentiment in the publication.

The methodology included downloading the full text, cleaning it and dividing it in suitable coherent segments, that would preserve the context. Moreover, segmentation allowed for granular insight in our analysis, providing us with the opportunity to identify shifts or variations in sentiment. Lastly, the collection of more data points offered a richer set of metrics to analyze, increasing validity and reliability of our analysis.

Through application of three different methods (*which are explained in more detail below*), we were able to cross-validate sentiment scores and ensure consistency across different analytical methods. In a single-method analysis, nuances in sentiment might have been overlooked or not captured due to noise etc. Hence, a multi-method analysis appeared more robust for our investigation.

Below, there is an overview of the results of each analysis.

Segment	TextBlob Polarity	TextBlob Subjectivity	VADER Negative	VADER Neutral	VADER Positive	VADER Compound	Flair Sentiment	Flair Confidence
Initial Paragraph	0.30142	0.550568	0.0	0.809	0.191	0.9358	NEGATIVE	0.968971
Second Larger Segmen	0.25303	0.579545	0.0	0.843	0.157	0.9766	POSITIVE	0.653849
Third Larger Segment	0.115463	0.493416	0.029	0.823	0.148	0.9985	POSITIVE	0.553345
Fourth Larger Segmen	0.1348	0.542008	0.044	0.826	0.13	0.9954	POSITIVE	0.978636
Fifth Larger Segment	0.152114	0.459408	0.029	0.794	0.177	0.9991	POSITIVE	0.778824
Final Segment	0.247598	0.46656	0.021	0.82	0.16	0.9923	NEGATIVE	0.986306

Fig. 1: Analysis Matrix

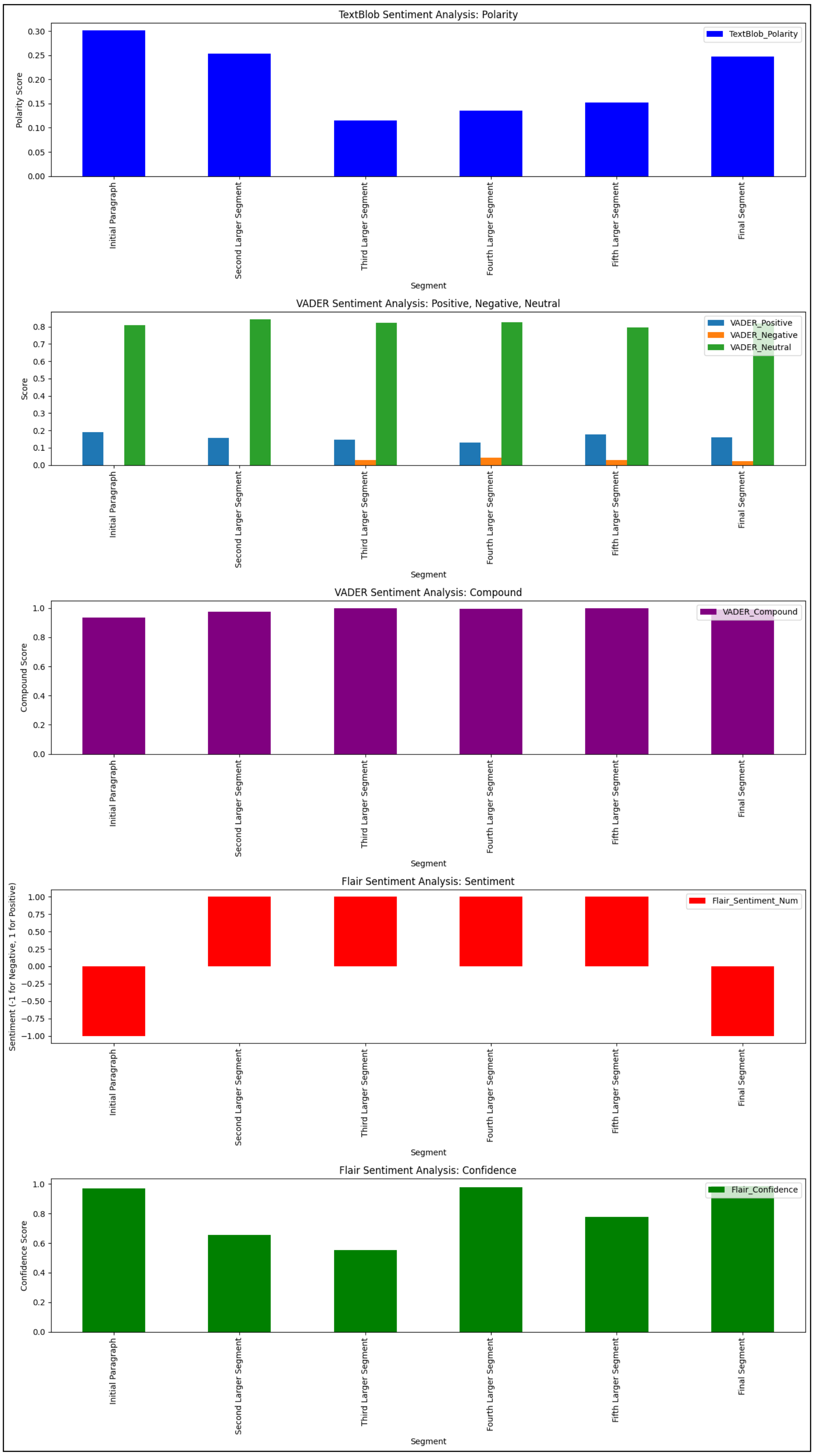

Fig. 2: Bar Chart Comparison

Fig. 3: Heatmap Distribution

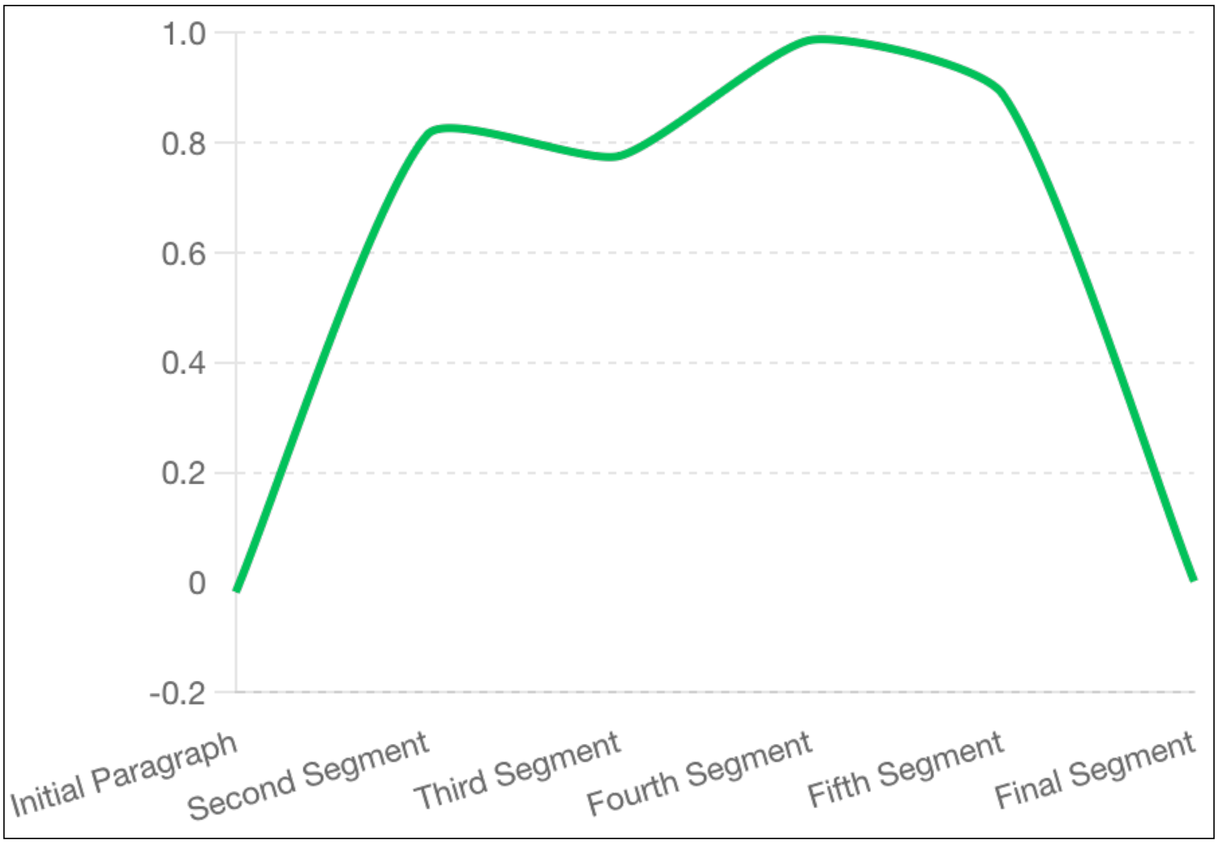

Fig. 4: Overall Sentiment Score by Segment (VADER Compound, Flair Normalized)

Legend and Discussion

TextBlob is a simple yet effective tool for natural language processing tasks. Its two key metrics are *polarity* and *subjectivity* score, where the former measures the degree of positivity or negativity on a scale from -1.0 to 1.0, with higher values indicating a more positive sentiment, and the latter indicates the amount of personal opinion versus factual information on a scale from 0.0 to 1.0, with higher values indicating more subjective content.

VADER (Valence Aware Dictionary and sEntiment Reasoner) is particularly attuned to the sentiment in social media contexts and is capable of capturing nuances, such as exclamation marks and capitalization. VADER provides four metric scores: positive, negative, neutral and a compound score, which calculates the aggregate of the overall sentiment into a single metric on a scale from -1.0 (most negative) to 1.0 (most positive). Consequently, higher values in VADER Negative indicate higher proportions of negative sentiment, higher values in VADER Positive indicate higher proportions of positive sentiment and VADER neutral indicate higher proportions of neither positive nor negative, but rather neutral sentiment. The aggregate of all three sums up to 1.0.

VADER Compound score is a normalized, weighted composite score that takes into account the proportion of negative, neutral and positive sentiment. It is thus not a direct average, which means due to weighting and normalization in the VADER algorithm, the values of each distinct metric are calculated against each other. Hence, in texts with predominantly neutral sentiment, a small proportion of positive score can result in a significantly higher composite score, if the counterweight metric (negative sentiment) is nearly absent. This way of calculation makes VADER Compound a highly expressive metric considering both the proportion and intensity of sentiment scores across a text.

Flair is a more sophisticated deep learning sentiment analysis tool with a binary sentiment classification and a confidence score that indicates the certainty of the classification on a scale from 0.0 to 1.0. In our visualization, we employed a numeric value of -1.0 for negative and 1.0 for positive for the binary classification.

As for the discussion of the results, starting with TextBlob, the calculated polarity scores exhibit consistently positive results, ranging from **0.115** to **0.301**, which indicates an overall positive sentiment. The subjectivity scores range from **0.459** to **0.58**, which suggests a relatively balanced mix of subjective and factual / neutral content throughout the publication.

VADER provided an overwhelmingly positive sentiment regarding the range of the compound score from **0.9358** to **0.9991**. While positive sentiment scores range from **0.13** to **0.191**, negative scores range from **0.0** to **0.044** and neutral scores range from **0.794** to **0.843**, the compound score factors in a weighted normalization of all three metrics, which explains that the high value in compound is based on predominantly neutral sentiment with tendencies to positive rather than negative throughout the publication.

The results of the analysis using Flair were slightly more nuanced. While the majority of segments was identified as positive, with confidence levels varying between **0.553** and **0.978**, two segments were identified as negative with high confidence scores **0.969** and **0.986**. This suggests, that Flair, which is capable of contextual analysis, identified specific phrases or contexts, that it could interpret as negative. Notably, those were the first and the last segment of the publication. Flair uses a neural network model, which is capable of capturing context and subtleties better, but might also be influenced by specific phrases that carry negative sentiment, even if the overall text is positive. Since in the first segment, the author addresses "*a lot of misconceptions*" and the final segment discusses "*shadowbanning*" (a theory experienced very negatively by users), where the term alone carries a notion of censorship and lack of transparency, this might deliver an explanation for Flair's negative classification.

Finally, the line plot featuring the overall sentiment score by segment (*Fig. 4*) is a compound metric we calculated from VADER Compound scores and Flair confidence, both derived directly from the sentiment analysis results. For this metric, we normalized the Flair numeric by the sentiment type. Positive sentiment was considered 1, negative was considered -1 to reflect the negative impact.

Thus, the general calculation is:

$$\text{Compound Sentiment Score} = \frac{\text{VADER Compound} + (\text{Flair Confidence} \times \text{Flair Sentiment Numeric})}{2}$$

Example calculation for the first segment (=*"initial paragraph"*):

VADER Compound: **0.9358**
Flair Confidence: **0.968971**
Flair Sentiment: **Negative (converted to -1)**

$$\text{Compound Sentiment Score}_{(Segment 1)} = \frac{0.9358 + (0.968971 \times -1)}{2} = \frac{0.9358 - 0.968971}{2} = -0.0165855$$

Justification for this metric:

Not only does this metric include the overall sentiment provided by VADER, but it also reflects the sentiment classification by Flair and likewise, its certainty. Moreover, the averaging process ensures that both VADER's and Flair's classifications are equally considered, and Flair's binary classification impacts the score through normalization to 1 and -1. Finally, the reason to choose VADER and Flair over Textblob was the latter's separation of subjectivity and sentiment scoring, which does not directly exhibit a composite sentiment score and is thus not as straightforward as the former two. In addition, VADER is tuned to social media contexts, and Flair is sophisticated through its contextual sensitivity, both of which are features that are useful to take into account in the calculation of an aggregate overall score. The need for such integrated analyses voted against an integration of Textblob in this particular metric.

In conclusion, our analysis of Adam Mosseri's 2023 publication shows an overarching positive sentiment. The combination of metrics and analyses conducted provides a robust multidimensional proof of the conveyed sentiment and solidifies the assumption, that, despite slight variations across the publication, the content is tailored to reinforce the narrative of user benefit and intentions to enhance users' experience on the platform, portraying the employment of algorithmic content curation as a choice that is beneficial to individuals, deliberately conveying a sense of positive impact and "*nothing to worry about*".

Qualitative Sentiment Analysis

For the qualitative analysis of the publications sentiment, we systematically examined the language and framing techniques used throughout the text.

1. Introduction with transparency claim

Quote: "*We want to do a better job of explaining how Instagram works. There are a lot of misconceptions out there, and we recognize that we can do more to help people, especially creators, understand what we do.*", "*...we recognize*", "*...shed more light*", "*...response to feedback*", "*...help improve the experiences*"

Analysis: This opening statement frames Instagram strongly as a user-focused and transparent platform, using phrases like "a better job of explaining" and "help people" (especially before a comma-separation), that specifically portray a commitment to openness. Furthermore, the term "misconceptions" subtly shifts the blame to the public, who obviously do not understand Instagram's operations. Thereby, all further claims of any potential wrongdoing are already to be easily subsumed under "misconceptions". Furthermore, employing the quoted highly positively connoted phrases present Instagram, from the introduction on, as a user-centered, user satisfaction-committed, input and criticism-acknowledging platform, that puts its emphasis on continuous evolution based on user input and creates a welcoming and inclusive sentiment, as expected by a transparent institution, almost framing Instagram as similar to a non-profit entity.

2. Algorithm and Personalization

Quote: "*Instagram doesn't have a singular algorithm that oversees what people do and don't see on the app. We use a variety of algorithms, classifiers, and processes, each with its own purpose.*"

Analysis: Firstly, by detailing the complexity of their systems, Instagram does appear sophisticated and thorough, fostering a particular sense of trust. Again, the "clarification" is introduced with the notion, that the statement is a hypothetical response to a "misconception", since it directly opens with a negation, instead of an affirmative sentence. Apart from that, the sentence ends, before another clarifying statement is made. Instagram doesn't (...) period. This suits the demystifying nature of the statement, reducing fear and scepticism about a "monolith" control mechanism, rather softening the public's concerns, in watering down the algorithm-based nature of content curation to carefully selected parts, that have an individualized experience for the user in mind, rather than a profit-oriented manipulative nature of content curation.

3. Customization and User Control

Quote: "*We rank things differently in these different parts of the app, and have added features and controls like Close Friends, Favorites and Following so you can further customize your experience.*", "*We want to make the most of people's time...*"

Analysis: Firstly, the emphasis on user control takes away the "purported" element of "remote-control" by algorithmic curation and creates a fictitious verbal distance to the actual purpose of the algorithm; instead, the reader is actively portrayed as the controlling element, exactly mirroring the mechanism behind algorithmic curation: "the user controls the algorithm" is the notion that is conveyed, rather than the algorithm being an actively manipulative element to content curation in the first place. Coupled with the image of eliciting the most out of users' time, the illustration of algorithm mechanics is further warped to reflect a sense of not only self-controlling one's content curation, but being supported by Instagram's implementation of algorithms. This reinforces agency and satisfaction, and likewise, it subtly portrays the platform itself as individually useful, since it helps "making the best of one's time" (*which can be seen as the very opposite to what is commonly referred to as "doomscrolling": the purposeless half-aware scrolling through content, while not realizing how much time passes, often leaving the user frustrated about their own unproductivity and waste of time*).

4. Community and Safety Emphasis

Quote: "*We always want to lean towards letting people express themselves, but when someone posts something that may jeopardize another person's safety, we step in.*"

Analysis: This dual-focus on self-expression and safety positions Instagram as a balanced platform that values individual freedom and community welfare. While this is a valuable approach, there is solely the content *creation* concerned in safety considerations, which can distract from the very thought, that content *curation* (that no user has an active influence on) might also pose detriment to community welfare.

5. Addressing Shadowbanning Concerns

Quote: "*Contrary to what you might have heard, it's in our interest as a business...*", "*...more we can do to increase transparency*"

Analysis: Opening the statement with another apparent "misconception", the choice of words in the paragraph is predominantly encouraging and conveys a sense of "collaboration" between Instagram and its users. By apparently aligning Instagram's business interests with users' success, the author creates a fictitious bond that is characterized by mutual trust and support, which is optimal for casting doubt on claims of manipulative practices and any "purportedly manipulative" practice.

6. Positive Word Association, Framing Techniques

Quote: "*...home base*", "*...friends*", "*...family*", "*...closest friends*", "*...discover*", "*...we do our best*", "*...you help improve the experience*" etc.

Analysis: Across the publication, there is an array of very positively associated terms and expressions, that evoke senses of comfort, safety, collaboration, openness, trust, intimacy, emotional attachment, connection, welcoming atmospheres, excitement, adventure, fun and joy.

All this positive vocabulary is intended to establish a highly positive association with the platform, intentionally integrating and including the individual as part of the evolution of the entire platform. Particularly the repeatedly reinforced sense of companionship between Instagram and the user, where the users are empowered to influence their own experience is a crucial linguistic tool to move the notion away from being a subject of manipulation in any way or form. This framing technique is consistently employed across the entire publication, continually reminding the reader of their highlighted influence and control, while diminishing the amount of control, that the algorithm has in the curation of content.

Furthermore, the language is aimed at creating a sense of belonging, where the platform's practices are "transparently" simplified (and, in addition, positioned alongside industry standards, through the statement "*With any ranking algorithm...*", which further solidifies the harmless, common and unsuspicious nature), suggesting openness, honesty, agency and first and foremost, always designed with the user's best interest in mind – thereby systematically downplaying potential conflicts of interest between the platform and its users and distracting from the own economic interests.

Conclusive Statement

Overall, the comprehensive sentiment analysis of Adam Mosseri's statements reveals a deliberate use of positive language and framing techniques to create a user-friendly image of Instagram as a platform, and attempts to convey transparency about its algorithmic practices. Both our quantitative and qualitative sentiment analyses revealed a predominantly positive sentiment, highlighting the strategic mitigation of concerns and criticisms, while systematically appealing to users' sense of trust. This portrayal tends to obscure the platform's economic motives and intentionally moves attention away from potentially manipulative, or generally negative aspects of its algorithms.

In the following section, we will critically review existing literature on algorithms and content curation, insights into industry standards on engagement maximization practices and general scientific knowledge on attention economy. Contrasting Mosseri's statements with those findings will provide more nuanced insights in how far the conveyed image is inaccurate and *shed more light on how Instagram works*.

Literature Review

According to Mosseri's portrayal, algorithms primarily serve as a tool working solely in favor of Instagram's users, designed to enhance the experience on the platform, by learning the user's preferences and displaying more content directly relevant to them. As we already established in the qualitative sentiment analysis, the publication in question thereby aims to position Instagram as a somewhat collaborative element to its users, aiming at no more than the most positive experience while using the app, and with no other motive to employ ranking and recommender algorithms than to help discover new relevant content, staying up to date with friends and family and to deliver its best version of itself. The way of overly positive illustration could almost make the reader forget for a moment, that Instagram pursues own economic interests, instead of being there to simply make its users' lives better.

Reality, however, tells us not only, that Instagram, parented by its owning company Meta Platforms, generates astronomical revenue, but also, that algorithms employed all across social media platforms these days, including Instagram, are indeed not *users' best friends* and serve the primary purpose of sustaining user attention, maximizing their engagement and doing their utmost to keep users on the app or website for as long as possible. The reason is simple: more time spent on the app means more targeted advertising displayed to each individual, which can in turn be directly translated into revenue for the platform.

Even through mere logic alone, one must already fall into doubt about the true intentions behind these algorithms. The notion, that a company of Instagram's magnitude would employ *the* most sophisticated sort of contemporary technology not for their own economic benefit, but rather for enhancing users' browsing experience is more than unrealistic; and even more so, raises the question as to how it generates financial revenue that market analysts expect to be north of USD 70 Billion in 2024 (WARC, 2023).

Investigations could reveal, that engagement mechanisms and algorithmic content curation in Instagram are strategically designed to maximize user engagement and sustain attention, which is the direct translation into economic benefit for the platform. Mosseri's assertion that Instagram's algorithms aim to personalize user experience contradicts findings that highlight the platform's focus on engagement metrics. For instance, studies show that likes and comments significantly influence content visibility (Purba & Yulia, 2021). This reveals a strategy centered around maximizing user interaction rather than merely enhancing user satisfaction. The more engagement users exhibit on the platform, the more advertising revenue can eventually be generated. Engagement metrics, such as comments, likes and shares play a critical role in the platform's use of predictive algorithms and enhancement of content visibility and user interaction. Through comprehensive extraction of distinct features, the investigation could support the notion, that Instagram's algorithmic architecture inherently aims to maximize and sustain attention and engagement (Tricomi et al., 2023). Especially in hindsight of the estimated USD 33.25 billion in 2022 (ibid.), it is difficult to rationalize away the economic motivation behind the algorithmic engagement-driven architecture. Contrary to Mosseri's statement that the algorithms are designed to help users make the most of their time, research clearly indicates that factors like image quality and posting time are optimized to sustain user attention (Wang et al., 2020). In turn, it can be assumed that the primary goal is to keep users engaged on the platform as long as possible, aligning with economic benefits for Instagram.

It could further be scientifically demonstrated, how deep learning models can be utilized to predict and enhance engagement on Instagram. By training personalized engagement prediction models on individual accounts, Wang et al. (2020) were able to show the potential for algorithms to sustain user engagement and attention through specific, highly engaging content. Their examination indicates, that the underlying strategy in Instagram's ecosystem supports a sophisticated data-driven approach to maximizing attention and prolonging engagement, aligning with and reflective of broad industry trends, rather than proprietary algorithm disclosure, as Mosseri's assertions might suggest. This research confirms the notion, that the economic incentives behind high engagement make for better visibility on the platform, that is directly tied to engagement-driven algorithms, which challenges Mosseri's portrayal of algorithms as tools for enhancing user experience.

In diffusion modeling, Purba et al. (2021) could show, that engagement is the most critical metric for realistic maximization of influence, suggesting, that Instagram's inherent architecture relies heavily on high user interaction. Through introduction of the *Engagement*

Grade (EG) and their diffusion models *IC-eg* and *LT-eg*, engagement metrics were incorporated to demonstrate, how engagement is the essential component of how content is propagated largely across Instagram, directly contradicting the portrayal of algorithmic curation for an improvement of individual user experience.

Contrary to Adam Mosseri's statements

> "*We want to make the most of people's time, and we believe that using technology to personalize everyone's experience is the best way to do that*" (Mosseri, 2023)

and

> "*Our intention is to help creators reach their audiences and get discovered so they can continue to grow and thrive on Instagram*" (Mosseri, 2023),

research on the mechanisms and algorithmic architecture clearly shows, that the design is built to surface content that maximizes engagement metrics, which can be seen as a strategy to keep users engaged for longer periods, rather than a mere focus on *helping* creators; likewise, rather than personalizing content curation for a better experience, the economic structure of Instagram, its practices and its sources of revenue clearly reveal, that prioritization of engagement metrics, such as likes, comments and shares foster frequent interaction, longer durations spent on the app and thereby directly align with Instagram's economic motives.

Findings from a 2021 study, conducted by Purba et al., substantiate these perspectives, because it could be found, that various features can be extracted to predict *Engagement Rates* and amplify combinations of favorable metrics to sustain user attention and increase engagement. This evidence further contradicts the image of algorithms as a user-friendly tool to discover more relevant content and supports the perspective of maximizing profitability first and foremost.

In this paper, we would like to address another, deeper layer of concern about how the statements made in the publication are vastly inaccurate, in the context of how users' behavior is being tracked and analyzed; the effort, precision and shocking accuracy with which algorithms can nowadays make estimations, classifications and predictions, is a topic that we are certainly not the first to mention. However, deliberately conveying the impression, that

> *there is merely some data collected for the purpose of getting to know the interests of the user better, in order to personalize the experience*

is so far from reality, that in the context of debunking the publicly shared notion of harmlessness and goodwill, they must be rectified, and to a large extent.

Shoshana Zuboff wrote in her 2019 seminal work *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, that at the time of the investigation and creation of her work, the depth and amount of extraction of individual behavioral data had been unprecedented in history, going as deep as classifying individual phrases and patterns of speech, which she labels the *behavioral surplus*, the meta-data or mid-level metrics, out of which classifications and predictions can be made. She clarifies, that while seemingly banal, through the amount and sophistication of data collection and analyses, including contemporary methodology, there is a depth in behavioral insights, that is unimaginable.

Citing big data scientist Michal Kosinski, she writes

> "*As Kosinski told an interviewer in 2015, few people understand that companies such as Facebook, Snapchat, Microsoft, Google and others have access to data that scientists would never be able to collect.*" (Zuboff, 2019, p. 261; Kosinski, 2015)

and further

> "*In his 2015 interview, Kosinski observed that 'all of our interactions are being mediated through digital products and services which basically means that everything is being recorded.' He even characterized his own work as 'pretty creepy': 'I actually want to stress that I think that many of the things that... one can do should certainly not be done by corporations of governments without users' consent.'*" (ibid.).

Delving deeper in Kosinski's research, accompanying this 2015 interview, we would like to draw attention to a 2015 study, that is concerned with entirely computer-based judgments on human personality, that are exactly based on such algorithmically extracted traits. They were solely based on Facebook-Likes in the investigation, yielding the following results:

After a number of x analyzed Facebook likes only, the computer did a significantly better and more accurate personality estimation and prognosis, than:

- a work colleague (x = 10)
- a friend (x = 70)
- a sibling or the subject's parents (x = 150)
- the subject's own spouse (x = 300).

While this (Youyou et al., 2015) serves as equally impressive and concerning evidence for the capability of behavioral prediction and the exploitation of gathered data in social media, we want to highlight, that this study does not serve as irrefutable evidence for Facebook's practices, and also not as evidence for how Instagram leverages user behavioral data.

However, it shows the vast potential of what can be done with such data, especially since the data-trail left behind on social media does not only consist of likes, but much more. Considering the accuracy that could be obtained from likes alone, it is fair to say, that through algorithms which continually monitor user behavior on a multi-scale basis, the potential in sophistication of behavioral prediction and manipulation can indeed be described as *creepy*.

Those insights are solidified by contemporary studies, that could reveal, that digital records about patterns of Facebook usage, or even as little as the language used in social media posts, are enough of a digital footprint to enable algorithms to extract highly detailed and accurate personality traits to build sophisticated personality profiles from them, upon which behavior can be precisely predicted. While users are unaware of those facts, the personal attributes that could be derived included sexual orientation, ethnicity, political and religious views, general personality traits, intelligence, overall hapiness, habits of substance abuse, parental separation, age and gender (Kosinski et al., 2013; Bachrach et al., 2012; Park et al., 2015; Zuboff, 2019).

Adam Mosseri himself makes several assertions on the depth of monitoring, and while making an effort to present their purpose as a benefit to Instagram's users, considering the former paragraph, we may cite some of there assertions with a healthy portion of doubt:

> "*Likes and comments are important signals for ranking content in Feed and Stories.*"
>
> "*We look at how often you interact with the person who posted, such as liking or commenting on their posts.*"
>
> "*We try to predict how likely you are to be interested in a photo or video based on past behavior.*"
>
> "*How long you spend looking at a post also helps us understand what is interesting to you.*"
>
> "*We consider your activity in Explore and Search to understand what you might be interested in.*"
>
> "*If you visit someone's profile or send them a message, it's a strong signal that you're interested in that person.*"
>
> "*We look at how often you share content, both publicly and privately.*"

While these admissions provide an overview of the monitoring depth implemented by Instagram, it is interesting to consider, that especially through the fourth statement, we can derive that there is real-time surveillance, extracting behavioral data as soon as the app is opened, which makes it obvious, that there is a continuous data-stream with every ever-so slight move, stored by Instagram and fed into the algorithm. If we consider the sophistication in behavioral predictions that could be derived from *likes alone*, and ten years ago, with less sophisticated information technology, then at this point, it must become obvious, how granular and accurate predictions and data processing potentially is today – and this *still* all from the perspective of the *obvious* and *official*, irrespective at this point of the potentially hidden data, methods and practices.

According to J. P. Titlow (2017), what makes Instagram decide about which content to show, out of the millions of options, is behavioral metadata, and irrespective of the content. That means, that the content of images is relatively unimportant for an algorithm in the decision how to rank the content. Behavioral metadata includes much more than, as Titlow puts it, "*if you like this, you'll like that*" logic: there is a highly complex interweb at play, well-hidden in the background, that extracts endless datapoints, in order to increase its users' overall engagement. According to Titlow, who argues Instagram "*is mining the multilayered social web between* users", this strategy worked out brilliantly: considering the inauguration of algorithmic personalization in the *Explore* feature of the app, two years after Facebook had acquired Instagram, experiments saw a surge of 400 % in engagement.

Considering the economic side of that medal, data extracted from *SignHouse, Business of Apps* and *Statista* suggest, that while Instagram was acquired by Facebook in 2012 for USD 1 billion, by late 2016 this skyrocketing engagement had already catapulted its valuation to USD 30 billion. Personalized algorithms fall well into that timeframe.

There is significant proof, that engagement maximization is at the heart of Instagram's business model, and certainly, sophisticated algorithms help well in extrapolating metrics obtained from user behavior. Mosseri's denial of content suppression to boost ad sales is contradicted by the platform's reliance on engagement metrics, which directly correlate with advertising revenue. Research solidifies that the economic structure of Instagram heavily relies on maximizing user engagement to drive profitability, challenging the notion that the algorithms are solely for user benefit, as we will show in the following paragraph.

Since parent company Meta, formerly Facebook, has an interest to extract utmost profits from its acquisition, the combined data-streams of Instagram and Facebook (*given an individual actually uses both*) would make for the ultimate resource of behavioral metadata (pun *not* intended) to create ever-so sophisticated profiles; remembering the need of no more than 300 Facebook likes to actually psychologically profile a person more accurately than their own spouse.

Reports of certain incidents let us safely assume, that this is exactly what happens, for instance by the account of T. Stenovec (2016), who argues, that Facebook and Instagram share user data among each other, to improve their personalization algorithms. According to Stenovec, this has been confirmed to *Tech Insider* by an Instagram spokesperson.

In her book, Shoshana Zuboff (2019) describes the complexity that lies in the data building the *surplus* in the context of the aftermath of the Cambridge Analytica scandal in 2018 – which in our context can be seen as one strong argument against the oversimplified and skewed portrayal of Instagram's algorithmic practices made by Mosseri (2023): as Zuboff explains, there is a vast asymmetry of knowledge between the intelligence Facebook gathers on its users and the knowledge thereof, obtained by its users themselves. Facebook argued, in the wake of criticism, that letting users freely access exactly this intelligence would "*require it to surmount 'huge technical challenges'*" (p. 453); furthermore, she states, that the data Facebook would provide to users did not include the data on the behavioral surplus gathered for the purpose of prediction products, eventually *sold* and employed for behavioral modification.

Instagram's algorithms are mechanisms that exert considerable control over the sort of content that is shown to users and in general, what content gains visibility, which highly benefits Instagram's commercial interests. Instead of solely enhancing the experience for users, the strategic interaction of actively shaping user behavior and maximizing visibility and engagement illustrates complex power dynamics and how maintenance of high engagement levels benefits the platform's profitability. The manipulation of content distribution on an individual scale has therefore been deemed a *visibility game* (Cotter, 2018).

Those findings align with the knowledge from investigations of Instagram's recommendation algorithms, that surfaced a significant influence of commercial optimization strategies, which, according to Mehlhose et al. (2021), did not differ significantly from those in other social networks: the clear priority is on engagement metrics that primarily drive commercial benefits. Jaakonmäki et al. (2017) came to a similar conclusion, when they established, that the algorithms embedded in Instagram, and generally in social media and its marketing ecosystems, are primarily geared towards the outcome of maximizing engagement and user attention. The implemented machine learning models prioritize features in their analyses of content, that amplify interaction and consequently, profitability for the platform.

Conclusive Statement

After careful review of the existing literature, Adam Mosseri's portrayal of Instagram's algorithmic practices, and especially their intentions, hold little to no credibility. While Mosseri publicly emphasizes user experience and creator support as motivation, evidence clearly contradicts those claims. It could be found, that Instagram's algorithms are meticulously designed to maximize user engagement and sustain their attention, predominantly to the content that algorithms curate for them. The latter is meanwhile far from chance or coincidence, instead it is carefully selected as a means to maximize profits. Factors such as likes, comments, image quality, and user history are permanently monitored in real-time and are strategically optimized to sustain prolonged interaction, well aligning with the platform's economic interests. In addition, as we demonstrated, the infrastructural demands in obtaining, retaining and processing the immense amounts of multilayered datapoints in a sheer astronomical complexity have not only the potential to make extremely, even worryingly accurate individual profile predictions and behavioral manipulations, they do also have to pay off reasonably to justify the tremendous effort. "Enhancing the experience" is therefore certainly not the primary motivation. Conclusions from the general advertising and social media marketing ecosystem confirm those perspectives, especially since it could be established, that the algorithmic architecture of Instagram's proprietary mechanisms does not significantly differ from others. These findings debunk the notion of a purely user-centric algorithmic design, revealing the underlying profit-driven motives and likewise, the deceptive nature of Mosseri's publication.

General Conclusion

As our research has been able to show, there is more to using algorithms than simply the best experience for the user. Our analysis reveals significant discrepancies between the image that Adam Mosseri likes to convey about Instagram's algorithmic practices and the realities. Both lie far from each other.

Our sentiment analysis demonstrates a deliberate use of positive language and framing techniques to create a user-friendly image of Instagram as a business, in attempting to intentionally steer the reader's attention away from the fact, that Instagram's – as any company's – primary intention is making a profit. While presenting Instagram as a tool, that tries its utmost to align with each individual user's goals, it is especially critical for Instagram to convey such a positive portrayal at a time where there is rising criticism of data mining practices, general suspicion of algorithmic content curation, data privacy concerns and the overall harmful practices exhibited by social media enterprises.

Understandably, there is a need for Instagram to maintain a good reputation, leading to potentially deceptive narratives that are meant to obscure the economic motives behind employing recommender algorithms. Therefore, user benefits are numerously emphasized throughout the statements, leading to a strongly positive sentiment across the publications. Yet, in the context of increasing scrutiny over data privacy and transparency claims, Adam Mosseri's statements primarily serve the purpose of soothing public concerns and to deflect criticism. Portraying algorithms as users' "friendly tool for a seamless and entertaining experience" however, skews the public image of what is really happening behind the curtain.

As our detailed examination of existing literature could show, data extraction and processing happens at a scale, that entirely eliminates any doubt about the real intentions, which are creating highly sophisticated profiles of each individual through real-time monitoring each and every action and creating prediction-ready user profiles from hundreds of thousands of datapoints, that are so sophisticated, that an algorithm perhaps knows the person behind those datapoints better than any human ever could; the complexity in processing that such datasets require simply exceeds the capacity of human interaction and cognition, hence well-informed scientists deemed their own findings *creepy*. Again, these practices, true not only for Instagram and Facebook, but the entire industry of social media marketing, are far beyond what Adam Mosseri attempts to convey. It can be safely said, that his publications are understandable, from Instagram's perspective, however they are far from legitimate and are an attempt to sugarcoat the use of algorithmic curation in the first place, while deliberately concealing their inherent intention, vastly downplaying their true scope of manipulation and malign potential.

Heavy doubt can further be cast upon the portrayal of a flourishing environment for creators and users around their close friends and family, positioning Instagram as a societal benefit. As we were able to expose, the entire publication was a mere means to present the platform in a more favorable light amid a general surge in criticism of the industry.

Two things however are certain; neither has Mosseri's statement shed any light on how Instagram works – on the contrary – nor has anything about the industry's practices changed until the present day.

5 Conflict of Interest Statement

As part of our commitment to diligence, academic integrity and transparency, we declare, that none of the authors of this publication have any economic involvement in social media networks or any related activities, and neither does the publisher of this work, the Zentrum für Medienpsychologie und Verhaltensforschung (ZeMV).

While there is no conflict of interest in traditional sense, we consider it important to fortify our credibility and to uphold academic rigor expected in our field. Therefore, while we emphasize, that our research adheres to rigorous academic standards and is impartial and objective, we highlight, that ZeMV is an entity that critically examines the impact of social media in general. This includes Instagram as well as any other social media platform.

In order to ensure our audience is fully informed of our institutional perspective, we hereby confirm that findings and conclusions presented in this publication are based on empirical data and thorough scholarly analyses and are devoid of personal biases, external influences and economic interests.

ZeMV, the publisher of this work, is a not-for-profit scientific entity, hosting scholars that conduct research in media psychology and behavioral science.

References

Bachrach, Y., Kohli, P., Kosinski, M., Stillwell, D., Graepel, T. (2012). Personality and Patterns of Facebook Usage. Microsoft Research. Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/FacebookPersonality_michal_29_04_12.pdf

Bellavista, P., Foschini, L., & Ghiselli, N. (2019). Analysis of Growth Strategies in Social Media: The Instagram Use Case. 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD).

CaPPr. (2015). Interview with Michal Kosinski on Personality and Facebook Likes, May 20, 2015. Retrieved from: https://www.youtube.com/watch?v=pJGuWKqwYRk

Cotter, K. (2018). Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. New Media & Society, 21(4), 895-913.

Daniel, C. Instagram Revenue and Growth Statistics (2024). (2023). SignHouse. Retrieved from: https://www.usesignhouse.com/blog/instagram-stats

Dixon, S. J. (2014). Instagram - statistics & facts. Statista. Retrieved from: https://www.statista.com/topics/1882/instagram/#topicOverview

Iqbal, M. (2024). Instagram Revenue and Usage Statistics. Business of Apps. Retrieved from: https://www.businessofapps.com/data/instagram-statistics/

Jaakonmäki, R., Müller, O., & Brocke, J. (2017). The impact of content, context, and creator on user engagement in social media marketing. HICSS 2017 Proceedings, 1-9.

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. Proceedings of the National Academy of Sciences of the United States of America 110,15, 5802–5.

Mehlhose, F. M., Petrifke, M., & Lindemann, C. (2021). Evaluation of graph-based algorithms for guessing user recommendations of the social network Instagram. 2021 IEEE 15th International Conference on Semantic Computing (ICSC), 409-414.

Mosseri, A. (2021). Shedding More Light on How Instagram Works. Retrieved from: https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works

Mosseri, A. (2023). Instagram Ranking Explained. Retrieved from: https://about.instagram.com/blog/announcements/instagram-ranking-explained

Oliveira, L. M., & Goussevskaia, O. (2020). Sponsored content and user engagement dynamics on Instagram. Proceedings of the 35th Annual ACM Symposium on Applied Computing, 124-131.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D., Ungar, L. H., Seligman, M. (2015). Automatic Personality Assessment Through Social Media Language. Journal of Personality and Social Psychology 108 (6), 934–52.

Purba, K. R., & Yulia, Y. (2021). Realistic influence maximization based on followers score and engagement grade on Instagram. Bulletin of Electrical Engineering and Informatics, 10, 1046-1053.

Purba, K. R., Asirvatham, D., & Murugesan, R. (2020). Classification of Instagram fake users using supervised machine learning algorithms. International Journal of Electrical and Computer Engineering (IJECE), 10(3), 2763-2772.

Purba, K. R., Asirvatham, D., & Murugesan, R. (2021). Instagram post popularity trend analysis and prediction. Proceedings of the International Conference on Information Management and Technology (ICIMTech 2020).

Purba, K. R., Asirvatham, D., & Murugesan, R. (2022). Influence maximization and diffusion models based on engagement and activeness on Instagram. Journal of King Saud University - Computer and Information Sciences, 34 (6), 2831-2839.

Skrubbeltrang, M. M., Grunnet, J., & Tarp, N. T. (2017). #RIPINSTAGRAM: Examining user's counter-narratives opposing the introduction of algorithmic personalization on Instagram. First Monday, 22(4).

Stenovec, T. (2016). How to stop Instagram ads from following you. Business Insider. Retrieved from: https://www.businessinsider.com/how-to-stop-instagram-ads-from-following-you-2016-3

Titlow, J. P. (2017). How Instagram Learns From Your Likes To Keep You Hooked. Fast Company. Retrieved from: https://www.fastcompany.com/40434598/how-instagram-learns-from-your-likes-to-keep-you-hooked

Tricomi, P. P., Chilese, M., Conti, M., & Sadeghi, A. (2023). Follow us and become famous! Insights and guidelines from Instagram engagement mechanisms. Proceedings of the 15th ACM Web Science Conference 2023.

Wang, L., Liu, R., & Vosoughi, S. (2020). Salienteye: Maximizing engagement while maintaining artistic style on Instagram using deep neural networks. Proceedings of the 2020 International Conference on Multimedia Retrieval.

WARC. (2023). Instagram forecast to hit $71bn revenue by 2024. Retrieved from: https://www.warc.com/content/feed/instagram-forecast-to-hit-71bn-revenue-by-2024/en-GB/8650

Wirtschaftspsychologie Aktuell. (2018). Facebook kennt dich besser als deine Freunde. Retrieved from: https://wirtschaftspsychologie-aktuell.de/magazin/leben/facebook-kennt-dich-besser-als-deine-freunde

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences, 112(4), 1036-1040.

Zou, L., Xia, L., Ding, Z., Song, J., Liu, W., & Yin, D. (2019). Reinforcement learning to optimize long-term user engagement in recommender systems. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Public Affairs.

Annex: Code Disclosure (Segment *n* = placeholder)

```
!pip install vaderSentiment

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import pandas as pd

analyzer = SentimentIntensityAnalyzer()

text_segment_n = """

"""

vader_scores_segment_1 = analyzer.polarity_scores(text_segment_1)

# DataFrame visualization
df_vader_segment_1 = pd.DataFrame([vader_scores_segment_1])
print(df_vader_segment_1)
```

```
!pip install flair

import pandas as pd
from flair.models import TextClassifier
from flair.data import Sentence

# Flair sentiment classifier
classifier = TextClassifier.load('sentiment')

text_segment_n = """

"""

sentence_segment_1 = Sentence(text_segment_1)

classifier.predict(sentence_segment_1)

sentiment_dict_segment_1 = sentence_segment_1.labels[0].to_dict()

# DataFrame visualization
df_segment_1 = pd.DataFrame([sentiment_dict_segment_1])
print(df_segment_1)
```

```
!pip install textblob

from textblob import TextBlob

text_segment_n = """

"""

blob_segment_1 = TextBlob(text_segment_1)
sentiment_segment_1 = blob_segment_1.sentiment

# DataFrame visualization
df_blob_segment_1 = pd.DataFrame([{
    'Polarity': sentiment_segment_1.polarity,
    'Subjectivity': sentiment_segment_1.subjectivity
}])
print(df_blob_segment_1)
```